RT-MAE-8904

AN INVESTIGATION OF THE PROPERTIES OF RAKING RATIO ESTIMATORS FOR CELL FREQUENCIES WITH SIMPLE RANDOM SAMPLING

by

Maria Cristina Martín

and

Wilton de O. Bussab

Palavras Chaves: Conditional bias and variance, Raking,
(Key words) ratio estimation, simple random sampling

Classificação AMS: 62D05, 62P12
(AMS Classification)

- Fevereiro 1989 -

An investigation of the properties of Raking Ratio Estimators for Cell Frequencies with Simple Random Sampling

by

Maria Cristina Martín [1] and *Wilton de Oliveira Bussab* [2]

Summary:

On the basis of a simple random sample from a population, on which a cross classification is defined with known marginal frequencies $N_{i.}$ and $N_{.j}$, one wishes to estimate the cell frequencies N_{ij}. Various authors have discussed so-called Raking Ratio Estimators, which are calculated iteratively. The purpose of this paper is to obtain asymptotic expressions for the bias, variances and covariances for the corresponding estimators of the cell frequencies for any numbers of iterations ($t \geq 0$) and to show a different result of that presented by Konijn (1981).

Key Words: Raking ratio estimation, Conditional bias and variance.

1.- Introduction

Consider a population of N units from which a simple random sample of n units is drawn. Supose both the sample and population units are cross-classified in a two-dimensional matrix that is defined in terms of basic characteristics with known marginal frequencies, say $N_{i.}$ and $N_{.j}$. Let n_{ij}, N_{ij} be respectively the sample and population counts in the (i,j)-th cell of this matrix ($N = \sum_i \sum_j N_{ij}$, $n = \sum_i \sum_j n_{ij}$).

This set up, following Deming and Stephan (1940), is shown in the tables (1.1) and (1.2).

(1) Facultad de Ciencias Exactas y Naturales - Universidad Nacional de La Pampa - Argentina

(2) Instituto de Matemática e Estatística - Universidade de São Paulo - Brasil

Table (1.1): Classification for the universe

	1	2	...	c	
1	N_{11}	N_{12}	...	N_{1c}	$N_{1.}$
2	N_{21}	N_{22}	...	N_{2c}	$N_{2.}$
i	⋮	⋮	N_{ij}	⋮	$N_{i.}$
r	N_{r1}	N_{r2}	...	N_{rc}	$N_{r.}$
	$N_{.1}$	$N_{.2}$	$N_{.j}$	$N_{.c}$	

Tabela (1.2): Classification for the sample

	1	2	...	c	
1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
i	⋮	⋮	n_{ij}	⋮	$n_{i.}$
r	n_{r1}	n_{r2}	...	n_{rc}	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.c}$	

Our objective, in the next section, will be to estimate the interior of the table and mainly to develop the asymptotic bias and variances of these estimators. We use to a two dimensional iterative procedure that will produce approximate sample-population adequacy simultaneously for two basic marginal distribution. This process is called Raking

Ratio Estimation Procedure (RREP). The resulting estimators for the cell frequencies are called Raking Ratio Estimators (RRE).

In Section 2 we describe the RREP. In Section 3 we develop the asymptotic bias, variance and covariances of the RRE for any iteration.

2.- The Raking Ratio Estimators Procedure: System of Weight and Estimators

In estimating any cell frequency of the universe, such as N_{ij}, we have three possibilities:

i.- from the over-all units of population, the estimator is $N\,\frac{n_{ij}}{n}$;

ii.- from the i-th row alone, the estimator is $N_{i\cdot}\,\frac{n_{ij}}{n_{i\cdot}}$;

iii.- from the j-th column alone, the estimator is $N_{\cdot j}\,\frac{n_{ij}}{n_{\cdot j}}$.

In general, the three estimators above will not be equal, and not necessarily the estimated marginal frequencies will be identical with the known marginal frequencies.

The **RREP** works as a sequence of iterations starting with rows or columns, and combines all three of the estimators just mentioned to give a new estimator which is more precise. Starting with rows, the weights are defined as follow:

Definition (2.1): Let $W_{ij}^{(t)}$ be the t-th iteration weight in cell (i,j). Then $W_{ij}^{(t)}$ is given by:

$$W_{ij}^{(t)} = \begin{cases} \frac{N}{n}, & \text{if } t=0, \text{ for all } i,\ j; \\ W_{ij}^{(t-1)} \frac{N_{\cdot j}}{\sum_{i=1}^{r} W_{ij}^{(t-1)} n_{ij}}, & \text{if } t \text{ even;} \\ W_{ij}^{(t-1)} \frac{N_{i\cdot}}{\sum_{j=1}^{c} W_{ij}^{(t-1)} n_{ij}}, & \text{if } t \text{ odd.} \end{cases} \tag{2.1}$$

Definition (2.2): The Raking Ratio Estimator **RRE** of the cell frequency N_{ij} in the t-th iteration of the procedure is

$$\widehat{N}_{ij}^{(t)} = W_{ij}^{(t)}\, n_{ij} \quad (i=1,\ldots,r; j=1,\ldots,c). \tag{2.2}$$

Using (2.1), (2.2) can be written as follows:

$$\widehat{N}_{ij}^{(t)} = \begin{cases} \frac{N}{n} n_{ij}, & \text{if } t=0; \\ N_{i\cdot} \frac{\widehat{N}_{ij}^{(t-1)}}{\widehat{N}_{i\cdot}^{(t-1)}}, & \text{if } t \text{ odd;} \\ N_{\cdot j} \frac{\widehat{N}_{ij}^{(t-1)}}{\widehat{N}_{\cdot j}^{(t-1)}}, & \text{if } t \text{ even.} \end{cases} \tag{2.3}$$

The weights used for the cells, and thus the estimated frequencies, are defined recursively, in such a way that at the odd iterations, the estimated marginal row frequencies coincide with the known marginal row frequencies (i.e., $\widehat{N}_{i\cdot}^{(2t-1)} = N_{i\cdot}$), and that at the even iterations, the estimated marginal column frequencies coincide with the known marginal column frequencies (i.e., $\widehat{N}_{\cdot j}^{(2t)} = N_{\cdot j}$). Moreover, the convergence of $N_{\cdot j}^{(2t-1)}$ to $N_{\cdot j}$ and $N_{i\cdot}^{(2t)}$ to $N_{i\cdot}$ is generally rapid.

The RREP thus is asymetric, i.e. the starting choice (rows and columns) affects the results. Brackstone and Rao (1979) (see also Arora and Brackstone (1977)) give some suggestions about this problem. In this paper, all the properties are derived starting the adjustement with rows.

3.- Properties of the Raking Ratio Estimator

In this section we explore the bias, variance and covariance of the estimators $\widehat{N}_{ij}^{(t)}$ for any t. Konijn (1981) derives these properties for $t = 0, 1$, and 2 and there, it was stated that " $Cov(\widehat{N}_{ij}^{(2)}, \widehat{N}_{i'j'}^{(2)})$ gives 0 for $i \neq i'$". Here we show that this is not true (see expression (3.3.4), theorem (3.3)).

In the calculus below, the subscript 2 denotes that we take the conditional operator given $n_{i\cdot}$, and the subscript 1 denotes that the means and variances are unconditional, that is, the variation in the rows $n_{i\cdot}$, and in the columns $n_{\cdot j}$ are present.

3.1.- Bias and Variance of the No-iteration Estimator

In simple random sampling, the RREP begins by setting

$$\widehat{N}_{ij}^{(0)} = N\,\frac{n_{ij}}{n}, \quad (i = 1, \ldots, r; j = 1, \ldots, c). \tag{3.1.1}$$

Theorem (3.1): The no-iteration estimator $\widehat{N}_{ij}^{(0)}$ is unbiased with variance:

$$V(\widehat{N}_{ij}^{(0)}) = \frac{(N-n)}{(N-1)}\frac{1}{n}N_{ij}(N - N_{ij}); \tag{3.1.2}$$

and

$$Cov(\widehat{N}_{ij}^{(0)}, \widehat{N}_{i'j'}^{(0)}) = -\frac{(N-n)}{(N-1)}\frac{1}{n}N_{ij}N_{i'j'}, \quad \text{for any } i, i', j, \text{ and } j'. \tag{3.1.3}$$

proof:

Since in the case $t = 0$ we are not using the additional information about the marginals, the theorem can be obtained by properties of the multinomial distribution. Nevertheless, we give the proof concerning to the mean and variance by conditioning to the row because will be useful for the higher iterations.

$\widehat{N}_{ij}^{(0)}$ is an unbiased estimator of N_{ij}:

$$E(\widehat{N}_{ij}^{(0)}) = E_1(E_2(N\frac{n_{ij}}{n})) = E_1\left(\frac{N}{n}\, n_{i\cdot}E_2(\frac{n_{ij}}{n_{i\cdot}})\right) = N\rho_{ij}E_1(\frac{n_{i\cdot}}{n}) = N\rho_{ij}\frac{N_{i\cdot}}{N} = N_{ij},$$

where $\rho_{ij} = \frac{N_{ij}}{N_{i\cdot}}$ is the relative frequency in the i-th row of the population.

To obtain the variance of $\widehat{N}_{ij}^{(0)}$ we use $V(\widehat{N}_{ij}^{(0)}) = V_1(E_2(\widehat{N}_{ij}^{(0)})) + E_1(V_2(\widehat{N}_{ij}^{(0)}))$. But

$$V_1(E_2(\widehat{N}_{ij}^{(0)})) = (N\rho_{ij})^2\left[\frac{N_{i\cdot}}{N}(1-\frac{N_{i\cdot}}{N})\frac{1}{n}(\frac{N-n}{N-1})\right] = \rho_{ij}^2\frac{N_{i\cdot}(N-N_{i\cdot})(N-n)}{n(N-1)}. \quad (3.1.4)$$

On the other hand:

$$E_1(V_2(\widehat{N}_{ij}^{(0)})) = E_1\left[\frac{N^2}{n^2}n_{i\cdot}^2V_2(\frac{n_{ij}}{n_{i\cdot}})\right] = E_1\left[\frac{N^2}{n^2}n_{i\cdot}^2\left(\frac{N_{ij}}{N_{i\cdot}}(1-\frac{N_{ij}}{N_{i\cdot}})\frac{1}{n_{i\cdot}}(\frac{N_{i\cdot}-n_{i\cdot}}{N_{i\cdot}-1})\right)\right]$$

$$= \frac{N^2}{n^2}\rho_{ij}(1-\rho_{ij})\frac{N_{i\cdot}}{N_{i\cdot}-1}E_1\left[n_{i\cdot}^2(\frac{1}{n_{i\cdot}}-\frac{1}{N_{i\cdot}})\right]. \quad (3.1.5)$$

Let

$$\alpha_i = N_{i\cdot}^2(\frac{1}{n_{i\cdot}}-\frac{1}{N_{i\cdot}}) \quad \text{and} \quad \beta_i = \frac{N_{i\cdot}}{N_{i\cdot}-1}\alpha_i. \quad (3.1.6)$$

Note that:

$$E\left[(\frac{n_{i\cdot}}{N_{i\cdot}})^2\alpha_i\right] = n^2(\frac{1}{n}-\frac{1}{N})\left[\frac{N_{i\cdot}}{N}-\frac{1}{N-1}(1-\frac{N_{i\cdot}}{N})\right],$$

and then, (3.1.5) is

$$E_1(V_2(\widehat{N}_{ij}^{(0)})) = \rho_{ij}(1-\rho_{ij})\frac{N_{i\cdot}(N-n)N}{n(N-1)}. \quad (3.1.7)$$

The variance of $\widehat{N}_{ij}^{(0)}$ is the sum of expressions (3.1.4) and (3.1.7). This sum satisfies the equality (3.1.2). ■

3.2.- Bias, Variance and Covariance of the One-iteration Estimator

In agreement with (2.3)

$$\widehat{N}_{ij}^{(1)} = N_{i\cdot}\frac{\widehat{N}_{ij}^{(0)}}{\widehat{N}_{i\cdot}^{(0)}} = N_{i\cdot}\frac{\frac{N}{n}n_{ij}}{\sum_{j=1}^{c}\frac{N}{n}n_{ij}} = N_{i\cdot}\frac{n_{ij}}{n_{i\cdot}}. \quad (3.2.1)$$

Theorem (3.2): The RRE $\widehat{N}_{ij}^{(1)}$ is unbiased with variance:

$$V(\widehat{N}_{ij}^{(1)}) = \rho_{ij}(1-\rho_{ij})B_i; \tag{3.2.2}$$

and

$$Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j'}^{(1)}) = \begin{cases} 0, & \text{if } i \neq i'; \\ -\rho_{ij}\,\rho_{ij'}\,B_i, & \text{if } i = i',\ j \neq j'; \end{cases} \tag{3.2.3}$$

with $B_i = E(\beta_i) = \left[\frac{N_{i.}(N-n)[N_{i.}(n-1)+N]}{(N_{i.}-1)n^2}\right]$.

proof:

$\widehat{N}_{ij}^{(1)}$ is an unbiased estimator of N_{ij}:

$$E(\widehat{N}_{ij}^{(1)}) = E_1(E_2(N_{i.}\frac{n_{ij}}{n_{i.}})) = E_1(N_{i.}\frac{N_{ij}}{N_{i.}}) = E_1(N_{ij}) = N_{ij}. \tag{3.2.4}$$

To obtain the variance of $\widehat{N}_{ij}^{(1)}$ note that $V_1(E_2(\widehat{N}_{ij}^{(1)})) = V_1(N_{ij}) = 0.$ and then

$$V(\widehat{N}_{ij}^{(1)}) = E_1(V_2(\widehat{N}_{ij}^{(1)})) = E_1\left(V_2(N_{i.}\frac{n_{ij}}{n_{i.}})\right)$$
$$= E_1\left[\rho_{ij}(1-\rho_{ij})\frac{N_{i.}}{N_{i.}-1}N_{i.}^2(\frac{1}{n_{i.}} - \frac{1}{N_{i.}})\right] = \rho_{ij}(1-\rho_{ij})B_i$$

this expression checks (3.2.2).

Now

$$Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j'}^{(1)}) = Cov_1\left(E_2(N_{i.}\frac{n_{ij}}{n_{i.}}), E_2(N_{i'.}\frac{n_{i'j'}}{n_{i'}})\right) + E_1\left(Cov_2(N_{i.}\frac{n_{ij}}{n_{i.}}, N_{i'.}\frac{n_{i'j'}}{n_{i'.}})\right).$$

Note that for any $i, i', j,$ and j' (see (3.2.2)) is

$$Cov_1\left(E_2(N_{i.}\frac{n_{ij}}{n_{i.}}), E_2(N_{i'.}\frac{n_{i'j'}}{n_{i'.}})\right) = Cov_1(N_{ij}, N_{i'j'}) = 0.$$

Besides that, for $i \neq i'$, by properties of the conditional multinomial distributions, we have $Cov_2\left(N_{i.}\frac{n_{ij}}{n_{i.}}, N_{i'.}\frac{n_{i'j'}}{n_{i'.}}\right) = 0$. Thus

$$Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j'}^{(1)}) = 0 \quad \text{if} \quad i \neq i'; \tag{3.2.5}$$

and for $i = i'$ and $j \neq j'$

$$Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij'}^{(1)}) = E_1\left[Cov_2(\frac{N_{i.}}{n_{i.}}n_{ij}, \frac{N_{i.}}{n_{i.}}n_{ij'})\right] = E_1\left[N_{i.}^2\left(-\frac{N_{i.}-n_{i.}}{N_{i.}-1}\frac{1}{n_{i.}}\rho_{ij}\rho_{ij'}\right)\right]$$

so that,

$$Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij'}^{(1)}) = -E_1(\beta_i\ \rho_{ij}\ \rho_{ij'}) = -\rho_{ij}\ \rho_{ij'}\ B_i.$$

It remains to calculate $B_i = E(\beta_i) = \frac{N_{i.}}{N_{i.}-1}E(\alpha_i)$, with β_i and α_i defined in (3.1.6). But for moderately large n and the condition $\frac{1}{N}$ negligible so that $\frac{N}{N-1} \approx 1$, it is easy to proof:

$$E(\alpha_i) \approx N^2\left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{N_{i.}}{N} + \frac{1}{n}(1 - \frac{N_{i.}}{N})\right],$$

(see Martín (1988), p. 16-17). ■

3.3. Bias, Variance and Covariance of the two-iteration estimator

When $t = 2$, using (2.3), we have

$$\widehat{N}_{ij}^{(2)} = N_{.j}\frac{\widehat{N}_{ij}^{(1)}}{\widehat{N}_{.j}^{(1)}}. \tag{3.3.1}$$

Theorem (2.3): The RRE $\widehat{N}_{ij}^{(2)}$ verifies the following properties:

$$B(\widehat{N}_{ij}^{(2)}) \approx -\frac{1}{N_{.j}}\left[B_i\rho_{ij}(1-\rho_{ij}) - \kappa_{ij}\sum_{i'} B_{i'}\rho_{i'j}(1-\rho_{i'j})\right]; \tag{3.3.2}$$

$$EQM(\widehat{N}_{ij}^{(2)}) \approx B_i\rho_{ij}(1-\rho_{ij})(1-2\kappa_{ij}) + \kappa_{ij}^2\sum_{i'} B_{i'}\rho_{i'j}(1-\rho_{i'j}); \tag{3.3.3}$$

and an approximation to the covariance between $\widehat{N}_{ij}^{(2)}$ and $\widehat{N}_{i'j'}^{(2)}$ is

$$\begin{cases} \kappa_{ij}(B_{i'}\rho_{i'j}\rho_{i'j'}) + \kappa_{i'j'}(B_i\rho_{ij}\rho_{ij'}) - \kappa_{ij}\kappa_{i'j'}\sum_a(B_a\rho_{aj}\rho_{aj'}), & \text{if } i \neq i', j \neq j'\ ; \\ -(1-\kappa_{ij}-\kappa_{ij'})(B_i\rho_{ij}\rho_{ij'}) - \kappa_{ij}\kappa_{ij'}\sum_a(B_a\rho_{aj}\rho_{aj'}), & \text{if } i = i', j \neq j'; \\ -\kappa_{ij}(B_{i'}\rho_{i'j}(1-\rho_{i'j})) - \kappa_{i'j}(B_i\rho_{ij}(1-\rho_{ij})) & \\ +\kappa_{ij}\kappa_{i'j}\sum_a(B_a\rho_{aj}(1-\rho_{aj})) & \text{if } i \neq i', j = j'. \end{cases} \tag{3.3.4}$$

where $\kappa_{ij} = \frac{E(\widehat{N}_{ij}^{(1)})}{E(N_{.j}^{(1)})} = \frac{N_{ij}}{N_{.j}}$ is the relative frequency in the j-th column of the population.

proof:

An approximation to the bias of the $\widehat{N}_{ij}^{(2)}$ may be obtained as the bias of a ratio estimator (see Martín (1988), Appendix A, for instance), and we have

$$B(\widehat{N}_{ij}^{(2)}) \approx -\frac{1}{N_{.j}}Cov\left(\widehat{N}_{.j}^{(1)}, \widehat{N}_{ij}^{(1)} - \frac{N_{ij}}{N_{.j}}\widehat{N}_{.j}^{(1)}\right)$$

$$= -\frac{1}{N_{.j}}\left[E_1\left(Cov_2(\widehat{N}_{.j}^{(1)}, \widehat{N}_{ij}^{(1)} - \kappa_{ij}\widehat{N}_{.j}^{(1)})\right) + Cov_1\left(E_2(\widehat{N}_{ij}^{(1)}), E_2(\widehat{N}_{ij}^{(1)} - \kappa_{i,j}\widehat{N}_{ij}^{(1)})\right)\right]$$

Since the conditional mean of $\widehat{N}_{ij}^{(1)}$ (and also $\widehat{N}_{.j}^{(1)}$) are non-random, we need only find the mean of the conditional covariance. In another words

$$
\begin{aligned}
B(\widehat{N}_{ij}^{(2)}) &\approx -\frac{1}{N_{.j}} E_1\left[Cov_2(\widehat{N}_{.j}^{(1)}, \widehat{N}_{ij}^{(1)} - \kappa_{ij}\widehat{N}_{.j}^{(1)})\right] \\
&= -\frac{1}{N_{.j}} E_1\left[Cov_2(\widehat{N}_{.j}^{(1)}, \widehat{N}_{ij}^{(1)}) - \kappa_{ij}V_2(\widehat{N}_{.j}^{(1)})\right. \\
&= -\frac{1}{N_{.j}} E_1\left[\sum_{i'} Cov_2(\widehat{N}_{i'j}^{(1)}, \widehat{N}_{ij}^{(1)}) - \kappa_{ij}V_2(\sum_{i'} \widehat{N}_{ij}^{(1)})\right].
\end{aligned} \tag{3.3.5}
$$

Since for $i \neq i'$, $\widehat{N}_{i'j}^{(1)}$ is conditionally independent of $\widehat{N}_{ij}^{(1)}$ we have

$$
\sum_{i'} Cov_2(\widehat{N}_{i'j}^{(1)}, \widehat{N}_{ij}^{(1)}) = Cov_2(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij}^{(1)}) = V_2(\widehat{N}_{ij}^{(1)}) \tag{3.3.6}
$$

and moreover

$$
V_2(\widehat{N}_{.j}^{(1)}) = \sum_{i'} V_2(\widehat{N}_{i'j}^{(1)}) + \sum_{i' \neq i} Cov_2(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j}^{(1)}) = \sum_{i'} V_2(\widehat{N}_{i'j}^{(1)}). \tag{3.3.7}
$$

From here, to obtain (3.3.2), it suffices to use, in an appropriate way, the equalites mentioned and then to remember that $E_1(V_2(\widehat{N}_{ij}^{(1)})) = V(\widehat{N}_{ij}^{(1)})$.

On the other hand, an approximation to the mean square error of $\widehat{N}_{ij}^{(2)}$ is

$$
\begin{aligned}
EQM(\widehat{N}_{ij}^{(2)}) &\approx V(N_{ij}^{(1)} - \kappa_{ij}N_{.j}^{(1)}) \\
&= V(\widehat{N}_{ij}^{(1)}) - 2\kappa_{ij}Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{.j}^{(1)}) + \kappa_{ij}^2 V(\widehat{N}_{.j}^{(1)}) \\
&= V(\widehat{N}_{ij}^{(1)}) - 2\kappa_{ij}V(\widehat{N}_{ij}^{(1)}) + \kappa_{ij}^2 \sum_{i'} V(\widehat{N}_{i'j}^{(1)})
\end{aligned}
$$

this last equality is obtained using (3.3.6) and (3.3.7). Using now (3.3.2), we may write (3.3.3).

Finally, we can find an approximation to $Cov(\widehat{N}_{ij}^{(2)}, \widehat{N}_{i'j'}^{(2)})$, for any i, i', j, and j' from

$$
\begin{aligned}
Cov(\widehat{N}_{ij}^{(2)}, \widehat{N}_{i'j'}^{(2)}) &\approx Cov(\widehat{N}_{ij}^{(1)} - \kappa_{ij}\widehat{N}_{.j}^{(1)}, \widehat{N}_{i'j'} - k_{i'j'}\widehat{N}_{.j'}^{(1)}) \\
&= Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j'}^{(1)}) - \kappa_{ij}Cov(\widehat{N}_{.j}^{(1)}, \widehat{N}_{i'j'}^{(1)}) - \kappa_{i'j'}Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{.j'}^{(1)}) \\
&\quad + \kappa_{ij}\kappa_{i'j'}Cov(\widehat{N}_{.j}^{(1)}, \widehat{N}_{.j'}^{(1)}).
\end{aligned}
$$

But

(a) for $i \neq i'$, $Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{i'j'}^{(1)})$ gives zero,

(b) $Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{.j'}^{(1)}) = \sum_a Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{aj'}^{(1)}) = Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij'}^{(1)})$, the last equality is obtained using that the sum is non-null if and only if $a = i$.

For the same argument, we have

$$Cov(\widehat{N}_{.j}^{(1)}, \widehat{N}_{i'j'}^{(1)}) = Cov(\widehat{N}_{i'j}^{(1)}, \widehat{N}_{i'j'}^{(1)}),$$

and

$$Cov(\widehat{N}_{.j}^{(1)}, \widehat{N}_{.j'}^{(1)}) = \sum_a Cov(\widehat{N}_{aj}^{(1)}, \widehat{N}_{aj'}^{(1)}).$$

Considering these facts and the different cases for i, i', j and j' we have an approximation to the covariance between $\widehat{N}_{ij}^{(2)}$ and $\widehat{N}_{i'j'}^{(2)}$, which is given by

$$\begin{cases} -\kappa_{ij}Cov(\widehat{N}_{i'j}^{(1)}, \widehat{N}_{i'j'}^{(1)}) - \kappa_{i'j'}Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij'}^{(1)}) & \\ +\kappa_{ij}\kappa_{i'j'}\sum_a Cov(\widehat{N}_{aj}^{(1)}, \widehat{N}_{aj'}^{(1)}), & \text{if } i \neq i',\ j \neq j' \\ (1 - \kappa_{ij} - \kappa_{ij'})Cov(\widehat{N}_{ij}^{(1)}, \widehat{N}_{ij'}^{(1)}) + \kappa_{ij}\kappa_{ij'}\sum_a Cov(\widehat{N}_{aj}^{(1)}, \widehat{N}_{aj'}^{(1)}), & \text{if } i = i', j \neq j' \\ -\kappa_{ij}V(\widehat{N}_{i'j}^{(1)}) - \kappa_{i'j}V(\widehat{N}_{ij}^{(1)}) + \kappa_{ij}\kappa_{i'j}\sum_a V(\widehat{N}_{aj}^{(1)}), & \text{if } i \neq i', j = j'. \end{cases}$$

From here, using (3.2.2) and (3.2.3) results (3.3.4). ■

2.3.4 Generalizations

An approximation to the bias of $\widehat{N}_{ij}^{(t)}$ for any t, may be obtained using the same approximation that the case $t = 2$ for $\widehat{N}_{ij}^{(t)} = N_{i.}\frac{\widehat{N}_{ij}^{(t-1)}}{\widehat{N}_{i.}^{(t-1)}}$ if t odd or $\widehat{N}_{ij}^{(t)} = N_{.j}\frac{\widehat{N}_{ij}^{(t-1)}}{\widehat{N}_{.j}^{(t-1)}}$ if t even. So, after the feasible development, we have:

$$B(\widehat{N}_{ij}^{(2t)}) \approx B(\widehat{N}_{ij}^{(2t-1)}) - \kappa_{ij}^{(2t-1)}B(\widehat{N}_{.j}^{(2t-1)}) - \frac{1}{E(\widehat{N}_{.j}^{2(t-1)})}\left(1 - \frac{B(\widehat{N}_{.j}^{(2t-1)})}{E(\widehat{N}_{.j}^{(2t-1)})}\right)$$
$$\left[\sum_a Cov(\widehat{N}_{aj}^{(2t-1)}, \widehat{N}_{ij}^{(2t-1)}) - \kappa_{ij}^{(2t-1)}V(\widehat{N}_{.j}^{(2t-1)})\right],$$

and

$$B(\widehat{N}_{ij}^{(2t+1)}) \approx B(\widehat{N}_{ij}^{(2t)}) - \rho_{ij}^{(2t)}B(\widehat{N}_{i.}^{(2t)}) - \frac{1}{E(\widehat{N}_{i.}^{(2t)})}\left(1 - \frac{B(\widehat{N}_{i.}^{(2t)})}{E(\widehat{N}_{i.}^{(2t)})}\right)$$
$$\left[\sum_b Cov(\widehat{N}_{ib}^{(2t)}, \widehat{N}_{ij}^{(2t)}) - \rho_{ij}^{(2t)}V(\widehat{N}_{i.}^{(2t)})\right],$$

with $\rho_{ij}^{(2t)} = \frac{E(\widehat{N}_{ij}^{(2t)})}{E(\widehat{N}_{i.}^{(2t)})}$ and $\kappa_{ij}^{(2t-1)} = \frac{E(\widehat{N}_{ij}^{(2t-1)})}{E(\widehat{N}_{.j}^{(2t-1)})}$.

Using the large-sample approximation for the mean square error of a ratio and noting that the estimated marginal frequencies are asymptotically unbiased (for example, in the case even that is $EQM(\widehat{N}_{ij}^{(t)}) \approx V(\widehat{N}_{ij}^{(t-1)} - \rho_{ij}^{(t-1)}\widehat{N}_{i.}^{(t-1)})$), it is easy to verify

$$EQM(\widehat{N}_{ij}^{(2t)}) \approx V(\widehat{N}_{ij}^{(2t-1)}) - \kappa_{ij}^{(2t-1)}\left[2\sum_a Cov(\widehat{N}_{ij}^{(2t-1)}, \widehat{N}_{aj}^{(2t-1)}) - \kappa_{ij}^{(2t-1)}V(\widehat{N}_{.j}^{(2t-1)})\right],$$

and

$$EQM(\widehat{N}_{ij}^{(2t+1)}) \approx V(\widehat{N}_{ij}^{(2t)}) - \rho_{ij}^{(2t)}\left[2\sum_b Cov(\widehat{N}_{ij}^{(2t)}, \widehat{N}_{ib}^{(2t)}) - \rho_{ij}^{(2t)}V(\widehat{N}_{i.}^{(2t)})\right].$$

For similar properties, an approximation to the covariance between $\widehat{N}_{ij}^{(t)}$ and $\widehat{N}_{i'j'}^{(t)}$ for any t is given by:

$$\begin{aligned} Cov(\widehat{N}_{ij}^{(2t)}, \widehat{N}_{i'j'}^{(2t)}) \approx\ & Cov(\widehat{N}_{ij}^{(2t-1)}, \widehat{N}_{i'j'}^{(2t-1)}) - \kappa_{ij}^{(2t-1)}\sum_a Cov(\widehat{N}_{aj}^{(2t-1)}, \widehat{N}_{i'j'}^{(2t-1)}) \\ & - \kappa_{i'j'}^{(2t-1)}\sum_a Cov(\widehat{N}_{ij}^{(2t-1)}, \widehat{N}_{aj'}^{(2t-1)}) \\ & + \kappa_{ij}^{(2t-1)}\kappa_{i'j'}^{(2t-1)}\sum_a\sum_b Cov(\widehat{N}_{aj}^{(2t-1)}, \widehat{N}_{bj'}^{(2t-1)}), \end{aligned}$$

and

$$\begin{aligned} Cov(\widehat{N}_{ij}^{(2t+1)}, \widehat{N}_{i'j'}^{(2t+1)}) \approx\ & Cov(\widehat{N}_{ij}^{(2t)}, \widehat{N}_{i'j'}^{(2t)}) - \rho_{ij}^{(2t)}\sum_b Cov(\widehat{N}_{ib}^{(2t)}, \widehat{N}_{i'j'}^{(2t)}) \\ & - \rho_{i'j'}^{(2t)}\sum_b Cov(\widehat{N}_{ij}^{(2t)}, \widehat{N}_{i'b}^{(2t)}) + \rho_{ij}^{(2t)}\rho_{i'j'}^{(2t)}\sum_a\sum_b Cov(\widehat{N}_{ia}^{(2t)}, \widehat{N}_{i'b}^{(2t)}). \end{aligned}$$

These two expressions serve as basis to derive the covariances expressions, but still we may distinguish three cases: covariances between estimated cell frequencies in the same row, covariances between estimated cell frequencies in the same column and covariances between estimated cell frequencies in different row and column (see (3.3.4) and recall that each step uses the results of the previous step).

References

ARORA, H.R. and BRACKSTONE, G. J. (1977): An Investigation of the Properties of Raking Ratio Estimators with Simple Random Sampling. *Survey Methodology*, 3, 62-83.

BRACKSTONE, G.J.and RAO, J.N.K. (1979): An Investigation of Raking Ratio Estimators.*Sankhyā*, Vol 41, Series C, Pt2, 97-114.

DEMING, W.E. and STEPHAN, F.F. (1940): On a Least Squares Adjustment of a Sampled Frecuency Table when the Expected Marginal Totals are known. *Ann. Math. Statist.*, 11, 427-444.

KONIJN, H.S. (1981): Biases, Variances and Covariances of Raking Ratio Estimators for Marginal and Cell Totals and Averages of Observed Characteristics. *Metrika*, 28, 109-201.

MARTIN, M.C. (1988): Procedimento Iterativo de Estimação por Totais Marginais Fixos. *Dissertação de Mestrado.* IME-USP, 135p.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; Inverse Gaussian Distribution: A Bayesian Approach, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; On the Hardy -Weinberg Equilibrium in Generalized ABO Systems, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; Generalizing Importance of Components for Multistate Monotone Systems, São Paulo, IME-USP, 1987, 12p.

8801 - PEREIRA, C.A.B. & WECHSLER, S.; On the Concept of P-value, São Paulo, IME-USP, 1988, 22p.

8802 - ZACKS, S., PEREIRA, C.A.B. & LEITE, J.G.; Bayes Sequential Estimation of the Size of a Finite Population, São Paulo,IME-USP, 1988, 23p.

8803 - BOLFARINE, H.; Finite Population Prediction Under Dynamic Generalized Linear Models, São Paulo, IME-USP, 1988, 21p.

8804 - BOLFARINE, H.; Minimax Prediction in Finite Populations, São Paulo, IME-USP, 1988, 18p.

8805 - SINGER, J.M. & ANDRADE, D.F.; On the Choice of Appropriate Error Terms for Testing the General Linear Hypothesis in Profile Analysis, São Paulo, IME-USP, 1988, 23p.

8806 - DACHS, J.N.W. & PAULA, G.A.; Testing for Ordered Rate Rations in Follow-up Studies with Incidence Density Data, São Paulo, IME-USP, 1988, 18p.

8807 - CORDEIRO, G.M. & PAULA, G.A.; Estimation, Significance Tests and Diagnostic Methods for the Non-Exponential Family Nonlinear Models, São Paulo, IME-USP, 1988, 29p.

8808 - RODRIGUES, J. & ELIAN, S.N.; The Coordinate - Free Estimation in Finite Population Sampling, São Paulo, IME-USP, 1988, 5p.

8809 - BUENO, V.C. & CUADRADO, R.Z.B.; On the Importance of Components for Continuous Structures, São Paulo, IME-USP, 1988, 14p.

8810 - ACHCAR, J.A., BOLFARINE, H & PERICCHI, L.R.; Some Applications of Bayesian Methods in Analysis of Life Data, São Paulo, IME-USP, 1988, 30p.

8811 - RODRIGUES, J.; A Bayesian Analysis of Capture-Recapture Experiments for a Closed Animal Population, São Paulo, IME-USP, 1988, 10p.

8812 - FERRARI, P.A.; Ergodicity for Spin Systems, São Paulo, IME-USP, 1988, 25p.

8813 - FERRARI, P.A. & MAURO, E.S.R.; A Method to Combine Pseudo-Random Number Generators Using Xor, São Paulo, IME-USP, 1988, 10p.

8814 - BOLFARINE, H. & RODRIGUES, J.; Finite Population Prediction Under a Linear Functional Superpopulation Model a Bayesian Perspective, São Paulo, IME-USP, 1988, 22p.

8815 - RODRIGUES, J. & BOLFARINE, H.; A Note on Asymptotically Unbiased Designs in Survey Sampling, São Paulo, IME-USP, 1988, 6p.

8816 - BUENO, V.C.; Bounds for the a Availabilities in a Fixed Time Interval for Continuous Structures Functions, São Paulo, IME-USP, 1988, 22 p.

8817 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Estimation for Time Series with Amplitude Modulated Observations: A Review, São Paulo, IME-USP, 1988, 16p.

8818 - CHAYES, J.T.; CHAYES, L,; GRIMMETT, G,R,; KESTEN, H. & SCHONMANN, R.H,; The Correlation Length for the High Density Phase of Bernoulli Percolation, São Paulo, IME-USP, 1988, 46p.

8819 - DURRETT, R,; SCHONMANN, R.H. & TANAKA, N.I.: The Contact Process on a Finite Set, III: The Critical Case, São Paulo, IME-USP, 1988, 31p.

8820 - DURRET, R,; SCHONMANN, R.H. & TANAKA, N.I.; Correlation Lengths for Oriented Percolation, São Paulo, IME-USP, 1988, 18p.

8821 - BRICMONT, J.; KESTEN, H,; LEBOWITZ, J.L. & SCHONMANN, R,H,; A Note on the Ising Model in High Dimensions, São Paulo, IME-USP, 1988, 21p.

8822 - KESTEN, H, & SCHONMANN, R,H,; Behavior in Large Dimensions of the Potts and Heisenberg Models, São Paulo, IME-USP, 1988, 61p.

8823 - DURRET, R, & TANAKA, N.I.; Scaling Inequalities for Oriented Percolation, São Paulo, IME-USP, 1988, 21p.

8901 - RODRIGUES, J.; Asymptotically Design - Unbiased Predictors to Two-Stage Sampling, São Paulo, IME-USP, 1989, 9p.

8902 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Analysis for Amplitude Modulated Time Series, São Paulo, IME-USP, 1989, 24p.

8903 - PAULA, G.A.; Influence Measures for Generalized Linear Models with Restrictions in Parameters, São Paulo, IME-USP, 1989, 18p.