Optimum ratio estimators for the population proportion-CMMSE-10

Juan Francisco Muñoz, Encarnación Álvarez, Antonio Arcos, Maria del Mar Rueda, Silvia Gonzalez, Agustín Santiago

▶ **To cite this version:**

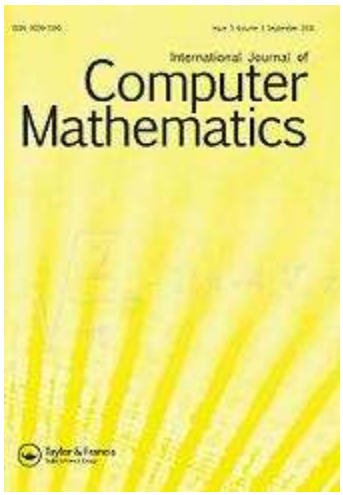
Juan Francisco Muñoz, Encarnación Álvarez, Antonio Arcos, Maria del Mar Rueda, Silvia Gonzalez, et al.. Optimum ratio estimators for the population proportion-CMMSE-10. International Journal of Computer Mathematics, 2011, pp.1. 10.1080/00207160.2011.587512 . hal-00722238

HAL Id: hal-00722238
https://hal.science/hal-00722238

Submitted on 1 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Optimum ratio estimators for the population proportion-CMMSE-10

Journal:	*International Journal of Computer Mathematics*
Manuscript ID:	GCOM-2010-0717-B.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	07-Feb-2011
Complete List of Authors:	Muñoz, Juan Francisco; University of Granada Álvarez, Encarnación; University of Granada Arcos, Antonio; University of Granada Rueda, Maria del Mar; University of Granada Gonzalez, Silvia; University of Jaen, Statistics and Operational Research Santiago, Agustín; University of Guerrero
Keywords:	Auxiliary information, ratio type estimator, survey sampling, population proportion estimation, simple random sample

SCHOLARONE™
Manuscripts

International Journal of Computer Mathematics
Vol. 00, No. 00, January 2008, 1–10

RESEARCH ARTICLE

Optimum ratio estimators for the population proportion-CMMSE-10

Juan Francisco Muñoz[a], Encarnación Álvarez[b], Antonio Arcos[c], María del Mar Rueda [d*], Silvia González[e] and A. Santiago[f]

[ab] *Department of Quantitative Methods for Economy and Enterprise, University of Granada, Spain* ; [cd] *Department of Statistics and Operational Research, University of Granada, Spain* ; [e] *Department of Statistics and Operational Research, University of Jaén, Spain* ; [e] *Faculty of Mathematics , University of Guerrero, Mexico*

(*v3.6 released September 2008*)

The problem of the estimation of a population proportion using auxiliary information has been recently studied by Rueda et al. (2011), which proposed several ratio estimators of the population proportion and studied some theoretical properties. In this paper, we define a new ratio estimator based on a linear combination of two ratio estimators defined by Rueda et al. (2011). The variance of the new estimator is calculated and it is used to obtain the optimum value into the linear combination in the sense of minimal variance. Theoretical and empirical studies show that the suggested ratio estimator performs better than alternative estimators.

Keywords: Auxiliary information, ratio type estimator, proportion estimation, finite population, simple random sampling

AMS Subject Classification: 62D05

1. Introduction

In the presence of auxiliary information, there exist many design-based approaches (see [4], [6], [2]) to improve the precision of estimators in comparison to customary methods, which do not involve auxiliary information. However, techniques involving auxiliary information have been discussed for quantitative variables, and the extension to the estimation of a population proportion requires further investigation. For example, one should be aware of the risks when confidence intervals are constructed for a population proportion, since limits outside $[0, 1]$ could be achieved.

We consider the scenario of a finite population $U = \{1, \ldots, N\}$ containing N units. Let $A_1, \ldots, A_N$ denote the values of an attribute of interest A, where $A_i = 1$ if ith unit possesses the attribute A and $A_i = 0$ otherwise. Let B denote an auxiliary attribute associated with A and values given by $B_1, \ldots, B_N$. We also assume that a sample s, of size n, is selected from U according to the well known simple random sampling without replacement (SRSWOR).

The aim is to estimate the population proportion of individuals that posses the attribute A, i.e. $P_A = N^{-1} \sum_{i=1}^{N} A_i$. Assuming a finite population, the naive

*Corresponding author. Email: mrueda@ugr.es

ISSN: 0020-7160 print/ISSN 1029-0265 online
© 2008 Taylor & Francis
DOI: 0020716YYxxxxxxxx
http://www.informaworld.com

For Peer Review Only

estimator of P_A, which makes no use of the auxiliary information, is given by $\widehat{p}_A = n^{-1}\sum_{i\in s} A_i$.

We assume that the population proportion of individuals that posses the attribute B, $P_B = N^{-1}\sum_{i=1}^{N} B_i$, is known from a census or estimated without error.

Rueda et al (2011) defined the following ratio estimator for P_A:

$$\widehat{p}_r = \widehat{R}P_B, \tag{1}$$

where $\widehat{R} = \widehat{p}_A/\widehat{p}_B$ is an estimator of the population ratio $R = P_A/P_B$ and $\widehat{p}_B = n^{-1}\sum_{i\in s} B_i$ is the sample proportion of individuals that posses the auxiliary attribute B.

Let A^c and B^c denote the complementary attributes of A and B, and consider the population two-way table given by

	B	B^c	
A	N_{11}	N_{12}	$N_{1\cdot}$
A^c	N_{21}	N_{22}	$N_{2\cdot}$
	$N_{\cdot 1}$	$N_{\cdot 2}$	N

(2)

where $N_{1\cdot} = \sum_{i=1}^{N} A_i$ is the number of units in the population that posses the attribute A, $N_{2\cdot}$ is the number of units in the population that not to posses the attribute A, etc. Analogously, N_{11} is the number of units in the population that simultaneously posses the attributes A and B, N_{12} is the number of units in the population that simultaneously posses the attributes A and B^c, etc. Classification (2) can be also defined at the sample level as

	B	B^c	
A	n_{11}	n_{12}	$n_{1\cdot}$
A^c	n_{21}	n_{22}	$n_{2\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	n

(3)

The estimator $\widehat{p}_r$ is a biased estimator of P_A and the asymptotic variance of $\widehat{p}_r$ is given by

$$AV(\widehat{p}_r) = \frac{N-n}{(N-1)n}\left(P_A Q_A + R^2 P_B Q_B - 2R\phi\sqrt{P_A Q_A P_B Q_B}\right), \tag{4}$$

where

$$\phi = \frac{N_{11}N_{22} - N_{12}N_{21}}{\sqrt{N_{1\cdot}N_{2\cdot}N_{\cdot 1}N_{\cdot 2}}}$$

is the Cramer's V coefficient ([1]) based on the two-way classification (2).

We observe that the customary estimator $\widehat{p}_A$ can be also obtained as $\widehat{p}_A = 1-\widehat{q}_A$, where $\widehat{q}_A = n^{-1}\sum_{i\in s} A_i^c$, hence $\widehat{p}_A$ has the same performance in the estimation of P_A then the performance of $\widehat{q}_A$ in the estimation of Q_A. However, this property is not satisfied by $\widehat{p}_r$, i.e. it can be easily seen that $\widehat{p}_r \neq 1-\widehat{q}_r$, where $\widehat{q}_r = \widehat{R}_c Q_B$ and $\widehat{R}_c = (\widehat{q}_A/\widehat{q}_B)$. For this reason, Rueda et al. (2011) defined the ratio estimator $\widehat{p}_{r.q} = 1-\widehat{q}_r$ for P_A and showed that $AV(\widehat{p}_r) < AV(\widehat{p}_{r.q})$ when $P_A < P_B$.

For Peer Review Only

2. The optimum ratio estimator

In this section, we define a new ratio type estimator using a linear combination of the ratio estimators $\widehat{p}_r$ and $\widehat{p}_{r.q}$ previously defined. The choice of the optimum weight value into the linear combination is achieved by minimizing the variance. Finally, some interesting theoretical properties are also obtained.

The new ratio type estimator is

$$\widehat{p}_{r.w} = w\widehat{p}_r + (1-w)\widehat{p}_{r.q}. \tag{5}$$

THEOREM 2.1

The optimum value for w in the sense of minimum variance into the class of estimators $\widehat{p}_{r.w}$ is

$$w_{opt} = \frac{AV(\widehat{p}_{r.q}) - cov(\widehat{p}_r, \widehat{p}_{r.q})}{AV(\widehat{p}_r) + AV(\widehat{p}_{r.q}) - 2cov(\widehat{p}_r, \widehat{p}_{r.q})}. \tag{6}$$

Proof

Next, we determine the optimum value of w by minimizing the variance of $\widehat{p}_{r.w}$. The asymptotic variance of $\widehat{p}_{r.w}$ is given by

$$AV(\widehat{p}_{r.w}) = AV(w\widehat{p}_r + (1-w)\widehat{p}_{r.q}) =$$

$$= w^2 AV(\widehat{p}_r) + (1-w)^2 AV(\widehat{p}_{r.q}) + 2w(1-w)cov(\widehat{p}_r, \widehat{p}_{rq}).$$

By denoting $V_1 = AV(\widehat{p}_r)$, $V_2 = V(\widehat{p}_{r.q})$ and $C = cov(\widehat{p}_r, \widehat{p}_{r.q})$, the variance of $\widehat{p}_{r.w}$ can be expressed as

$$AV(\widehat{p}_{r.w}) = w^2 V_1 + (1-w)^2 V_2 + 2w(1-w)C.$$

The first derivative of $AV(\widehat{p}_{r.w})$ with respect to w is

$$\frac{\partial AV(\widehat{p}_{r.w})}{\partial w} = 2wV_1 - 2(1-w)V_2 + 2(1-2w)C = 0;$$

$$w_{opt} = \frac{V_2 - C}{V_1 + V_2 - 2C}.$$

The second derivative is

$$\frac{\partial AV(\widehat{p}_{r.w})}{\partial^2 w} = 2V_1 + 2V_2 - 4C = 2(V_1 + V_2 - 2C) = 2AV(\widehat{p}_r - \widehat{p}_{r.q}) > 0,$$

and we conclude that w_{opt} really minimizes $AV(\widehat{p}_{r.w})$.

■

For Peer Review Only

Therefore, the optimum ratio estimator in the sense of minimum variance into the class (5) is

$$\widehat{p}_{r.OPT} = w_{opt}\widehat{p}_r + (1 - w_{opt})\widehat{p}_{r.q}.$$

In practice, $\widehat{p}_{r.OPT}$ could be unknown, since w_{opt} depends on population variances, which are generally unknown. In this situation, we can use the estimator

$$\widehat{p}_{r.opt} = \widehat{w}_{opt}\widehat{p}_r + (1 - \widehat{w}_{opt})\widehat{p}_{r.q}, \tag{7}$$

where

$$\widehat{w}_{opt} = \frac{\widehat{V}(\widehat{p}_{r.q}) - \widehat{cov}(\widehat{p}_r, \widehat{p}_{r.q})}{\widehat{V}(\widehat{p}_r) + \widehat{V}(\widehat{p}_{r.q}) - 2\widehat{cov}(\widehat{p}_r, \widehat{p}_{r.q})}. \tag{8}$$

Following Särndal et al. (1992) pg 372, the variance of $\widehat{p}_{r.w}$ can be expressed as

$$AV(\widehat{p}_{r.w}) = (V_1 + V_2 - 2C)\left(w - \frac{V_2 - C}{V_1 + V_2 - 2C}\right)^2 + \frac{V_1V_2 - C^2}{V_1 + V_2 - 2C},$$

and we can deduce that the variance of the optimum estimator is

$$AV(\widehat{p}_{r.OPT}) = \frac{V_1V_2 - C^2}{V_1 + V_2 - 2C}.$$

An estimator of the variance of the optimum estimator can be obtained as

$$\widehat{V}(\widehat{p}_{r.opt}) = \frac{\widehat{V}(\widehat{p}_r)\widehat{V}(\widehat{p}_{r.q}) - \widehat{cov}^2(\widehat{p}_r, \widehat{p}_{r.q})}{\widehat{V}(\widehat{p}_r) + \widehat{V}(\widehat{p}_{r.q}) - 2\widehat{cov}(\widehat{p}_r, \widehat{p}_{r.q})}.$$

We observe that $AV(\widehat{p}_{r.OPT})$ depends on the covariance $C = cov(\widehat{p}_r, \widehat{p}_{r.q})$. Theorem 2 gives an expression for C.

THEOREM 2.2 *The covariance between the ratio estimators $\widehat{p}_r$ and $\widehat{p}_{r.q}$ is*

$$cov(\widehat{p}_r, \widehat{p}_{r.q}) = \frac{N - n}{N - 1}\frac{1}{n}\left(P_AQ_A + RR_cP_BQ_B - (R + R_c)\phi\sqrt{P_AQ_AP_BQ_B}\right),$$

where $R_c = Q_A/Q_B$ is the population ratio of the complementary proportions of the attributes A and B.

Proof

Using Taylor series (see Särndal et al. 1992, pg 178), $\widehat{R}$ can be expressed as

$$\widehat{R} \cong R + \frac{1}{nP_B}\sum_{i \in s}(A_i - RB_i) = R + \frac{1}{P_B}(\widehat{p}_A - R\widehat{p}_B),$$

and similarly

$$\widehat{R}_c \cong R_c + \frac{1}{Q_B}(\widehat{q}_A - R_c\widehat{q}_B).$$

For Peer Review Only

Using the previous expressions we obtain

$$C = cov(\widehat{p}_r, 1 - \widehat{q}_r) = -cov(\widehat{R}P_B, \widehat{R}_c Q_B) =$$

$$-P_B Q_B cov\left(R + \frac{1}{P_B}(\widehat{p}_A - R\widehat{P}_B), R_c + \frac{1}{Q_B}(\widehat{q}_A - R_c\widehat{q}_B)\right) =$$

$$= V(\widehat{p}_A) + RR_c V(\widehat{p}_B) - (R + R_c)cov(\widehat{p}_A, \widehat{p}_B) =$$

$$= \frac{N-n}{N-1}\frac{1}{n}\left(P_A Q_A + RR_c P_B Q_B - (R + R_c)\phi\sqrt{P_A Q_A P_B Q_B}\right).$$

■

An estimator of the covariance $cov(\widehat{p}_r, \widehat{p}_{r.q})$ is

$$\widehat{cov}(\widehat{p}_r, \widehat{p}_{r.q}) = \frac{1-f}{n-1}\left(\widehat{p}_A\widehat{q}_A + \widehat{R}\widehat{R}_c\widehat{p}_B\widehat{q}_B - (\widehat{R} + \widehat{R}_c)\widehat{\phi}\sqrt{\widehat{p}_A\widehat{q}_A\widehat{p}_B\widehat{q}_B}\right),$$

where

$$\widehat{\phi} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}}.$$

THEOREM 2.3
The optimum weight w_{opt} in expression (6) can be expressed as

$$w_{opt} = \frac{R_c - \beta}{R_c - R},$$

where

$$\beta = \frac{cov(\widehat{p}_A, \widehat{p}_B)}{V(\widehat{p}_B)}.$$

Proof

Knowing that

$$V_1 = V(\widehat{p}_A) + R^2 V(\widehat{p}_B) - 2Rcov(\widehat{p}_A, \widehat{p}_B),$$

$$V_2 = V(\widehat{q}_A) + R_c^2 V(\widehat{q}_B) - 2R_c cov(\widehat{q}_A, \widehat{q}_B) =$$

$$= V(\widehat{p}_A) + R_c^2 V(\widehat{p}_B) - 2R_c cov(\widehat{p}_A, \widehat{p}_B)$$

For Peer Review Only

and

$$C = V(\widehat{p}_A) + RR_cV(\widehat{p}_B) - (R + R_c)cov(\widehat{p}_A, \widehat{p}_B),$$

the numerator and the denominator of w_{opt} in (6) are given by

$$V_2 - C = V(\widehat{p}_B)(R_c^2 - RR_c) - cov(\widehat{p}_A, \widehat{p}_B)[2R_c - (R - R_c)] =$$

$$= V(\widehat{p}_B)R_c(R_c - R) - cov(\widehat{p}_A, \widehat{p}_B)(R_c - R) =$$

$$= (R_c - R)[V(\widehat{p}_B)R_c - cov(\widehat{p}_A, \widehat{p}_B)]$$

and

$$V_1 + V_2 - 2C = V(\widehat{p}_B)(R^2 + R_c^2 - 2RR_c) - cov(\widehat{p}_A, \widehat{p}_B)[2R + 2R_c - 2(R + R_c)] =$$

$$= V(\widehat{p}_B)(R_c - R)^2$$

By replacing these expressions in (6) we obtain

$$w_{opt} = \frac{V_2 - C}{V_1 + V_2 - 2C} = \frac{(R_c - R)[V(\widehat{p}_B)R_c - cov(\widehat{p}_A, \widehat{p}_B)]}{V(\widehat{p}_B)(R_c - R)^2} =$$

$$= \frac{V(\widehat{p}_B)R_c - cov(\widehat{p}_A, \widehat{p}_B)}{(R_c - R)V(\widehat{p}_B)} = \frac{R_c - \beta}{R_c - R}.$$

■

Following theorem 2.3, the estimated optimum weight $\widehat{w}_{opt}$ given by (8) can be calculated as

$$\widehat{w}_{opt} = \frac{\widehat{R}_c - \widehat{\beta}}{\widehat{R}_c - \widehat{R}}, \tag{9}$$

where

$$\widehat{\beta} = \frac{\widehat{cov}(\widehat{p}_A, \widehat{p}_B)}{\widehat{V}(\widehat{p}_B)}.$$

From expression (9) we conclude that $\widehat{w}_{opt} = 1$, that is $\widehat{p}_{r.opt} = \widehat{p}_r$, if $\widehat{\beta} = \widehat{R}$, whereas $\widehat{w}_{opt} = 0$, that is $\widehat{p}_{r.opt} = \widehat{p}_{r.q}$, if $\widehat{\beta} = \widehat{R}_c$. In other words, the ratio estimator $\widehat{p}_r$ has a larger weight into the optimum estimator $\widehat{p}_{r.opt}$ as $\widehat{\beta}$ is closer to $\widehat{R}$. On the

For Peer Review Only

other hand, the ratio estimator $\widehat{p}_{r.q}$ has a larger weight into the optimum estimator $\widehat{p}_{r.opt}$ as $\widehat{\beta}$ is closer to $\widehat{R}_c$.

THEOREM 2.4

The asymptotic variance of the optimum ratio estimator $\widehat{p}_{r.OPT}$ can be calculated as

$$AV(\widehat{p}_{r.OPT}) = V(\widehat{p}_A)(1-\phi^2).$$

Proof

The asymptotic variance of $\widehat{p}_{r.OPT}$ is

$$AV(\widehat{p}_{r.OPT}) = \frac{V_1V_2 - C}{V_1 + V_2 - 2C},$$

where the denominator, as seen in proof of theorem 2.3, can be obtained as

$$V_1 + V_2 - 2C = V(\widehat{p}_B)(R - R_c)^2.$$

Next, we obtain the numerator of $AV(\widehat{p}_{r.OPT})$. For the sake of simplicity, we denote $V_A = V(\widehat{p}_A)$, $V_B = V(\widehat{p}_B)$ and $C_{AB} = cov(\widehat{p}_A, \widehat{p}_B)$. We had that

$$V_1 = V_A + R^2V_B - 2RC_{AB},$$

$$V_2 = V_A + R_c^2V_B - 2R_cC_{AB}$$

and

$$C = V_A + RR_cV_B - (R + R_c)C_{AB}.$$

First,

$$V_1V_2 = V_A^2 + R_c^2V_AV_B - 2R_cV_AC_{AB} + R^2V_AV_B + R^2R_c^2V_B^2$$

$$-2R^2R_cV_BC_{AB} - 2RV_AC_{AB} - 2RR_c^2V_BC_{AB} + 4RR_cC_{AB}^2.$$

The square of the covariance can be expressed as

$$C^2 = V_A^2 + R^2R_c^2V_B^2 + (R + R_c)^2C_{AB}^2 + 2V_ARR_cV_B$$

$$-2(R + R_c)V_AC_{AB} - 2RR_c(R + R_c)V_BC_{AB} =$$

$$= V_A^2 + R^2R_c^2V_B^2 + R^2C_{AB}^2 + R_c^2C_{AB}^2 + 2RR_cC_{AB}^2 + 2V_ARR_cV_B$$

For Peer Review Only

$$-2RV_AC_{AB} - 2R_cV_AC_{AB} - 2R^2R_cV_BC_{AB} - 2RR_c^2V_BC_{AB}.$$

Then, the numerator of $AV(\widehat{p}_{r.OPT})$ is

$$V_1V_2 - C^2 = V_AV_B(R_c^2 - 2RR_c + R^2) - C_{AB}^2(R^2 + R_c^2 - 2RR_c) =$$

$$= (V_AV_B - C_{AB}^2)(R - R_c)^2.$$

The variance of $\widehat{p}_{r.OPT}$ can be also obtained as

$$AV(\widehat{p}_{r.OPT}) = \frac{V_AV_B - C_{AB}^2}{V_B} = \frac{V(\widehat{p}_A)V(\widehat{p}_B) - cov(\widehat{p}_A, \widehat{p}_B)^2}{V(\widehat{p}_B)}. \quad (10)$$

Replacing $V(\widehat{p}_A)$, $V(\widehat{p}_B)$ and $cov(\widehat{p}_A, \widehat{p}_B)$ in (10) by their respective expressions under SRSWOR we obtain

$$AV(\widehat{p}_{r.OPT}) = \frac{N-n}{N-1}\frac{1}{n}\left[\frac{P_AQ_AP_BQ_B - \phi^2P_AQ_AP_BQ_B}{P_BQ_B}\right] =$$

$$= \frac{N-n}{N-1}\frac{1}{n}P_AQ_A(1-\phi^2) = V(\widehat{p}_A)(1-\phi^2).$$

■

Theoretical comparison between the ratio estimator $\widehat{p}_{r.OPT}$ and the simple expansion estimator $\widehat{p}_A$ is fairly simple using theorem 2.4. In fact, $\widehat{p}_{r.OPT}$ is more efficient than $\widehat{p}_A$, since $1 - \phi^2 \leq 1$, and both estimators has the same performance when $\phi^2 = 0$.

Using theorem 2.4, an estimator of the optimum ratio type estimator variance is

$$\widehat{V}(\widehat{p}_{r.opt}) = \widehat{V}(\widehat{p}_A)(1 - \widehat{\phi}^2).$$

3. Simulation study

In this section, the proposed optimum ratio estimator $\widehat{p}_{r.opt}$ is compared numerically with alternative proportion estimators. Simulation studies are based on several simulated populations which cover a wide number of possible scenarios, including small and large proportions, small and large Cramer's V coefficients between the attribute of interest and the auxiliary attributes, etc. Simulated populations are briefly described as follows.

A total of 30 populations of $N = 1000$ units were generated to study the effect of different aspects on the estimators of a population proportion. Populations were generated as a random sample of 1000 units from a Bernoulli distribution with parameter $p = \{0.1, 0.25, 0.5, 0.75.0.9\}$, and the attributes of interest were thus achieved with the aforementioned population proportions. Auxiliary attributes were also generated by using the same distribution, but we randomly change a given proportion of values in order to the Cramer's V coefficient between the attribute of interest and the auxiliary attribute goes from 0.5 to 0.9. Since $P_A < P_B$ when $P_A = 0.25$, we also generated populations with $P_A = 0.25$ and $P_A > P_B$, which

For Peer Review Only

Table 1. Proportion (%) of number of times that estimators takes values outside the limits $[0, 1]$. Values associated to estimators $\widehat{p}_{r.e}$ and $\widehat{p}_{r.opt}$ are all 0%, and they are omitted.

		$n = 20$		$n = 60$		$n = 100$	
P_A	ϕ	$\widehat{p}_r$	$\widehat{p}_d$	$\widehat{p}_r$	$\widehat{p}_d$	$\widehat{p}_r$	$\widehat{p}_d$
0.1	0.5	0.00	12.31	0.00	0.22	0.00	0.20
	0.6	0.00	11.73	0.00	0.75	0.00	0.20
	0.7	0.00	3.88	0.00	0.15	0.00	0.06
	0.8	0.00	4.16	0.00	0.02	0.00	0.00
	0.9	0.00	0.52	0.00	0.00	0.00	0.00
0.9	0.5	16.64	12.31	4.15	0.22	1.30	0.20
	0.6	13.10	11.73	2.63	0.75	0.50	0.20
	0.7	6.91	3.88	0.90	0.15	0.12	0.06
	0.8	4.24	4.16	0.15	0.02	0.00	0.00
	0.9	1.83	0.52	0.00	0.00	0.00	0.00

allow us to study the effect of the relation between P_A and P_B on the various estimators, specially on the estimator $\widehat{p}_r$.

For each of the 30 populations, $D = 10000$ samples, with size $n = 100$, were selected to compare the various estimators in terms of relative bias (RB) and relative efficiency (RE), where

$$RB = \frac{E[\widehat{p}] - P_A}{P_A} \quad ; \quad RE = \frac{MSE[\widehat{p}_A]}{MSE[\widehat{p}]},$$

$\widehat{p}$ is a given estimator and the empirical expectation ($E[\cdot]$) and the empirical mean square error ($MSE[\cdot]$) are given by

$$E[\widehat{p}] = \frac{1}{D}\sum_{d=1}^{D}\widehat{p}(d) \quad ; \quad MSE[\widehat{p}] = \frac{1}{D}\sum_{d=1}^{D}(\widehat{p}(d) - P_A)^2,$$

$\widehat{p}(d)$ denotes the estimator $\widehat{p}$ calculated at the dth simulation run. Values of RE larger than 1 indicate that the estimator $\widehat{p}$ is more efficient than the customary estimator $\widehat{p}_A$, which is considered as the reference estimator in the efficiency studies.

We considered the proposed optimum ratio estimator $\widehat{p}_{r.opt}$, the ratio estimators $\widehat{p}_r$ and $\widehat{p}_{r.e}$ proposed by Rueda et al. (2011) and the difference estimator given by

$$\widehat{p}_d = \widehat{p}_A + (P_B - \widehat{p}_B).$$

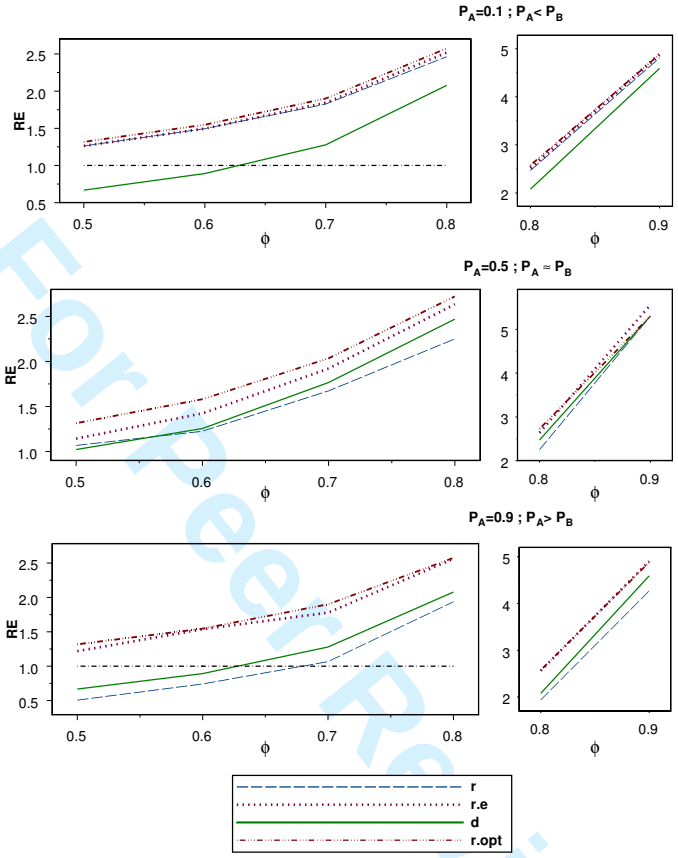
Values of RB in this simulation study are within a reasonable range, i.e, they are all less than 1% and are thus omitted. Figures 1 and 2 reports the values of RE for the various estimators and samples selected under SRSWOR. We observe that the proposed optimum estimator is more efficient than alternative estimators, whereas other estimators such as $\widehat{p}_r$ and $\widehat{p}_d$ can be less efficient than the customary estimator $\widehat{p}_A$. The gain in efficiency of the estimators based upon auxiliary information increases considerably as ϕ is closer to 0.9. For this reason and for clarity in figures, values of RE are separately plotted according to ϕ.

Table 1 reports the proportion of number of times that estimators take values outside the limits $[0, 1]$. For small as well as for large samples, we observe that the proposed optimum ratio estimator and $\widehat{p}_{r.e}$ always take values within the interval $[0, 1]$, whereas estimators $\widehat{p}_r$ and $\widehat{p}_d$ can give estimates outside $[0, 1]$, with a percentage larger than 10% for small samples.

For Peer Review Only

4. Figures

Figure 1. Values of Relative Efficiency (RE) for the various estimators of P_A. ϕ goes from 0.5 to 0.9 and P_A takes the values 0.1, 0.5 and 0.9.



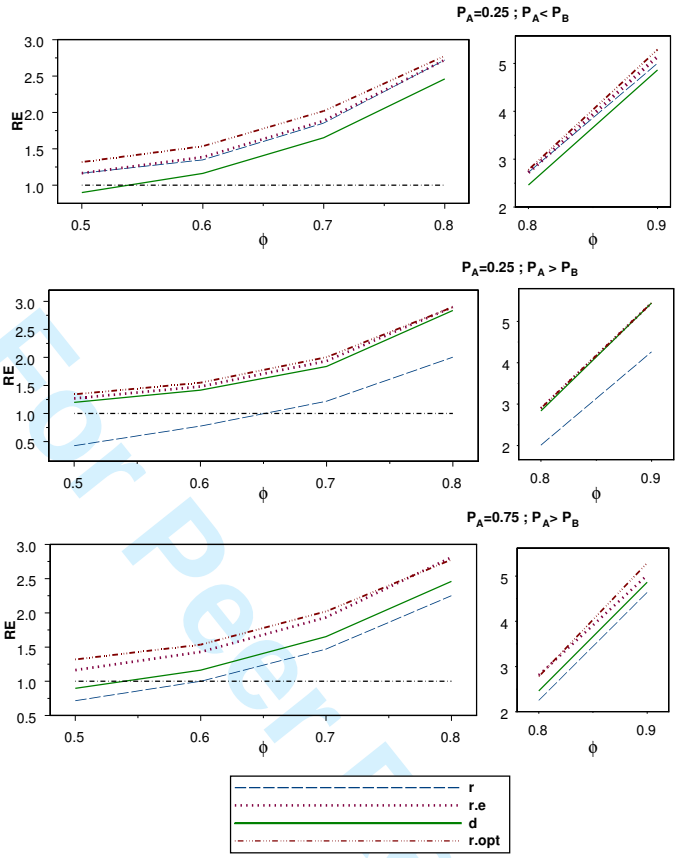
Acknowledgments. The authors would like to thank the referees for their many helpful comments and suggestions. This work is partially supported by Ministerio de Educación y Ciencia (contract No. MTM2009-10055)

References

[1] H. Cramer, *Mathematical methods of statistics*, Princenton, Pricenton University, 1946.
[2] S. Singh, *Advanced sampling theory with applications: How Michael "selected" Amy,*, The Netherlands, 2003.
[3] C.E. Särndal, B.Swensson and J. Wretman , *Model Assisted Survey sampling*, Springer Verlag, 1992.
[4] J.C. Deville, C.E. Särndal, *Calibration Estimators in Survey Sampling*, J. Am. Stat. Assoc. **87** (1992) 376-382.
[5] M.M. Rueda, J.F. Muñoz, A. Arcos, E. Álvarez, and S. Martínez, *Estimators and confidence intervals for the proportion using binary auxiliary information with applications to pharmaceutical studies* , Journal of Biopharmaceutical Statistics (2011). In press.
[6] C.E. Särndal, *The calibration approach in survey theory and practice*, Surv. Methodol. **33** (2007) 99-119.

For Peer Review Only

Figure 2. Values of Relative Efficiency (RE) for the various estimators of P_A. ϕ goes from 0.5 to 0.9 and P_A takes the values 0.25 and 0.75.



For Peer Review Only