A Multi-proportion Randomized Response Model Using the Inverse Sampling

Marcella Polisicchio and Francesco Porro

Abstract This paper describes a new procedure to unbiasedly estimate the proportions of t population groups, which at least one is very small and then it can be considered a rare group. This procedure guarantees the privacy protection of the interviewees, as it is based on an extension of the Warner randomized response model. As the estimation regards rare groups, the sampling design considered is the inverse sampling. Some characteristics of the proposed estimators are investigated.

Keywords Inverse sampling • Randomized response • Sensitive questions

1 Introduction

Sample surveys are fundamental to better understand the society we live in, and the reliability on the results is a crucial point of the whole process. The surveys regarding sensitive issues, stigmatizing attributes, or in general behaviors which are not accepted by the majority of the population, can sometimes embarrass the interviewees, and therefore they can be difficult to perform. When dealing with such personal information, the refusal to respond or even worse, intentional incorrect answers must be taken into account by the researcher, since it is evident that the non-sampling error induced can bear unreliable estimates (for instance, see Cochran 1963).

Warner (1965) introduced an ingenious method that overcomes this issue. He proposed a randomized response technique, which provides an unbiased estimator for the unknown proportion π_A of the persons bearing a stigmatizing characteristic A, with no privacy violation of the interviewee. There is no privacy violation in

---

M. Polisicchio • F. Porro (✉)

Università degli Studi di Milano-Bicocca, Milano, Italy

e-mail: marcella.polisicchio@unimib.it; francesco.porro1@unimib.it

the sense that the interviewer receives an answer ("yes" or "no") to the question "Is the statement on the card you drew, true for you?", but he does not know what the statement says. Even if one can believe that nowadays such technique has to be considered overcome, for example, by the computer-assisted self-interview, in the literature there are some studies which prove the contrary (see, for example, Van der Heijden et al. 2000; Van der Heijden and Bockenholt 2008). From the innovative paper of Warner (1965) a lot of improvements have been implemented and a lot of more refined techniques have been developed. Multiple randomized response devices (see Mangat and Singh 1990; Singh 2002; Gjestvang and Singh 2006), modified randomized response techniques (see Mangat 1994; Kuk 1990), models taking into account the probability of lying (see Mangat and Singh 1995), models with one or more scrambled variables (see Gupta et al. 2002; Gjestvang and Singh 2006) and models with unrelated questions (see Singh and Mathur 2006; Pal and Singh 2012)—just to report some examples—are all models or techniques originated by the intuition of Warner. In particular, an interesting extension to the case where the aim is the estimation of the proportions of t-related mutually exclusive population groups, with at least one and at most $t-1$ of which exhibit sensitive characteristics, is due to Abul-Ela et al. (1967).

The procedure suggested by Warner and the extension of Abul-Ela et al. are based on the "sample proportion" estimator, within a framework where the sample size is fixed a priori. The use of such estimator can have some drawbacks. The first is that, drawing an "unlucky" sample, it can happen that the sample proportion estimator assumes a value equal to zero: in this case the proportion estimate of the group bearing the sensitive attribute depends only on the parameter settings of randomized response device and does not depend on the empirical results. Another situation where the sample proportion estimator can have some troubles is when the sampling costs are expensive, since the drawing of a sample with no elements with the sensitive attribute implies a discard of the sample and a new drawing: this can cause a relevant increasing of the total cost of the survey. The last point is that if the investigated proportion is small (meaning that the survey deals with rare populations) the literature provides more performant estimators: one of these is based on the inverse sampling technique, introduced by Haldane (see Haldane 1945a,b). In such technique, the sample size is not fixed a priori, since the sampling continues until a certain quantity, say k, of elements with the considered attribute are drawn from the population. Many papers in the literature developed the inverse sampling (for further details, see Finney 1949; Best 1974; Mikulski and Smith 1976; Sathe 1977; Sahai 1980; Prasad and Sahai 1982; Pathak and Sathe 1984; Mangat and Singh 1991; Singh and Mathur 2002b; Chaudhuri 2011; Chaudhuri et al. 2011a, just to mention some references).

The aim of this paper is to use the randomized response technique and the inverse sampling together. More in detail, this paper proposes a method to estimate the t proportions of related mutually exclusive population groups, with at least one rare group. This method is based on the inverse sampling technique. Indeed in the literature the randomized response technique and the inverse sampling have been already used together, but as far as the authors know, only to approach the case

where one proportion has to be estimated (refer to Singh and Mathur 2002a for further details).

In order to simplify the explanation, the present paper will describe the case where $t = 3$, that is the trinomial case, since the generalization to the case $t > 3$ straightforwardly follows.

The plan of the paper is the following: the next section provides a brief explanation of the randomized response technique of Warner, and the extension of Abul-Ela et al. to the trinomial case. In Sect. 3 the new procedure based on the inverse sampling, for obtaining the proportion estimators is described, and three propositions about their features are stated and proved in detail. Section 4 provides an efficiency comparison with the estimators proposed by Abul-Ela et al. Section 5 is devoted to the estimation of the variances and the covariance of the proposed estimators. In order to improve the proportion estimates, three shrinkage estimators are proposed in Sect. 6. The last section describes some final remarks.

2 The Randomized Response Model of Warner and the Extension to the Trinomial Case of Abul-Ela et al.

Let $\mathbb{S}_1$ be a population, consisting of two mutually exclusive groups. The persons in the first group, say A, bear a sensitive attribute. The aim of the survey is the estimation of the proportion π_A of the group A. Since the direct question "Are you in group A?" can be embarrassing for the interviewee, Warner (1965) proposed a procedure equivalent to the following one. A deck of cards is given to each interviewee. On each card is reported only one of these two statements:

(1) Statement I: "I belong to the group A".
(2) Statement II: "I do not belong to the group A".

The proportion p (with $p \neq 1/2$) of cards with the statement I is known and fixed before the beginning of the survey by the researcher. Obviously, the proportion of cards with the statement II is then $1 - p$. The interviewee is requested to shuffle the deck, to draw a card, and to answer "yes" or "no" if the statement on the card is true or no for him/her, respectively. Since the interviewer does not see the card, he does not know the statement on it, and therefore the privacy of the respondent is not violated. Using this setting, the proportion π_A can be unbiasedly estimated by

$$\hat{\pi}_A = \frac{\hat{\lambda} - (1 - p)}{2p - 1},$$

where $\hat{\lambda}$ is the sample proportion estimator of "yes" answers.

As aforementioned, Abul-Ela et al. extended the Warner procedure, in order to estimate the proportions of t groups: in Abul-Ela et al. (1967) they described in detail the case $t = 3$. Here, a summary of such procedure is reported.

Let $\mathbb{S}_2$ be a population with three exhaustive and mutually exclusive groups (say A, B, and C). Let π_i $(i = 1, 2, 3)$ be the true unknown proportions of the three groups, that is:

$$\pi_1 = \text{the true proportion of group A in the population;}$$

$$\pi_2 = \text{the true proportion of group B in the population;}$$

$$\pi_3 = \text{the true proportion of group C in the population;}$$

with

$$\pi_i \in [0, 1], \quad \text{and} \quad \sum_{i=1}^{3} \pi_i = 1. \tag{1}$$

Since the independent parameters to be estimated are two, from the population two independent non-overlapping random samples with replacement are drawn: the first one with size m_1, the second one with size m_2.

Remark 1. It is worth underlining that here the two sample sizes m_1 and m_2 are fixed a priori by the researcher before the beginning of the survey.

The randomized devices are two decks of cards, one for each sample. There are three kinds of cards in each deck. For the deck i, $(i = 1, 2)$, let

p_{i1} be the proportion of cards with the statement: "I belong to group A";

p_{i2} be the proportion of cards with the statement: "I belong to group B";

p_{i3} be the proportion of cards with the statement: "I belong to group C".

Obviously it holds that

$$p_{ij} \in [0, 1], \quad \text{and} \quad \sum_{j=1}^{3} p_{ij} = 1, \ i = 1, 2. \tag{2}$$

Merely for a technical reason that will be evident in the following, the proportions p_{ij} of the cards in the two decks must satisfy the restriction:

$$(p_{11} - p_{13})(p_{22} - p_{23}) \neq (p_{12} - p_{13})(p_{21} - p_{23}). \tag{3}$$

Abul-Ela et al. (1967) proposed to consider the random variables X_{ir} $(i = 1, 2)$, defined by

$$X_{ir} = \begin{cases} 0 \text{ if the } r\text{-th interviewee in sample } i \text{ says "no"} \\ 1 \text{ if the } r\text{-th interviewee in sample } i \text{ says "yes"} \end{cases}$$

where $r \in \{1, 2, \ldots, m_1\}$ if $i = 1$, and $r \in \{1, 2, \ldots, m_2\}$ if $i = 2$. The probabilities of "yes" answer for the r-th interviewee in the two samples are

$$\mathbb{P}(X_{1r} = 1) = \mathbb{P}(X_{1r} = 1|A)\mathbb{P}(A) + \mathbb{P}(X_{1r} = 1|B)\mathbb{P}(B) + \mathbb{P}(X_{1r} = 1|C)\mathbb{P}(C)$$
$$= p_{11}\pi_1 + p_{12}\pi_2 + p_{13}\pi_3,$$

and

$$\mathbb{P}(X_{2r} = 1) = \mathbb{P}(X_{2r} = 1|A)\mathbb{P}(A) + \mathbb{P}(X_{2r} = 1|B)\mathbb{P}(B) + \mathbb{P}(X_{2r} = 1|C)\mathbb{P}(C)$$
$$= p_{21}\pi_1 + p_{22}\pi_2 + p_{23}\pi_3,$$

where $\mathbb{P}(X_{ir} = 1|A)$ denotes the probability that the r-th interviewee in the sample i ($i = 1, 2$) says "yes", given that he/she belongs to group A and $\mathbb{P}(A)$ is the probability that a person selected belongs to the group A. For the restrictions (2) and (1), the above probabilities become

$$\mathbb{P}(X_{1r} = 1) = (p_{11} - p_{13})\pi_1 + (p_{12} - p_{13})\pi_2 + p_{13},$$

and

$$\mathbb{P}(X_{2r} = 1) = (p_{21} - p_{23})\pi_1 + (p_{22} - p_{23})\pi_2 + p_{23}.$$

Denoting $\mathbb{P}(X_{1r} = 1)$ by λ_1 and $\mathbb{P}(X_{2r} = 1)$ by λ_2, from the equations

$$\begin{cases} \lambda_1 = (p_{11} - p_{13})\pi_1 + (p_{12} - p_{13})\pi_2 + p_{13} \\ \lambda_2 = (p_{21} - p_{23})\pi_1 + (p_{22} - p_{23})\pi_2 + p_{23} \\ \pi_3 = 1 - (\pi_1 + \pi_2) \end{cases} \quad (4)$$

it derives that

$$\begin{cases} \pi_1 = C \cdot [(\lambda_1 - p_{13})(p_{22} - p_{23}) - (\lambda_2 - p_{23})(p_{12} - p_{13})] \\ \pi_2 = -C \cdot [(\lambda_1 - p_{13})(p_{21} - p_{23}) - (\lambda_2 - p_{23})(p_{11} - p_{13})] \\ \pi_3 = 1 - C[(\lambda_1 - p_{13})(p_{22} - p_{21}) - (\lambda_2 - p_{23})(p_{12} - p_{11})], \end{cases} \quad (5)$$

where the constant C is given by

$$C = [(p_{11} - p_{13})(p_{22} - p_{23}) - (p_{12} - p_{13})(p_{21} - p_{23})]^{-1}. \quad (6)$$

Remark 2. The restriction (3) guarantees that the constant C is well defined.

The equations in (5) describe the unknown proportions π_1, π_2 and π_3 in terms of the quantities λ_1 and λ_2, which can be estimated using the sample proportion estimators. Then, denoting by Y_1 the number of "yes" answers in the first sample, and by Y_2 the number of "yes" answers in the second sample, the random variables

$$\hat{\lambda}_1^A = \frac{Y_1}{m_1} \qquad \text{and} \qquad \hat{\lambda}_2^A = \frac{Y_2}{m_2} \tag{7}$$

can be used to estimate the probabilities λ_1 and λ_2. Since the sampling is with replacement, the random variables Y_1 and Y_2 have binomial distribution, with parameters (λ_1, m_1) and (λ_2, m_2), respectively. Replacing these estimators in (5), Abul-Ela et al. obtained the following estimators for the proportions π_i ($i = 1, 2, 3$):

$$\begin{cases} \hat{\pi}_1^A = C \cdot \left[\left(\hat{\lambda}_1^A - p_{13} \right) (p_{22} - p_{23}) - \left(\hat{\lambda}_2^A - p_{23} \right) (p_{12} - p_{13}) \right] \\ \hat{\pi}_2^A = -C \cdot \left[\left(\hat{\lambda}_1^A - p_{13} \right) (p_{21} - p_{23}) - \left(\hat{\lambda}_2^A - p_{23} \right) (p_{11} - p_{13}) \right] \\ \hat{\pi}_3^A = 1 - C \left[\left(\hat{\lambda}_1^A - p_{13} \right) (p_{22} - p_{21}) - \left(\hat{\lambda}_2^A - p_{23} \right) (p_{12} - p_{11}) \right], \end{cases} \tag{8}$$

where the last equation is obtained from $\hat{\pi}_3^A = 1 - (\hat{\pi}_1^A + \hat{\pi}_2^A)$. Such estimators are unbiased and they have some interesting features (see Abul-Ela et al. 1967; Chaudhuri 2011 for further details). It is worth noting that the variances of the estimators in (8) can be directly evaluated by

$$\begin{aligned} \text{var}(\hat{\pi}_1^A) &= C^2 \left[(p_{22} - p_{23})^2 \varphi_1^2 + (p_{12} - p_{13})^2 \varphi_2^2 \right] \\ \text{var}(\hat{\pi}_2^A) &= C^2 \left[(p_{21} - p_{23})^2 \varphi_1^2 + (p_{11} - p_{13})^2 \varphi_2^2 \right] \\ \text{var}(\hat{\pi}_3^A) &= C^2 \left[(p_{22} - p_{21})^2 \varphi_1^2 + (p_{12} - p_{11})^2 \varphi_2^2 \right], \end{aligned} \tag{9}$$

where φ_i^2 denotes the variance of the sample proportion estimators $\hat{\lambda}_i^A$ defined in (7):

$$\varphi_i^2 = \text{var}(\hat{\lambda}_i^A) = \frac{\lambda_i (1 - \lambda_i)}{m_i}, \quad i = 1, 2. \tag{10}$$

3 The Introduction of the Inverse Sampling Technique

This section describes a procedure to estimate the proportions π_i ($i = 1, 2, 3$). As far as the authors know, there is no mention in the literature of such procedure, then it can be considered innovative.

As stated in the introduction, such procedure is based on the inverse sampling technique, where the sample size is not fixed a priori. The needed setting is the same one illustrated in the previous section: the population, the two decks with proportions of the cards, and the randomized response technique are exactly the same ones previously described.

The crucial difference from the procedure of Abul-Ela et al. is that, here, the sizes of the two samples are not fixed a priori. The drawing (with replacement) of

persons in the first sample continues until that k_1 "yes" answers are obtained, and the drawing (with replacement) of persons in the second sample finishes when k_2 "yes" answer are reported. The values of the parameters k_1 and k_2 are decided by the researcher before the beginning of the survey: the most interesting situations are when k_1 and k_2 are greater or equal to 2. In such cases, the sample sizes, say N_1 and N_2, are random variables with negative binomial distribution, with parameters (λ_1, k_1) and (λ_2, k_2), respectively. In the proposed procedure, only the sampling with replacement is examined. The case where only distinct units are considered poses further difficulties, and hence it will not be treated in this paper (for further details, see Chaudhuri 2011; Chaudhuri et al. 2011b).

The criterion for fixing the value of the parameter k_i $(i = 1, 2)$ depends on many aspects, for example, the population size, the sampling costs, the time and resources needed for the sampling procedure. For this reason, the determination of these parameters is an important issue and it varies from case to case. A possible criterion, using data from previous surveys (if available) is to select the value of the parameter k_i $(i = 1, 2)$ as the product between a reasonable (and cost-bearable) expected sample size and the previous estimate of λ_i $(i = 1, 2)$. In such way the sample size is random, but at least the expected sample size can be kept under control, since the expectation of the size for the sample i is given by k_i/λ_i $(i = 1, 2)$.

Using the notation of the previous section, in order to obtain estimators for the proportions π_i $(i = 1, 2, 3)$, two estimators of λ_1 and λ_2 are needed.

In the literature, when dealing with the inverse sampling scheme, the most used estimator for the proportion of a group in the population is given by the ratio between the prefixed number of "successes" that concludes the sampling minus one and the sample size minus one. This estimator is unbiased and its variance can be evaluated as the sum of a series. These and further important and useful features are discussed and proved in Haldane (1945a,b), Best (1974), and Chaudhuri (2011).

In the present framework dealing with the inverse sampling, the sample proportion estimators used by Abul-Ela et al. cannot be utilized. Instead of them, the following two estimators can be used:

$$\hat{\lambda}_1 = \frac{k_1 - 1}{N_1 - 1} \qquad \text{and} \qquad \hat{\lambda}_2 = \frac{k_2 - 1}{N_2 - 1}, \tag{11}$$

where k_1 and k_2 are the fixed numbers of "yes" answers which conclude the samplings and N_1 and N_2 are the random variables representing the sizes of the two samples. Following the same approach of Abul-Ela et al., the plug-in estimators of the unknown proportions π_i $(i = 1, 2, 3)$ can then be obtained as:

$$\begin{cases} \hat{\pi}_1 = C \cdot \left[\left(\hat{\lambda}_1 - p_{13}\right)(p_{22} - p_{23}) - \left(\hat{\lambda}_2 - p_{23}\right)(p_{12} - p_{13})\right] \\ \hat{\pi}_2 = -C \cdot \left[\left(\hat{\lambda}_1 - p_{13}\right)(p_{21} - p_{23}) - \left(\hat{\lambda}_2 - p_{23}\right)(p_{11} - p_{13})\right] \\ \hat{\pi}_3 = 1 - C \left[\left(\hat{\lambda}_1 - p_{13}\right)(p_{22} - p_{21}) - \left(\hat{\lambda}_2 - p_{23}\right)(p_{12} - p_{11})\right], \end{cases} \tag{12}$$

where the estimators $\hat{\lambda}_1$, $\hat{\lambda}_2$ are defined in (11), and the third estimator $\hat{\pi}_3$ is derived as $\hat{\pi}_3 = 1 - (\hat{\pi}_1 + \hat{\pi}_2)$.

These estimators have several interesting characteristics: some of them are stated in the following propositions.

Proposition 1. *The estimators defined in (12) are unbiased for the unknown proportions π_i ($i = 1, 2, 3$).*

Proof. For the first estimator $\hat{\pi}_1$ it holds that

$$
\begin{aligned}
\mathbb{E}(\hat{\pi}_1) &= \mathbb{E}\left[C\left[\left(\hat{\lambda}_1 - p_{13}\right)(p_{22} - p_{23}) - \left(\hat{\lambda}_2 - p_{23}\right)(p_{12} - p_{13}) \right] \right] \\
&= C\left[(p_{22} - p_{23})\left[\mathbb{E}\left(\hat{\lambda}_1\right) - p_{13} \right] - (p_{12} - p_{13})\left[\mathbb{E}\left(\hat{\lambda}_2\right) - p_{23} \right] \right].
\end{aligned}
$$

Since $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are unbiased (as proved in Haldane 1945b), it follows:

$$
\mathbb{E}(\hat{\pi}_1) = C\left[(p_{22} - p_{23})(\lambda_1 - p_{13}) - (p_{12} - p_{13})(\lambda_2 - p_{23}) \right],
$$

hence, using (4) and (6):

$$
\begin{aligned}
\mathbb{E}(\hat{\pi}_1) &= C\left[(p_{22} - p_{23})[(p_{11} - p_{13})\pi_1 + (p_{12} - p_{13})\pi_2] + \right. \\
&\quad \left. -(p_{12} - p_{13})[(p_{21} - p_{23})\pi_1 + (p_{22} - p_{23})\pi_2] \right] \\
&= \pi_1 C\left[(p_{22} - p_{23})(p_{11} - p_{13}) - (p_{12} - p_{13})(p_{21} - p_{23}) \right] \\
&= \pi_1.
\end{aligned}
$$

Through an analogous procedure it can be shown that also $\hat{\pi}_2$ is unbiased:

$$
\mathbb{E}(\hat{\pi}_2) = -C\left[(p_{21} - p_{23})\left[\mathbb{E}\left(\hat{\lambda}_1\right) - p_{13} \right] - (p_{11} - p_{13})\left[\mathbb{E}\left(\hat{\lambda}_2\right) - p_{23} \right] \right] = \pi_2,
$$

while for the last estimator $\hat{\pi}_3$ it holds:

$$
\mathbb{E}(\hat{\pi}_3) = \mathbb{E}[1 - (\hat{\pi}_1 + \hat{\pi}_1)] = 1 - \pi_1 - \pi_2 = \pi_3.
$$

Remark 3. By construction, the two estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are independent, as the first one is referred to the first sample, while the latter is referred to the second sample and the two samples are independent and non-overlapping. On the contrary, the two estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ are not independent since each of them depends on both $\hat{\lambda}_1$ and $\hat{\lambda}_2$.

Now, the variance of the estimators defined in (12) can be evaluated, as the following proposition states.

Proposition 2. *The variances of the estimators defined in (12) are:*

$$\text{var}(\hat{\pi}_1) = C^2 \left[(p_{22} - p_{23})^2 \, \delta_1^2 + (p_{12} - p_{13})^2 \, \delta_2^2 \right]$$
$$\text{var}(\hat{\pi}_2) = C^2 \left[(p_{21} - p_{23})^2 \, \delta_1^2 + (p_{11} - p_{13})^2 \, \delta_2^2 \right] \tag{13}$$
$$\text{var}(\hat{\pi}_3) = C^2 \left[(p_{22} - p_{21})^2 \delta_1^2 + (p_{12} - p_{11})^2 \delta_2^2 \right],$$

where δ_i^2 denotes the variance of the estimator $\hat{\lambda}_i$ defined in (11):

$$\delta_i^2 = \text{var}(\hat{\lambda}_i) = \lambda_i^2 \sum_{r=1}^{+\infty} \binom{k_i + r - 1}{r}^{-1} (1 - \lambda_i)^r, \quad i = 1, 2. \tag{14}$$

Proof. For the first estimator $\hat{\pi}_1$, from the definition (12) it follows that

$$\text{var}(\hat{\pi}_1) = C^2 \text{var} \left[\left(\hat{\lambda}_1 - p_{13} \right) (p_{22} - p_{23}) - \left(\hat{\lambda}_2 - p_{23} \right) (p_{12} - p_{13}) \right].$$

As stated in the Remark 3, since $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are independent, it holds that

$$\text{var}(\hat{\pi}_1) = C^2 \left[(p_{22} - p_{23})^2 \text{var} \left(\hat{\lambda}_1 \right) + (p_{12} - p_{13})^2 \text{var} \left(\hat{\lambda}_2 \right) \right]. \tag{15}$$

As briefly mentioned, Best (1974) proved that the variances δ_i^2 of the estimators $\hat{\lambda}_i$ defined in (11) are given by the formula (14), therefore replacing of such expressions in (15) concludes the proof for $\hat{\pi}_1$. With a similar procedure the variance of $\hat{\pi}_2$ and $\hat{\pi}_3$ can be evaluated as in (13).

In order to have a graphical overview on the behaviour of these variances when the parameters are modified, in Figs. 1 and 2 some graphs are drawn. In such graphs, the variance of $\hat{\pi}_1$ is considered, but the analogous remarks hold for $\hat{\pi}_2$. In all the figures, the settings of the randomized response devices (that is, the proportions of the cards in the two decks) are equal and given by

$$
\begin{array}{ll}
p_{11} = 0.1 & p_{21} = 0.2 \\
p_{12} = 0.2 & p_{22} = 0.5 \\
p_{13} = 0.7 & p_{23} = 0.3.
\end{array}
$$

Figure 1 shows how the variance of the estimator $\hat{\pi}_1$ changes as the values of the parameters k_1 (on the left) and k_2 (on the right) are modified. In both the graphs, the value of π_2 is fixed and equal to 0.3, while the value of π_1 is reported in the legenda, and the value of π_3 equals $1 - (\pi_1 + \pi_2)$. The difference between the two graphs is that in the one on the left, the value of k_2 is fixed and equal to 5, while in the other one k_1 is fixed and equal to 5.
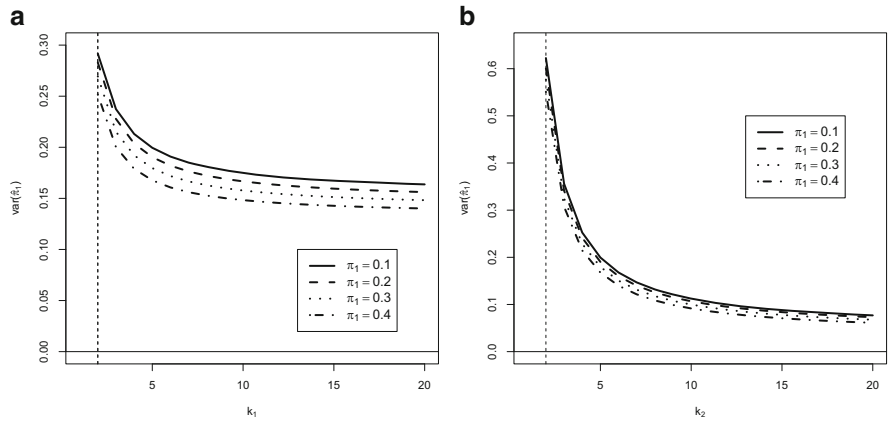
Fig. 1 The behaviour of the variance of $\hat{\pi}_1$, when k_1 (**a**) and when k_2 (**b**) varies, for different value of π_1
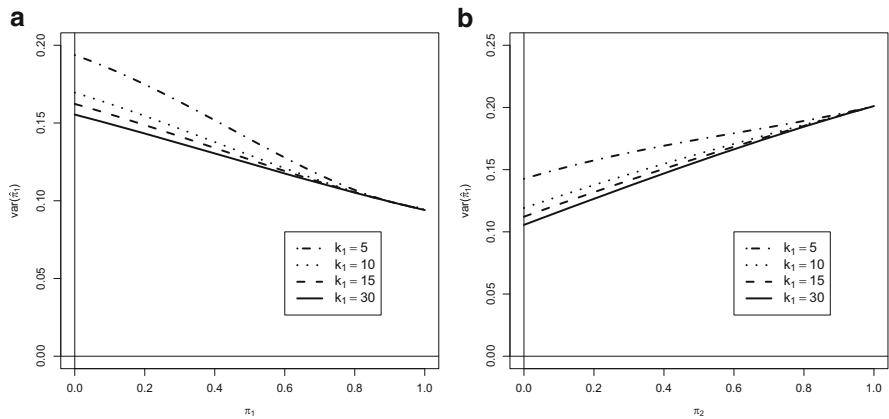

Fig. 2 The behavior of the variance of $\hat{\pi}_1$, when π_1 (**a**) and when π_2 (**b**) varies, for different value of k_1

The graph on the left in Fig. 1 shows that as k_1 increases, the variance of $\hat{\pi}_1$ decreases: such behaviour is not surprising, since a large k_1 means a large size of the first sample and then, at least intuitively, less variability of the estimators. The same arguments can be applied for the graph on the right side: here, moreover it is evident that not only $\hat{\pi}_2$, but also $\hat{\pi}_1$, as aspected, depends on the value of the parameter k_2. The different rate of decrease of var($\hat{\pi}_1$) in the two graphs can be explained by the different settings of the two decks of cards. Figure 2 shows how the variance of $\hat{\pi}_1$ depends on the true values of the proportions π_1 and π_2. In the graph on the left π_2 is set equal to 0.3, while in the other one the value of π_1 is 0.3.

Remark 4. The variances in the Proposition 2 are not easy to manage and for some values of the parameters, they are difficult to be evaluated, even only numerically. To overcome this issue, the finding of an upper bound for the variances may be useful, especially in the applications.

In Singh and Mathur (2005) eleven upper bounds for the variance of the estimators defined in (11) are compared: the main result is that the following upper bound proposed in Sahai (1983)

$$W_1(k) = \frac{\lambda}{6k} \left[\sqrt{A^2 - 12k\lambda B} - A \right]$$

where

λ is the unknown proportion of the "successes";

k is the number of "successes" that concludes the sampling procedure;

$$A = k^2 + k(3\lambda - 1) - 3\lambda(1 - \lambda) - \frac{6(1 - \lambda)^2}{k + 1};$$

$$B = (1 - \lambda) \left[\frac{(1 - \lambda)(k - 1)}{k + 1} - (k + 2) \right],$$

is the closest to the exact variance for smaller size for k (i.e. $k \leq 7$). For large k (i.e $k \geq 8$), the following upper bound, due to Pathak and Sathe (1984),

$$W_2(k) = \frac{\lambda^2(1 - \lambda)}{k} \left[1 + \frac{2(1 - \lambda)}{k - 2} \times \right.$$

$$\left. \times \left(1 - \frac{6\lambda}{k - 3(1 - \lambda) + 1 + \sqrt{(k - 5(1 - \lambda) + 1)^2 + 16\lambda(1 - \lambda)}} \right) \right]$$

is the best choice.

Applying such result to the estimators defined in (12), the three upper bounds U_i for their variances can be obtained, since:

$$\text{var}(\hat{\pi}_1) \leq U_1 = C^2 \left[(p_{22} - p_{23})^2 \, W(k_1) + (p_{12} - p_{13})^2 \, W(k_2) \right]$$

$$\text{var}(\hat{\pi}_2) \leq U_2 = C^2 \left[(p_{21} - p_{23})^2 \, W(k_1) + (p_{11} - p_{13})^2 \, W(k_2) \right] \qquad (16)$$

$$\text{var}(\hat{\pi}_3) \leq U_3 = C^2 \left[(p_{22} - p_{21})^2 \, W(k_1) + (p_{12} - p_{11})^2 \, W(k_2) \right],$$

where

$$W(k) = \begin{cases} W_1(k) & \text{for } k \leq 7 \\ W_2(k) & \text{for } k \geq 8. \end{cases}$$

As described in the Remark 3 the two estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ are not independent since each of them depends on both $\hat{\lambda}_1$ and $\hat{\lambda}_2$: for this reason the evaluation of the covariance between $\hat{\pi}_1$ and $\hat{\pi}_2$ needs an investigation.

Proposition 3. *The covariance between the estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ defined in (12) is*

$$\text{cov}(\hat{\pi}_1, \hat{\pi}_2) = C^2[(p_{22} - p_{23})(p_{23} - p_{21})\delta_1^2 + (p_{12} - p_{13})(p_{13} - p_{11})\delta_2^2].$$

Proof. In the previous proposition it has been proved that

$$\text{var}(\hat{\pi}_3) = \text{var}(\hat{\pi}_1 + \hat{\pi}_2) = C^2\left[(p_{22} - p_{21})^2\delta_1^2 + (p_{12} - p_{11})^2\delta_2^2\right]$$

where δ_i^2 denotes the variance of the estimator $\hat{\lambda}_i$ $(i = 1, 2)$, defined in (11). By the definition of the variance of the sum of two random variables, it holds that

$$C^2\left[(p_{22} - p_{21})^2\delta_1^2 + (p_{12} - p_{11})^2\delta_2^2\right] = \text{var}(\hat{\pi}_1) + \text{var}(\hat{\pi}_2) + 2\,\text{cov}(\hat{\pi}_1, \hat{\pi}_2).$$

Recalling from the previous proposition that

$$\text{var}(\hat{\pi}_1) = C^2\left[(p_{22} - p_{23})^2\,\delta_1^2 + (p_{12} - p_{13})^2\,\delta_2^2\right]$$

and

$$\text{var}(\hat{\pi}_2) = C^2\left[(p_{21} - p_{23})^2\,\delta_1^2 + (p_{11} - p_{13})^2\,\delta_2^2\right],$$

after some algebra it follows that

$$\text{cov}(\hat{\pi}_1, \hat{\pi}_2) = C^2[(p_{22} - p_{23})(p_{23} - p_{21})\delta_1^2 + (p_{12} - p_{13})(p_{13} - p_{11})\delta_2^2].$$

Remark 5. As before with the variances, the expression of the covariance can be difficult to manage. However, using the aforementioned upper bounds U_1 and U_2 and the Cauchy–Schwartz inequality, it is possible to determinate the following interval, where the covariance lies in:

$$\left[-\sqrt{U_1 U_2}\; ;\; \sqrt{U_1 U_2}\right].$$

The estimators $\hat{\pi}_i$ $(i = 1, 2, 3)$ proposed in this section have beyond the features stated in the previous propositions at least two other important characteristics. The first one is that the values of these estimators always depend on the sample information: this basic issue is not true for the estimators proposed by Abul-Ela et al., since if a sample contains no "yes" answers, the associated estimators sample proportion λ_i^A assumes value zero, and therefore the estimators $\hat{\pi}_i$ $(i = 1, 2, 3)$ capture no information from this sample. A possible solution to this drawback can be the discard of such "unlucky" sample, but if the sampling costs are not negligible, such operation can largely increase the total cost of the survey, and therefore in

particular situations cannot be performed. Using the procedure proposed in this paper, such situation cannot occur, since the estimators $\hat{\lambda}_i$ $(i = 1, 2)$ cannot be equal to zero.

The last remark is about the impression that the proposed estimators $\hat{\pi}_i$ $(i = 1, 2, 3)$ do not really differ so much from the ones introduced by Abul-Ela et al. From the formal point of view such idea can arise, as the expressions (12) look very similar to the ones in (8), but indeed the two procedures deal with two very different experiments: the one of Abul-Ela et al. is based on sample sizes prefixed, while the procedure proposed in this paper has random sample sizes.

4 An Efficiency Comparison

Since the proposed procedure provides three new estimators for the proportions π_i $(i = 1, 2, 3)$, a comparison with the ones proposed by Abul-Ela et al. (1967) should be interesting. All of these estimators are unbiased, therefore they can be compared, through an analysis of their variances. The expressions of $\text{var}(\hat{\pi}_i^A)$ $(i = 1, 2, 3)$ are reported in (9), while the $\text{var}(\hat{\pi}_i)$ $(i = 1, 2, 3)$ are described in (13).

This section compares the estimators $\hat{\pi}_1^A$ and $\hat{\pi}_1$, but the same investigation can be performed for the other estimators, obtaining analogous results.

Since the estimators $\hat{\pi}_1^A$ and $\hat{\pi}_1$ refer to two different sample designs, a reasonable comparison of their variance can be performed, assuming that the expected sample size is the same in the two sample designs.

Figure 3 shows on the left the variance of the estimator $\hat{\pi}_1^A$, and on the right, the variance of the estimator $\hat{\pi}_1$, in function of the true value of the proportion π_1.

In both the figures, the setting of the randomized response devices (i.e. the cards in the two decks) is the same one used for the figures in the previous section, and π_2 is equal to 0.3.

In the figure on the right, each curve corresponds to a particular choice of the parameters. The value of the parameter k_2 is fixed and equal to 6, while k_1 assumes three values: 3 (solid line), 10 (dashed line) and 100 (dotted line). As noted in the previous section, the variance curves are ordered: the curve corresponding to a large k_1 lies below the curve corresponding to a small k_1.

The variance of the estimator $\hat{\pi}_1^A$ of Abul-Ela et al. does not depend on k_1 and k_2, but it depends on the two sample sizes m_1 and m_2. In order to design three curves comparable to those of $\hat{\pi}_1$, the values of m_1 and m_2 have been set, applying the following rule:

$$m_1 = \frac{k_1}{\lambda_1} \qquad \text{and} \qquad m_2 = \frac{k_2}{\lambda_2}. \tag{17}$$

In other words, the values of m_1 in the first curve (in solid) on the left side are calculated as the ratio between $k_1 = 3$ and the value of λ_1, obtained using formula (4), which varies in function of π_1. In such way, as mentioned in the previous
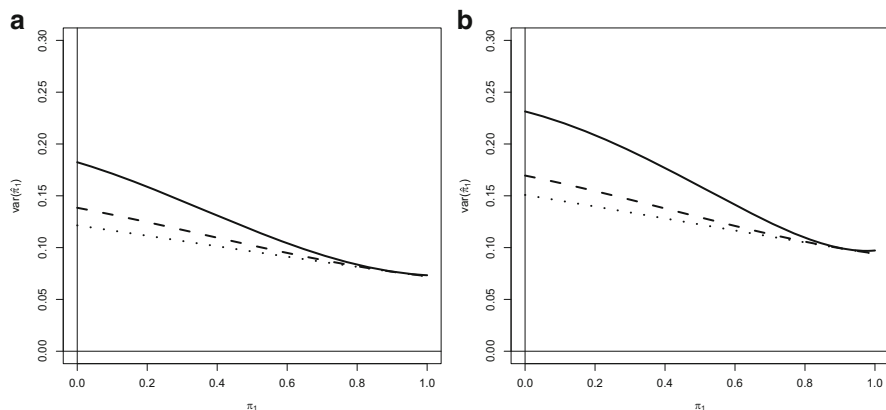
Fig. 3 The behaviour of the variance of $\hat{\pi}_1^A$ (**a**) and of $\hat{\pi}_1$ (**b**), for different values of the parameters, when π_1 varies

sections, for a fixed π_1, the parameters of two corresponding curves in the two figures guarantee that the expectations of the sample size in the two experiments are equal.

The comparison of the curves makes clear basically two points: first, the variance of $\hat{\pi}_1^A$ is always smaller than the variance of $\hat{\pi}_1$; second, the behaviour of the variance curves of $\hat{\pi}_1$ and $\hat{\pi}_1^A$, as the true proportion π_1 varies, is quite similar.

The first point is not surprising: the estimators proposed in this paper are based on the inverse sampling, which outperforms other sampling designs in particular situations, but it does not reduce the uncertainty and therefore the variability of the estimators.

Also the second point can be explained. The relationship between the variance of $\hat{\lambda}_i$ (denoted by δ_i^2) and the variance of $\hat{\lambda}_i^A$ (denoted by φ_i^2) for $(i = 1, 2)$ can be made explicit by

$$
\begin{aligned}
\delta_i^2 &= \text{var}(\hat{\lambda}_i) \\
&= \lambda_i^2 \sum_{r=1}^{+\infty} \binom{k_i + r - 1}{r}^{-1} (1 - \lambda_i)^r \\
&= \frac{\lambda_i^2 (1 - \lambda_i)}{k_i} + \lambda_i^2 \sum_{r=2}^{+\infty} \binom{k_i + r - 1}{r}^{-1} (1 - \lambda_i)^r \\
&= \text{var}(\hat{\lambda}_i^A) \frac{\lambda_i m_i}{k_i} + \lambda_i^2 \sum_{r=2}^{+\infty} \binom{k_i + r - 1}{r}^{-1} (1 - \lambda_i)^r \\
&= \varphi_i^2 \frac{\lambda_i m_i}{k_i} + \xi(k_i, \lambda_i).
\end{aligned}
\tag{18}
$$

Using the formula (18), the variance of $\hat{\pi}_1$ becomes:

$$\text{var}(\hat{\pi}_1) = C^2[(p_{22} - p_{23})^2 \delta_1^2 + (p_{12} - p_{13})^2 \delta_2^2]$$

$$= C^2 \left[(p_{22} - p_{23})^2 \left(\varphi_1^2 \frac{\lambda_1 m_1}{k_1} + \xi(k_1, \lambda_1) \right) + \right.$$

$$\left. + (p_{12} - p_{13})^2 \left(\varphi_2^2 \frac{\lambda_2 m_2}{k_2} + \xi(k_2, \lambda_2) \right) \right]$$

$$= C^2 \left[(p_{22} - p_{23})^2 \left(\varphi_1^2 \frac{\lambda_1 m_1}{k_1} \right) + (p_{12} - p_{13})^2 \left(\varphi_2^2 \frac{\lambda_2 m_2}{k_2} \right) \right] +$$

$$+ C^2 \left[(p_{22} - p_{23})^2 \xi(k_1, \lambda_1) + (p_{12} - p_{13})^2 \xi(k_2, \lambda_2) \right].$$

If the parameters m_1 and m_2 are chosen using the rule (17), then the var($\hat{\pi}_1$) assumes the form

$$\text{var}(\hat{\pi}_1) = \text{var}(\hat{\pi}_1^A) + C^2 \left[(p_{22} - p_{23})^2 \xi(k_1, \lambda_1) + (p_{12} - p_{13})^2 \xi(k_2, \lambda_2) \right].$$

It is worth noting that the quantity ξ introduced in (18) is the sum

$$\xi(k_i, \lambda_i) = \lambda_i^2 \sum_{r=2}^{+\infty} \binom{k_i + r - 1}{r}^{-1} (1 - \lambda_i)^r$$

$$= \frac{2! \lambda_i^2 (1 - \lambda_i)^2}{k_i (k_i + 1)} + \frac{3! \lambda_i^2 (1 - \lambda_i)^3}{k_i (k_i + 1)(k_i + 2)}$$

$$+ \frac{4! \lambda_i^2 (1 - \lambda_i)^4}{k_i (k_i + 1)(k_i + 2)(k_i + 3)} + \cdots$$

which decreases as k_i increases. This behaviour of ξ is the reason why in Fig. 3 the difference between var($\hat{\pi}_1$) and var($\hat{\pi}_1^A$) reduces, as k_1 increases.

The comparison highlights that the proposed estimators can be a valid alternative to those introduced by Abul-Ela et al., in sample surveys dealing with rare groups.

5 Unbiased Estimators for the Variances and the Covariance

In some cases, it is useful to have at disposal an estimator for the variances of $\hat{\pi}_i$ ($i = 1, 2, 3$). The following formula provides an unbiased estimator of the variance δ_i^2 ($i = 1, 2$) of the estimators defined in (11) (see Sukhatme et al. 1984 for further details):

$$\widehat{\text{var}(\hat{\lambda}_i)} = \widehat{\delta_i^2} = \frac{\hat{\lambda}_i(1 - \hat{\lambda}_i)}{N_i - 2}, \qquad \text{with } N_i > 2, \qquad i = 1, 2.$$

Using this result, the following estimators are therefore unbiased estimators for the variances $\text{var}(\hat{\pi}_i)$ $(i = 1, 2, 3)$:

$$\widehat{\text{var}(\hat{\pi}_1)} = C^2 \left[(p_{22} - p_{23})^2 \, \widehat{\delta_1^2} + (p_{12} - p_{13})^2 \, \widehat{\delta_2^2} \right]$$

$$\widehat{\text{var}(\hat{\pi}_2)} = C^2 \left[(p_{21} - p_{23})^2 \, \widehat{\delta_1^2} + (p_{11} - p_{13})^2 \, \widehat{\delta_2^2} \right] \qquad (19)$$

$$\widehat{\text{var}(\hat{\pi}_3)} = C^2 \left[(p_{22} - p_{21})^2 \widehat{\delta_1^2} + (p_{12} - p_{11})^2 \widehat{\delta_2^2} \right].$$

Analogously, an unbiased estimator for the $\text{cov}(\hat{\pi}_1, \hat{\pi}_2)$ is instead given by:

$$\widehat{\text{cov}(\hat{\pi}_1, \hat{\pi}_2)} = C^2 [(p_{22} - p_{23})(p_{23} - p_{21})\widehat{\delta_1^2} + (p_{12} - p_{13})(p_{13} - p_{11})\widehat{\delta_2^2}]. \qquad (20)$$

A detailed analysis of the features of the estimators defined in (19) and (20) is an interesting and promising subject, but it is out of the scope of the present paper.

6 A Shrinkage Estimator

This section provides three shrinkage estimators for the proportions π_1, π_2 and π_3 described in the previous sections. The basic idea is to reduce the support of the estimators defined in (12), discarding the inadmissible values. Since the considered estimators are estimators of proportions, the values outside of the interval $[0, 1]$ have to be removed. A drawback of such kind of shrinkage is that it transforms unbiased estimators in estimators that can be biased: the balance between such defect and the advantage deriving by the modification of the range has to be evaluated by the researcher, taking into account the final scope of the survey.

The shrinkage method presented in this section provides three new estimators $\hat{\pi}_1^S, \hat{\pi}_2^S$ and $\hat{\pi}_3^S$ such that:

- The range of each $\hat{\pi}_i^S$ is the interval $[0, 1]$.
- They sum to 1: $\sum_{i=1}^{3} \hat{\pi}_i^S = 1$.

To simplify the notation, according to the definitions in (12) the following five sets are defined:

$$L_1 = \{(\hat{\lambda}_1, \hat{\lambda}_2) \in]0, 1]^2 : \hat{\pi}_1(\hat{\lambda}_1, \hat{\lambda}_2) \leq 0\};$$

$$L_2 = \{(\hat{\lambda}_1, \hat{\lambda}_2) \in]0, 1]^2 : \hat{\pi}_2(\hat{\lambda}_1, \hat{\lambda}_2) \leq 0\};$$

$$Z_1 = \{(\hat{\lambda}_1, \hat{\lambda}_2) \in]0, 1]^2 : \hat{\pi}_1(\hat{\lambda}_1, \hat{\lambda}_2) \geq 1\};$$

$$Z_2 = \{(\hat{\lambda}_1, \hat{\lambda}_2) \in]0, 1]^2 : \hat{\pi}_2(\hat{\lambda}_1, \hat{\lambda}_2) \geq 1\};$$

$$R = \{(\hat{\lambda}_1, \hat{\lambda}_2) \in]0, 1]^2 : \hat{\pi}_1(\hat{\lambda}_1, \hat{\lambda}_2) + \hat{\pi}_2(\hat{\lambda}_1, \hat{\lambda}_2) > 1\}.$$

As in the previous sections, the value of the last estimator $\hat{\pi}_3$ is obtained at the end, using the values assumed by the others.

The shrinkage procedure is based on the assumption that the first group is the most relevant or significant, therefore the corresponding proportion estimator $\hat{\pi}_1$ has a more important role than the others. For this reason, as importance as possible is given to the empirical evidence regarding the first unknown proportion π_1 of group A. This is realized by the shrinkage, which does not modify the value of $\hat{\pi}_1$ whether it is admissible, hence:

$$\hat{\pi}_1^S(\hat{\lambda}_1, \hat{\lambda}_2) = \begin{cases} 0 & \text{for } (\hat{\lambda}_1, \hat{\lambda}_2) \in L_1 \\ 1 & \text{for } (\hat{\lambda}_1, \hat{\lambda}_2) \in Z_1 \\ \hat{\pi}_1(\hat{\lambda}_1, \hat{\lambda}_2) & \text{otherwise.} \end{cases}$$

The value of $\hat{\pi}_2$ is not modified in only one case, otherwise it changes as indicated below:

$$\hat{\pi}_2^S(\hat{\lambda}_1, \hat{\lambda}_2) = \begin{cases} 1 & \text{for } (\hat{\lambda}_1, \hat{\lambda}_2) \in L_1 \cap Z_2 \\ \hat{\pi}_2(\hat{\lambda}_1, \hat{\lambda}_2) & \text{for } (\hat{\lambda}_1, \hat{\lambda}_2) \in (R \cup L_2 \cup Z_2)^C \\ 1 - \hat{\pi}_1^S(\hat{\lambda}_1, \hat{\lambda}_2) & \text{for } (\hat{\lambda}_1, \hat{\lambda}_2) \in (R^C \cup L_1 \cup Z_1)^C \\ 0 & \text{otherwise.} \end{cases}$$
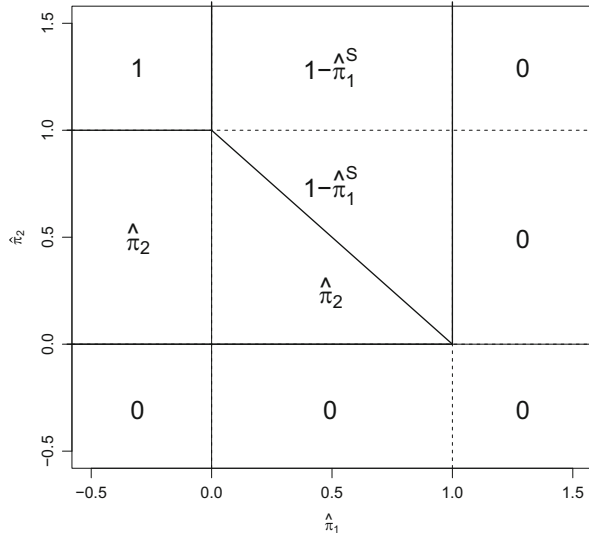
Finally, the estimator $\hat{\pi}_3$ follows:

$$\hat{\pi}_3^S = 1 - [\hat{\pi}_1^S + \hat{\pi}_2^S].$$

This shrinkage can be applied when the estimation of the proportion π_1 of the first group is more important and more sensible than the others. It is hence crucial to utilize in the estimation procedure all the information conveyed in the two samples about it. Figure 4 shows the values of the shrinkage estimator $\hat{\pi}_2^S$, as function of the values assumed by $\hat{\pi}_1$ and $\hat{\pi}_2$.

Remark 6. In the case where the most important proportion is related to the second group (group B), a different shrinkage procedure can be applied, giving more importance to the empirical result for the unknown proportion π_2 of group B. This new procedure can be performed, by the replacement of $\hat{\pi}_1$ with $\hat{\pi}_2$. Therefore, if needed, the first estimator to be shrunk is $\hat{\pi}_2$:

Fig. 4 The values of $\hat{\pi}_2^S$ for different values of the couple $(\hat{\pi}_1, \hat{\pi}_2)$

$$\hat{\pi}_2^S(\hat{\lambda}_1, \hat{\lambda}_2) = \begin{cases} 0 & \text{for} (\hat{\lambda}_1, \hat{\lambda}_2) \in L_2 \\ 1 & \text{for} (\hat{\lambda}_1, \hat{\lambda}_2) \in Z_2 \\ \hat{\pi}_2(\hat{\lambda}_1, \hat{\lambda}_2) & \text{otherwise.} \end{cases}$$

After that, just in case, the value of $\hat{\pi}_1$ changes as follows:

$$\hat{\pi}_1^S(\hat{\lambda}_1, \hat{\lambda}_2) = \begin{cases} 1 & \text{for} (\hat{\lambda}_1, \hat{\lambda}_2) \in L_2 \cap Z_1 \\ \hat{\pi}_1(\hat{\lambda}_1, \hat{\lambda}_2) & \text{for} (\hat{\lambda}_1, \hat{\lambda}_2) \in (R \cup L_1 \cup Z_1)^C \\ 1 - \hat{\pi}_2^S(\hat{\lambda}_1, \hat{\lambda}_2) & \text{for} (\hat{\lambda}_1, \hat{\lambda}_2) \in (R^C \cup L_2 \cup Z_2)^C \\ 0 & \text{otherwise.} \end{cases}$$

As before the value of $\hat{\pi}_3$ is then computed by:

$$\hat{\pi}_3^S = 1 - [\hat{\pi}_1^S + \hat{\pi}_2^S].$$

7 Conclusions

This paper describes a new procedure to obtain the estimators of the proportions of t population groups, which at least one is rare. As mentioned in the introduction, the description proposed in Sect. 2 deals with the trinomial case, but the extension to the case of more than three groups straightforwardly follows. The characteristics stated and proved in the previous sections highlight the potential of the estimators proposed and confirm that they can be successfully utilized in particular situations.

Indeed, this has to be considered a preliminary work, because further investigations are needed. A detailed comparison with other estimators and the comparison with other randomized response models are two examples of possible directions that must be explored in order to complete the analysis of these estimators, which look like very interesting and very useful in real situations.

References

Abul-Ela, A.L.A., Greenberg, B.G., Horvitz, D.G.: A multi-proportions randomized response model. J. Am. Stat. Assoc. **62**, 990–1008 (1967)

Best, D.J.: The variance of the inverse binomial estimator. Biometrika **61**, 385–386 (1974)

Cochran, W.G.: Sampling Techniques. Wiley, New York (1963)

Chaudhuri, A.: Randomized Response and Indirect Questioning Techniques in Surveys. CRC Press, Boca Raton (2011)

Chaudhuri, A., Bose, M., Dihidar, K.: Estimation of a sensitive proportion by Warners's randomized response data through inverse sampling. Stat. Pap. **52**, 343–354 (2011)

Chaudhuri, A., Bose, M., Dihidar, K.: Estimation sensitive proportions by Warners's randomized techinique using multiple randomized responses for distinct persons sampled. Stat. Pap. **52**, 111–124 (2011)

Finney, D.J.: On a method of estimating frequencies. Biometrika **36**, 233–234 (1949)

Gjestvang, C.R., Singh, S.: A new randomized response model. J. R. Stat. Soc. Ser. B **68**, 523–530 (2006)

Gupta, S., Gupta, B., Singh, S.: Estimation of sensitivity level of personal interview survey questions. J. Stat. Plan. Inference **100**, 239–247 (2002)

Haldane, J.B.S.: A labour-saving method of sampling. Nature (Lond.) **155**, 49–50 (1945)

Haldane, J.B.S.: On a method of estimating frequencies. Biometrika **33**, 222–225 (1945)

Kuk, A.Y.C.: Asking sensitive questions indirectly. Biometrika **77**, 436–438 (1990)

Mangat, N.S.: An improved randomized response strategy. J. R. Stat. Soc. Ser. B **56**, 93–95 (1994)

Mangat, N.S., Singh, R.: An alternative randomized response procedure. Biometrika **77**, 439–442 (1990)

Mangat, N.S., Singh, R.: Alternative approach to randomized response surveys. Statistica **51**, 327–332 (1991)

Mangat, N.S., Singh, R.: A note on the inverse binomial randomized response procedure. J. Indian Soc. Agric. Stat. **47**, 21–25 (1995)

Mikulski, P.W., Smith, P.J.: A variance bound for unbiased estimation in inverse sampling. Biometrika **63**, 216–217 (1976)

Pal, S., Singh, S.: A new unrelated question randomized response model. Statistics **46**, 99–109 (2012)

Pathak, P.K., Sathe, Y.S.: A new variance formula for unbiased estimation in inverse sampling. Sankhya Indian J. Stat. Ser. B **46**, 301–305 (1984)

Prasad, G., Sahai, S.: Sharper variance upper bound for unbiased estimation in inverse sampling. Biometrika **69**, 286 (1982)

Sahai, A.: Improved variance bounds for unbiased estimation in inverse sampling. J. Stat. Plan. Inference **4**, 213–216 (1980)

Sahai, A.: On a systematic sharpening of variance bounds of MVU estimator in inverse binomial sampling. Statistica **43**, 621–624 (1983)

Sathe, Y.S.: Sharper variance bounds for unbiased estimation in inverse sampling. Biometrika **64**, 425–426 (1977)

Singh, S.: A new stochastic randomized response model. Metrika **56**, 131–142 (2002)

Singh, H.P., Mathur, N.: An alternative randomized response technique using inverse sampling. Calcutta Stat. Assoc. Bull. **53**, 211–212 (2002)

Singh, H.P., Mathur, N.: On Mangat's improved randomized response strategy. Statistica **62**, 397–403 (2002)

Singh, H.P., Mathur, N.: On inverse binomial randomized response technique. J. Indian Soc. Agric. Stat. **59**, 192–198 (2005)

Singh, H.P., Mathur, N.: An improved estimation procedure for estimating the proportion of a population possessing sensitive attribute in unrelated question randomized response technique. Brasilian J. Probab. Stat. **20**, 93–110 (2006)

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C.: Sampling Theory of Surveys with Applications, 3rd edn. Iowa State University Press, Ames, Iowa (USA) and the Indian Society of Agricultural Statistics, New Delhi, India (1984)

Van der Heijden, P.G.M., Bockenholt, U.: Applications of randomized response methodology in e-commerce. In: Jank, W., Shmueli, G. (eds.) Statistical Methods in e-Commerce Research, pp. 401–416. Wiley, New York (2008)

Van der Heijden, P.G.M., Van Gils, G., Bouts, J., Hox, J.J.: A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. Sociol. Methods Res. **28**, 505–537 (2000)

Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc. **60**, 63–69 (1965)