Demography India
Vol22.,No2(1993).,pp. 235-240

A. K. Bhattacharyya*

Bias and Sampling Variability of Some Fertility Measures under Stratified Random Sampling

Introduction

WE have been investigting the sampling variability of fertility and reproduction measures likeTFR,GRR and NRR under different sampling schemes. In our earlier paper (Bhattacharyya and Mullick 1987) we had shown that the estimates of these measures are unbiased under simple random sampling scheme upto third order of approximation. However, under cluster sampling scheme these are biased (Bhattacharyya 1989) and we had derived revised estimators with expressions for MSEs which could be used for giving their confidence intervals. In this paper, we investigate the bias of these measures under stratified random sampling scheme. In particular, we attempt to estimate the bias and give revised estimators of these rates with MSEs. These can be used for estimating confidence intervals as in the prevmious cases.

Notations and Definitions

Let there be F females in the child-bearing age group. Suppose they are divided into L strata. Let F_h denote the size of the hth stratum. Suppose further that a simple random sample without replacement of f_h females is chosen from the hth stratum. Let

F_{hx} = Number of females of age x of the hth stratum;

${}_aF_{hx}$ = Number of females in the age group $(x, x + a)$ in the hth stratum;

B_{hx} = Number of single births to F_{hx} females during the last year;

${}_aB_{hx}$ = Number of single births to ${}_aF_{hx}$ females during the last year;

f_{hx} = Number of females of age x in the hth stratum in the sample;

${}_af_{hx}$ = Number of females of age $(x, x + a)$ in the hth stratum in the sample;

b_{hx} = Number of single births to F_{hx} females during the last year in the sample;

${}_ab_{hx}$ = Number of single births to ${}_af_{hx}$ females during the last year in the sample.

Let $B_x = \sum_h B_{hx}$, $F_x = \sum_h f_{hx}$, $I_x = \dfrac{B_x}{F_x}$

* Population Studies Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Calcutta- 700 035.

I_x is the age-specific fertility rate.

Let
$$i_x = \frac{\sum_h \frac{F_h}{f_h} b_{hx}}{\sum_h \frac{F_h}{f_h} f_{hx}} = \frac{b_x}{f_x} \tag{1}$$

Clearly, b_x and f_x are unbiased estimators of B_x and F_x. $T\hat{F}R = \sum_x i_x$. Using similar notations for grouped data, $T\hat{F}R\,(G) = 5 \cdot \sum_x 5^{i_x}$.

Some Preliminary Results

Here, we give expressions for $B(i_x)$, $MSE(i_x)$ and Mean Product Error $(MPE(i_x, i_y)) = E(i_x - I_x)(i_y - I_y)$ which we shall need for deriving the bias and MSE of TFR, GRR and NRR.

We notice that $Eb_x = B_x$ and $Ef_x = F_x$. Hence we can use the Δ method of expansion for i_x writing $\Delta_1 = (b_x - B_x)/B_x$ and $\Delta_2 = (f_x - F_x)/F_x$ and assuming $|\Delta_1| < |$ and $|\Delta_2| < |$ which will hold in most practical examples we ultimately get, neglecting third and higher order terms,

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

$$MSE\ (i_x) = I_x^2 \left\{ \frac{1}{B_x^2} V(b_x) + \frac{1}{F_x^2} V(f_x) - \frac{2}{B_x F_x} \text{Cov}\ (b_x, f_x) \right\} \tag{3}$$

$$MPE(i_x, i_y) = I_x I_y \left\{ \frac{\text{Cov}\ (b_x, b_y)}{B_x B_y} - \frac{\text{Cov}\ (b_x, f_y)}{B_x F_y} - \frac{\text{Cov}\ (f_x, b_y)}{F_x B_y} + \frac{\text{Cov}\ (f_x, f_y)}{F_x F_y} \right\} \tag{4}$$

Substituting the expressions for variance and conariance terms which can be found out by using simple random sampling within the Strata (Bhattacharyya and Mullick 1987), we get

$$B(i_x) = I_x \left[\frac{1}{F_x^2} \sum_h \left\{ \frac{(F_h^2)}{f_h} \frac{(F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h}\right) \right\} - \frac{1}{F_x B_x} \sum_h \left\{ \frac{F_h^2}{f_h} \frac{(F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h}\right) \right\} \right] \tag{5}$$

Similarly,

$$MSE\ (i_x) = I_x^2 \left[\frac{1}{B_x^2} \sum_h \left\{ \frac{F_h^2}{f_h} \frac{(F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \left(1 - \frac{B_{hx}}{F_h}\right) \right\} + \frac{1}{F_x^2} \sum_h \left\{ \frac{F_h^2}{f_h} \frac{(F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h}\right) \right\} \right.$$

$$- \frac{1}{F_x B_x} \sum_h \left\{ \frac{F_h^2}{f_h} \frac{(F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h} \right) \right\} \Bigg] \tag{6}$$

and

$$MPE\,(i_x, i_y) = -I_x I_y \left[\sum_h \frac{F_h^2 (F_h - f_h)}{f_h (F_h - 1)} \left\{ \frac{1}{B_x b_y} \frac{B_{hx}}{F_h} \cdot \frac{B_{hy}}{F_h} \right. \right.$$

$$\left. \left. - \frac{1}{B_x F_y} \frac{B_{hx}}{F_h} \frac{F_{hy}}{F_h} - \frac{1}{F_x B_y} \frac{F_{hx}}{F_h} \cdot \frac{B_{hy}}{F_h} + \frac{1}{F_x F_y} \cdot \frac{F_{hx}}{F_h} \cdot \frac{F_{hy}}{F_h} \right\} \right] \tag{7}$$

We find that $B(i_x)$ is non-zero and we give estimators of Bias, $MSE(i_x)$ and $MPE(i_x, i_y)$, by substituting unbiased estimators of $B_x, F_x, F_{hx}, B_{hx}, B_{hy}$ and F_{hy} in the respective expressions. We explicity give the expressions for $\hat{B}(i_x)$, $\hat{MSE}(i_x)$ and $\hat{MPE}(i_x, i_y)$ as they will be needed to find out the revised estimators and their MSE's.

Estimators of $B(i_x)$, $MSE(i_x)$ and $MPE(i_x, i_y)$

$$\hat{B}(i_x) = i_x \left[\sum_h \frac{F_h^2 (F_h - f_h)}{f_h (F_h - 1)} \left\{ \frac{1}{f_x^2} \frac{f_{hx}}{f_h} \left(1 - \frac{f_{hx}}{f_h} \right) \right. \right.$$

$$\left. \left. - \frac{1}{f_x b_x} \frac{b_{hx}}{f_h} \left(1 - \frac{f_{hx}}{f_h} \right) \right\} \right] \tag{8}$$

$$\hat{MSE}\,(i_x) = i_x^2 \left[\sum_h \frac{F_h^2 (F_h - f_h)}{f_h (F_h - 1)} \left\{ \frac{1}{b_x^2} \frac{b_{hx}}{f_h} \left(1 - \frac{b_{hx}}{f_h} \right) \right. \right.$$

$$\left. \left. + \frac{1}{f_x^2} \frac{f_{hx}}{f_h} \left(1 - \frac{f_{hx}}{f_h} \right) - \frac{2}{f_x b_x} \frac{b_{hx}}{f_h} \left(1 - \frac{f_{hx}}{f_h} \right) \right\} \right] \tag{9}$$

$$\hat{MPE}\,(i_x, i_y) = -i_x i_y \left[\sum_h \frac{F_h^2 (F_h - f_h)}{f_h (F_h - 1)} \left\{ \frac{1}{b_x b_y} \frac{b_{hx} b_{hy}}{f_h^2} - \frac{1}{b_x f_y} \frac{b_{hx} f_{hy}}{f_h^2} \right. \right.$$

$$\left. \left. - \frac{1}{f_x b_y} \frac{f_{hx} b_{hy}}{f_h^2} + \frac{1}{f_x f_y} \frac{f_{hx} f_{hy}}{f_h^2} \right\} \right] \tag{10}$$

Revised Estimates of Age-Specific Fertility Rates and their Estimator of MSE

As shown in the case of cluster sampling scheme, we have

$$\hat{I}_x = i_x - \hat{B}\,(i_x) \tag{11}$$

$$\hat{MSE}(\hat{I}_x) = \hat{MSE}(i_x) - \{\hat{B}(i_x)\}^2 \tag{12}$$

$$\hat{MPE}(\hat{I}_x, \hat{I}_y) = \hat{MPE}(i_x, i_y) - \hat{B}(i_x)\hat{B}(i_y) \tag{13}$$

Revised Estimator of TFR, GRR and NRR

We denote the corrected estimators by subscript c :

$$TFR_c = TFR - \sum_x \hat{B}(i_x) \tag{14}$$

$$TFR_c(G) = TFR(G) - 5 \cdot \sum_x \hat{B}({}_5 i_x) \tag{15}$$

$$GRR_c = GRR - \sum_x \hat{B}(f_{i_x}) \tag{16}$$

$$GRR_c(G) = GRR(G) - 5 \cdot \sum_x \hat{B}({}_5 f_{i_x}) \tag{17}$$

$$NRR_c = NRR - \frac{1}{f_{l_0}} \sum_x f_{L_x} \hat{B}(f_{i_x}) \tag{18}$$

$$NRR_c(G) = NRR(G) - \frac{1}{f_{l_0}} \sum_x {}_5 f_{L_x} \hat{B}({}_5 f_{i_x}) \tag{19}$$

Estimators of the MSEs of the Revised Fertility Rates

As in the case of cluster sampling, we get similar expressions for the MSEs of the revised fertility rates

$$\hat{MSE}(TFR_c) = \sum_x \hat{MSE}(i_x) + 2 \sum_{x>y} \hat{MPE}(i_x, i_y) - \left\{\sum_x \hat{B}(i_x)\right\}^2 \tag{20}$$

$$\hat{MSE}(TFR_c(G)) = 25 \cdot \left[\sum_x \hat{MSE}({}_5 i_x) + 2 \sum_{x>y} \hat{MPE}({}_5 i_x, {}_5 i_y) - \left\{\sum_x \hat{B}({}_5 i_x)\right\}^2\right] \tag{21}$$

$$MSE(GRR_c) = \sum_x \hat{MSE}(f_{i_x}) + 2 \sum_{x>y} \hat{MPE}(f_{i_x}, f_{i_y}) - \left\{\sum_x \hat{B}(f_{i_x})\right\}^2 \tag{22}$$

$$\hat{MSE}(GRR_c(G)) = 25 \cdot \left[\sum_x \hat{MSE}({}_5 f_{i_x}) + 2 \sum_{x>y} \hat{MPE}({}_5 f_{i_x}, {}_5 f_{i_y}) - \left\{\sum_x \hat{B}({}_5 f_{i_x})\right\}^2\right] \tag{23}$$

$$\hat{MSE}(NRR_c) = \frac{2}{f_{l_0}^2}\left[\sum_x f_{Lx}^2 \hat{MSE}(f_{i_x}) + 2 \sum_{x>y} f_{L_x} f_{L_y} \hat{MPE}(f_{i_x}, f_{i_y}) - \left\{\sum_x f_{L_x} \hat{B}(f_{i_x})\right\}^2\right] \tag{24}$$

$$\hat{MSE}\,(NRR_c\,(G)) = \frac{1}{f_{l_0}^2}\left[\sum_x {}_5f_x^2\,\hat{MSE}\;{}_5f_{i_x} + 2\sum_{x>y} {}_5f_{L_x}\,{}_5f_{L_y}\,\hat{MPE}\,({}_5f_{i_x}, {}_5f_{i_y}) - \left\{\sum_x {}_5f_{L_x}\,\hat{B}\,({}_5f_{i_x})\right\}^2\right] \quad (25)$$

Conclusion

In this paper, while investigating the bias and sampling variablity of fertility measures under stratified random sampling, we have given the explicit expressions for the $\hat{MSE}$s of corrected estimators so that they can be used for giving confidence intervals.

References

Bhattacharyya, A. K. and Mullick, S. K. (1987), On the bias and mean square error of some fertility measures. *Demography India* 16 (2) : 288-294.

Bhattacharyya, A. K. (1989), Sampling bias MSE and confidence intervals for some fertility measures under cluster sampling scheme. *Sankhya, Series B*, 51(1) : 134-139.

APPENDIX

In this appendix, we give various expressions in this paper. We note that

$$E(b_{hx}) = \frac{f_h}{F_h} B_{hx} \tag{1}$$

$$E(f_{hx}) = \frac{f_h}{F_h} B_{hx} \tag{2}$$

$$V(f_{hx}) = \frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h}\right) \tag{3}$$

$$V(b_{hx}) = \frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \left(1 - \frac{B_{hx}}{F_h}\right) \tag{4}$$

$$\mathrm{Cov}(b_{hx}, f_{hx}) = \frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \left(1 - \frac{F_{hx}}{F_h}\right) \tag{5}$$

$$\mathrm{Cov}(b_{hx}, b_{hy}) = -\frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \cdot \frac{B_{hy}}{F_h} \tag{6}$$

$$\mathrm{Cov}(b_{hx}, f_{hy}) = -\frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{B_{hx}}{F_h} \cdot \frac{F_{hy}}{F_h} \tag{7}$$

$$\mathrm{Cov}(f_{hx}, b_{hy}) = -\frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \cdot \frac{B_{hy}}{F_h} \tag{8}$$

$$\mathrm{Cov}(f_{hx}, f_{hy}) = -\frac{f_h (F_h - f_h)}{(F_h - 1)} \frac{F_{hx}}{F_h} \cdot \frac{F_{hy}}{F_h} \tag{9}$$

These exprssions can be found out from first principles or may be derived by using 3-dimensional hypergeometric distribution.