A. K. Bhattacbaryya and S. K. Mullick

On the Bias and Mean Square Error of Some Fertility Measures

1. Introduction

WHILE the formulae for different fertility indicators are commonly given in details in demographic literature, the methods of deriving the sampling fluctuations of estimated rates based on sample studies are lacking. In the present study we try to investigate the nature of bias and sampling variability of the estimators of some of the indicators of fertility such as total fertility rate (TFR), gross reproduction rate (GRR) and net reproduction rate (NRR). We give estimators of MSEs of these rates which can be used for giving confidence intervals for these rates.

It is to be noted that sample TFR is the sum of age specific fertility rates. Moreover age specific fertility rates are ratio of two estimators namely total number of births and number of females. We know that generally ratio estimators are biased estimates of a populat on parameter. Here in particular we investigate the bias of age specific fertility rates for simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRSWOR). Using ^ method of expansion of a ratio estimator we get that these biases are zero upto third order of approximation. This is an important observation that such cases do not necessarily happen for general ratio estimators. As a consequence, upto third order of approximation the biases of TFR, GRR and NRR are zero. Hence the mean square errors (MSE) are approximately the variances- We get expressions of these MSEs upto third order of approximation. We propose estimators for the MSEs of TFR, GRR and NRR based on these approximations. These MSEs can be used to give confidence intervals for the sample TFR, GRR and NRR. Till now these fertility rates wore estimated from samples but no formulae were available for the sampling variances to give confidence intervals. Hence we give explicit expressions for the estimators of MSEs for TFR, GRR and NRR.

1. *Notations and Definitions*

We assume that $U_1, U_2, \ldots, U_F$, is the population of females in the child bearing age range (w_1, w_2). We draw a sample of size f from this population by SRSWR or SRSWOR (to be mentioned specifically when needed). Let i_x denote the sample age specific fertility rate at age x last birth day (l.b.d.) and $'i_x$ denote the sample age specific fertility rate for female births only at age x l.b.d. Here I_x denote the population age specific fertility rate at age x, and $'I_x$ denote the population age specific fertility rate for female births only. The notations for grouped ages are self explanatory. We denote by b_x the number of births for the mothers of age x in the sample and B_x denote the same for the population, whereas f_x denote the number of females at age x in the sample and F_x denote the same in the population. The sample estimate of I_x is i_x which is explicitly b_x/f_x.

We moreover assume that there can be at most one birth to a female because we restrict ourselves to one year data. The incidence of multiple births being negligible this assumption will not create a major problem in the estimation of age specific fertility rates as well as the TFR, GRR, and NRR. In the next section we give some useful results which will be used for the subsequent derivations.

2. Some Preliminary Results

In this section we give expressions for $B(i_x)$ and $E(i_x - I_x)(i_y - I_y)$ which we shall need for deriving the bias and MSE of TFR, GRR and NRR, we shall use the notation WR for with replacement and WOR for without replacement.

We notice that $Ecb_x = B_x$ and $Ecf_x = F_x$ where $c = F/f$. Hence we can use the Δ-method [1] of expansion for i_x. By writing $\Delta_1 = (cb_x - B_x)/B_x$ and $\Delta_2 = (cf_x - F_x)/F_x$ and assuming $|\Delta_1| < 1$ and $|\Delta_2| < 1$ which will hold in most practical examples we ultimately get, neglecting third and higher order terms,

$$B(i_x) = I_x c^2 \left\{ \frac{1}{F_x^2} V(f_x) - \frac{1}{F_x B_x} \text{Cov}\,(b_x, f_x) \right\} \qquad (2.1)$$

$$\text{MSE}\,(i_x) = I_x^2\, c^2 \left\{ \frac{1}{B_x^2} V(b_x) + \frac{1}{F_x^2} V(f_x) - \frac{2}{B_x F_x} \text{Cov}\,(b_x, f_x) \right\} \qquad (2.2)$$

$$E(i_x - I_x)(i_y - I_y) = I_x I_y c^2 \left\{ \frac{\text{Cov}\,(b_x, b_y)}{B_x B_y} - \frac{\text{Cov}\,(b_x, f_y)}{B_x F_y} - \frac{\text{Cov}\,(f_x, b_y)}{F_x B_y} + \frac{\text{Cov}\,(f_x, f_y)}{F_x F_y} \right\} \qquad (2.3)$$

Substituting the expressions for the variances and covariances which can be found from the multinomial model for the case of SRSWR and generalised hypergeometric model for the case of SRSWOR we get the following results.

$$B(i_x)_{\text{WR}} = 0 \tag{2.4}$$

$$B(i_x)_{\text{WOR}} = 0 \tag{2.5}$$

$$\text{MSE}\,(i_x)_{\text{WOR}} = I_x^2\,\frac{F}{f}\left\{\frac{1}{B_x}\left(1-\frac{B_x}{F}\right)-\frac{1}{F_x}\left(1-\frac{F_x}{F}\right)\right\} \tag{2.6}$$

$$\text{MSE}\,(i_x)_{\text{WOR}} = I_x^2\,\frac{(F-f)}{(F-1)}\,\frac{F}{f}\left\{\frac{1}{B_x}\left(1-\frac{B_x}{F}\right)-\frac{1}{F_x}\left(1-\frac{F_x}{F}\right)\right\} \tag{2.7}$$

$$E(i_x - I_x)\,(i_y - I_y)_{\text{WR}} = 0 \tag{2.8}$$

$$E(i_x - I_x)\,(i_y - I_y)_{\text{WOR}} = 0 \tag{2.9}$$

All the above expressions are obtained from (2.1) to (2.3) and hence are exact neglecting third and higher order terms. One thing to mention is that $\text{Cov}(b_x, f_x)$ is not straightforward. We define $g_x = f_x - b_x$. Now

$$(b_x, g_x, f - b_x - g_x) \sim \text{Trinomial}\left(f,\ \frac{B_x}{F},\ \frac{F_x - B_x}{F},\ \frac{F - F_x}{F}\right)$$

and

$$\text{Cov}\,(f_x, b_x) = \text{Cov}\,(b_x, g_x + b_x) = V(b_x) + \text{Cov}\,(b_x, g_x).$$

Similarly, for SRSWOR we find $\text{Cov}\,(b_x, f_x)$ from 3 dimensional hypergeometric distribution.

A salient point to note is that unlike the general case of ratio estimators which are biased even upto third order of approximation $B(i_x)$ turns out to be zero upto third order of approximation. This happens because the expressions for $V(b_x)$, $\text{Cov}\,(b_x, f_x)$ etc. turn out to be such that ultimately the expressions cancel out to zero.

3. Estimators for MSE (i_x)

Now we propose that $\text{MSE}(i_x)$ be estimated by putting unbiased estimates of F_x and B_x in the expressions for $\text{MSE}(i_x)$. Thus we get,

$$\widehat{\text{MSE}}\,(i_x)_{\text{WR}} = i_x^2\left\{\frac{1}{b_x}\left(1-\frac{b_x}{f}\right)-\frac{1}{f_x}\left(1-\frac{f_x}{f}\right)\right\} \tag{3.1}$$

$$\widehat{\text{MSE}}(i_x)_{\text{WOR}} = i_x^2\cdot\frac{(F-f)}{(F-1)}\left\{\frac{1}{b_x}\left(1-\frac{b_x}{f}\right)-\frac{1}{f_x}\left(1-\frac{f_x}{f}\right)\right\} \tag{3.2}$$

4. Bias and MSE of TFR and Estimators of MSE

The sample TFR is given by $\text{TFR} = \sum_{w_1}^{w_2} i_x$ for ungrouped data and for grouped data it is given by $\text{TFR}(G) = 5 \cdot \Sigma\, {}_5i_x$, where ${}_5i_x$ is the age specific fertility for the age group $[x, x+5]$. Therefore Bias of TFR is given by

$$B(\text{TFR}) = E\left(\sum_{w_1}^{w_2} i_x - \sum_{w_1}^{w_2} I_x\right) = \sum_{w_1}^{w_2} B(i_x)$$

Thus, we have

$$B(\text{TFR})_{\text{WR}} \simeq 0 \tag{4.1}$$

and

$$B(\text{TFR})_{\text{WOR}} \cong 0 \tag{4.2}$$

Since everything remains same for the case of the grouped data only i_x is replaced by ${}_5i_x$, B_x is replaced by ${}_5B_x$ and F_x is replaced by ${}_5F_x$; we get

$$B(\text{TFR}(G))_{\text{WR}} \simeq 0 \tag{4.3}$$

$$B(\text{TFR}(G))_{\text{WOR}} \cong 0 \tag{4.4}$$

Thus the MSE of TFR is the approximate variance

$$\begin{aligned} \text{MSE}(\text{TFR}) &= E(\Sigma\, i_x - \Sigma\, I_x)^2 \\ &= \sum_x E(i_x - I_x)^2 + 2 \sum_{x>y} E(i_x - I_x)(i_y - I_y) \end{aligned}$$

Since the second term is zero by Section 2. We get

$$\text{MSE}(\text{TFR}) \cong \sum_x E(i_x - I_x)^2 = \sum_x \text{MSE}(i_x) \tag{4.5}$$

Thus we get

$$\text{MSE}(\text{TFR})_{\text{WR}} \simeq \sum_x I_x^2 \frac{F}{f}\left\{\frac{1}{B_x}\left(1 - \frac{B_x}{F}\right) - \frac{1}{F_x}\left(1 - \frac{F_x}{F}\right)\right\} \tag{4.6}$$

$$\text{MSE}(\text{TFR})_{\text{WOR}} \cong \sum_x I_x^2 \frac{(F-f)F}{(F-1)f}\left\{\frac{1}{B_x}\left(1 - \frac{B_x}{F}\right) - \frac{1}{F_x}\left(1 - \frac{F_x}{F}\right)\right\} \tag{4.7}$$

Similarly for grouped data

$$\text{MSE}(\text{TFR}(G))_{\text{WR}} = 25\left[\sum_x {}_5I_x^2 \frac{F}{f}\left\{\frac{1}{{}_5B_x}\left(1 - \frac{{}_5B_x}{F}\right) - \frac{1}{{}_5F_x}\left(1 - \frac{{}_5F_x}{F}\right)\right\}\right] \tag{4.8}$$

$$\text{MSE}\,(\text{TFR}(G))_{\text{WOR}} = 25\Big[\sum_x {}_5I_x \frac{(F-f)F}{(F-1)f}\Big\{\frac{1}{{}_5B_x}\Big(1-\frac{{}_5B_x}{F}\Big) - \frac{1}{{}_5F_x}\Big(1-\frac{{}_5F_x}{F}\Big)\Big\}\Big] \qquad (4.9)$$

The estimators for the MSEs are obtained by replacing the MSE(i_x)'s by their estimators.

$$M\hat{S}E\,(\text{TFR})_{\text{WR}} = \sum_x i_x^2\Big\{\frac{1}{b_x}\Big(1-\frac{b_x}{f}\Big) - \frac{1}{f_x}\Big(1-\frac{f_x}{f}\Big)\Big\} \qquad (4.10)$$

$$M\hat{S}E\,(\text{TFR})_{\text{WOR}} = \sum_x i_x^2\frac{(F-f)}{(F-1)}\Big\{\frac{1}{b_x}\Big(1-\frac{b_x}{f}\Big) - \frac{1}{f_x}\Big(1-\frac{f_x}{f}\Big)\Big\} \qquad (4.11)$$

$$M\hat{S}E\,(\text{TFR}(G))_{\text{WR}} = 25\Big[\sum_x {}_5i_x^2\Big\{\frac{1}{{}_5b_x}\Big(1-\frac{{}_5b_x}{f}\Big) - \frac{1}{{}_5f_x}\Big(1-\frac{{}_5f_x}{f}\Big)\Big\}\Big] \qquad (4.12)$$

$$M\hat{S}E\,(\text{TFR}(G))_{\text{WOR}} = 25\Big[\sum_x {}_5i_x^2\frac{(F-f)}{(F-1)}\Big\{\frac{1}{{}_5b_x}\Big(1-\frac{{}_5b_x}{f}\Big) - \frac{1}{{}_5f_x}\Big(1-\frac{{}_5f_x}{f}\Big)\Big\}\Big] \qquad (4.13)$$

5. Bias and MSE of GRR and their Estimators

GRR for ungrouped data is given by $\sum_{w_1}^{w_2} {}^fi_x$ and for grouped data it is given by $5 \cdot \sum_x {}_5^fi_x$.

Thus we see in case GRR b_x and B_x is replaced by ${}^fb_x$ and ${}^fB_x$, respectively where ${}^fb_x$ represents female births in the sample, ${}^fB_x$ is the same for the population. Hence the Bias of GRR will be approximately zero as in the case of TFR. The MSE and their estimators can be obtained as in the case of TFR. Since the expression for MSE can be obtained from (4.6) to (4.9) by replacing births by female births, we only mention the estimators of MSE for reference :

$$M\hat{S}E\,(\text{GRR})_{\text{WR}} = \sum_x {}^fi_x^2\Big\{\frac{1}{{}^fb_x}\Big(1-\frac{{}^fb_x}{f}\Big) - \frac{1}{f_x}\Big(1-\frac{f_x}{f}\Big)\Big\} \qquad (5.1)$$

$$M\hat{S}E\,(\text{GRR})_{\text{WOR}} = \sum_x {}^fi_x^2\frac{(F-f)}{(F-1)}\Big\{\frac{1}{{}^fb_x}\Big(1-\frac{{}^fb_x}{f}\Big) - \frac{1}{f_x}\Big(1-\frac{f_x}{f}\Big)\Big\} \qquad (5.2)$$

$$\hat{MSE}\,(GRR(G))_{WR} = 25\left[\sum_x {}_5^f i_x^2 \left\{ \frac{1}{{}_5^f b_x}\left(1 - \frac{{}_5^f b_x}{f}\right) - \frac{1}{{}_5 f_x}\left(1 - \frac{{}_5 f_x}{f}\right)\right\}\right] \tag{5.3}$$

$$\hat{MSE}\,(GRR\,(G))_{WOR} = 25\left[\sum_x {}_5^f i_x^2 \frac{(F-f)}{(F-1)}\left\{\frac{1}{{}_5^f b_x}\left(1 - \frac{{}_5^f b_x}{f}\right) - \frac{1}{{}_5 f_x}\left(1 - \frac{{}_5 f_x}{f}\right)\right\}\right] \tag{5.4}$$

6. Bias and MSE of NRR and their Estimators

The NRR for ungrouped data is given by $1/{}^f l_0\ \Sigma\ {}^f i_x \cdot {}^f L_x$ and for grouped data NRR is given $1/{}^f l_0\ \Sigma\ {}_5^f i_x \cdot {}_5^f L_x$. Since ${}^f l_0$ is the female cohort and ${}^f L_x$ is the female life table stationary population at x, we have ${}^f l_0$ and ${}_5^f L_x$ as constants. Thus as usual we get bias of NRR approximately zero. MSEs can be similarly found out.

The estimators of the MSE of NRR will be given by the following expressions.

$$\hat{MSE}\,(NRR)_{WR} = \frac{1}{{}^f l_0^2}\left[\sum_x {}^f L_x^2\ {}^f i_x^2 \left\{\frac{1}{{}^f b_x}\left(1 - \frac{{}^f b_x}{f}\right) - \frac{1}{f_x}\left(1 - \frac{f_x}{f}\right)\right\}\right] \tag{6.1}$$

$$\hat{MSE}\,(NRR)_{WOR} = \frac{1}{{}^f l_0^2}\left[\sum_x {}^f L_x^2\ {}^f i_x^2 \frac{(F-f)}{(F-1)}\left\{\frac{1}{{}^f b_x}\left(1 - \frac{{}^f b_x}{f}\right) - \frac{1}{f_x}\left(1 - \frac{f_x}{f}\right)\right\}\right] \tag{6.2}$$

$$\hat{MSE}\,(NRR\,(G))_{WR} = \frac{1}{{}^f l_0^2}\left[\sum_x {}_5^f L_x^2\ {}_5^f i_x^2 \left\{\frac{1}{{}_5^f b_x}\left(1 - \frac{{}_5^f b_x}{f}\right) - \frac{1}{{}_5 f_x}\left(1 - \frac{{}_5 f_x}{f}\right)\right\}\right] \tag{6.3}$$

$$\hat{MSE}\,(NRR(G))_{WOR} = \frac{1}{{}^f l_0^2}\left[\sum_x {}_5^f L_x^2\ {}_5^f i_x^2 \frac{(F-f)}{(F-1)}\left\{\frac{1}{{}_5^f b_x}\left(1 - \frac{{}_5^f b_x}{f}\right) - \frac{1}{{}_5 f_x}\left(1 - \frac{{}_5 f_x}{f}\right)\right\}\right] \tag{6.4}$$

(We note that the $\hat{MSE}$s are always non-negative for TFR, GRR, and NRR.)

It is interesting to note that sample TFR, GRR and NRR has lesser approximate variance under SRSWOR than SRSWR. Thus it seems it is reasonable to use SRSWOR. The confidence intervals can be based on the inequalities like $P \{ | \hat{TFR} - TFR | < l\} > 1 - (MSE(TFR)/t2)$ and replacing MSE by $\hat{MSE}$s.

Acknowledgement

We are indebted to Dr. R. Bairagi for proposing the problem and his keen interest while preparing the paper. We are also indebted to Dr. Rahul Mukherjee and Dr. Arijit Chowdhury for going through the preliminary draft and their suggestions for improving the final write up of the paper. We are also grateful to Mr. S. Guha Roy for reviewing the paper before sending for publication, to Mr, K. P. Misra for valuable discussions and to Mr. P. K. Swar for typing this paper with great care.

References

1. Murthy, M. N., 1967, *Sampling Theory and Methods,* Statistical Publishing Society, Calcutta.
2. GOOD, A. M., Gupta, M. K. and Dasgupta, B., 1979, *Fundamentals of Statistics, Vol. II,* The World Prest Private Ltd., Calcutta.
3. Kiih, L., Namboodiri N. K. and Filial, K., 1962, The ratio bias in Surveys. *JASA* 57(300): 863-876, December.