

Surveys with Negative Questions for Sensitive Items

Fernando Esponda^{a,b}, Victor M. Guerrero^c

^a*Current: Instituto Tecnológico Autónomo de México (ITAM), Computer Science Department, México D.F. 01080, México*

^b*Previous: Yale University, Computer Science Department, New Haven CT 06520, USA*

^c*Instituto Tecnológico Autónomo de México (ITAM), Statistics Department, México D.F. 01080, México*

Abstract

This paper proposes a strategy for administering a survey that is mindful of sensitive data and individual privacy. The survey seeks to estimate the population proportion of a sensitive variable and does not depend on anonymity, cryptography, or in legal guarantees for its privacy preserving properties. Our technique presents interviewees with a question and t possible answers, and asks participants to eliminate one of $t - 1$ alternatives at random. We introduce a specific setup that requires just a single coin as randomizing device, and that limits the amount of information each respondent is exposed to by presenting to her/him only a subset of the question's alternatives. Finally we conduct a simulation study to provide evidence of the robustness against response and non-response bias of the suggested procedure.

Key words: Privacy preserving surveys; Estimating population proportions; Randomized response techniques; Sensitive attribute; Polychotomous variable

1. Introduction

In this paper, we propose to ask *negative questions* as a technique for conducting surveys that is mindful of participant's privacy. *Negative question surveys* allow participants to keep the target datum undisclosed by asking them,

Email address: fernando.esponda@itam.mx, Phone: +52 (55) 5628 4060 (Fernando Esponda)

To appear in Statistics & Probability Letters, ISSN: 0167-7152, Elsevier

instead, to make a series of decisions with the datum in mind. In this way, the frequency of drug use can be calculated without respondents admitting to using any drugs, the popular opinion on abortion can be measured without asking for anyone’s specific position, and an election can be run without any of the voters explicitly stating their preference. Our objective is to create a reliable technique for sensitive subject studies that is resilient to the usual types of bias and which elicits participation by providing transparent privacy guarantees. A suit of techniques that share the same motivation as ours—to protect privacy, promote participation, and increase survey accuracy—and which have a similar procedures are known as *Randomized response techniques* (RRTs) see Warner (1965); Horvitz et al. (1967); Abul-Ela et al. (1967); Bourke and Dalenius (1973); Chaudhuri and Mukerjee (1988) for its foundations or Kim and Warde (2005); Gjestvang and Singh (2006) for more recent work.

In what follows, we present negative question technique, discuss its privacy and explain how the sought data is extracted from it. We analyze, through an agent-based simulation, the impact of different kinds of biases on survey accuracy and conclude by summarizing the main characteristics of our method.

2. Negative question surveys

The type of survey we consider consists of a questionnaire with a single question (or statement) and t categories from which to choose an answer (or alternative). The survey is administered to a sample of n individuals drawn uniformly at random with replacement from the population.

We refer to it as a *positive question survey* or as a direct response survey when the subjects are asked to reveal which category they belong to. We call it a *negative question survey* when the requirement is to disclose a category (a single one, for the current work) to which they *do not* belong. A negative questionnaire can be obtained by simply negating the question of a positive questionnaire; however, the precise manner in which a negative question is posed must be considered carefully in order to maximize survey accuracy and is beyond the

scope of this work. The categories in the direct response survey are exhaustive and mutually exclusive, so that one and only one option is true; in the negative question survey, one and only one category is false for a particular individual; the object of both is to estimate the proportions of the population that belong to each category. Consider the two versions of the following survey (only one of the two questions is presented to the respondent):

(*Positive question*) I earn: (*Negative question*) I **do not** earn:

☐ Less than 30,000 dollars a year

☐ Between 30,000 and 60,000 dollars a year

☐ More than 60,000 dollars a year

An individual whose income is 20,000 dollars should select the first option when answering the positive question or either the second or third option when confronted with its negative version.

It is intuitively clear that the amount of information disclosed while answering a negative questionnaire is less than or equal to what is relinquished during the direct response version. We formalize this notion by quantifying the information surrendered to an observer, e.g., the interviewer, when only one option, s , is selected by a respondent in a negative questionnaire. Using Shannon's uncertainty measure, we compute this quantity as the difference in information of two positive questionnaires: the information from the positive version (first term of the following equation) minus the information from the same questionnaire once s is no longer an option (second term of the equation)

$$I(1, \dots, t | X \neq s) = - \sum_{i=1}^t p_i \log p_i + \sum_{i \neq s} P(X = i | X \neq s) \log P(X = i | X \neq s),$$

where $p_i = P(X = i)$ is the probability that option i is true in a direct response survey and $P(X = i | X \neq s)$ is the probability i is true after s has been removed as an option, i.e., after finding out it is false. These probabilities reflect the observer's prior and posterior beliefs regarding the respondent.

It is easy to see from the above expression that the information given away in a negative questionnaire is at most what is surrendered in its positive counterpart, that is $I(1, \dots, t | X \neq s) \leq -\sum_{1 \leq i \leq t} p_i \log p_i$. This fact is interpreted as saying that a negative question survey increases the interviewee's privacy with respect to its positive counterpart.

2.1. Estimating proportions using negative input

We now show how to estimate the proportions of the population that positively belong to each of the t categories. The analysis follows a similar reasoning as the one used in Chaudhuri and Mukerjee (1988) for randomized response techniques.

Let the random variables X and Y , each taking the values $1, \dots, t$, denote the true and reported categories, and let $\pi_i = P(X = i)$ be the proportion of the population that positively belongs to category i (with $\sum \pi_i = 1$). Now, let us consider n independent repetitions of an experiment in which only one of the t mutually exclusive events $Y = 1, \dots, Y = t$ occurs, with $\lambda_i = P(Y = i)$. Let us also assume the probabilities $\lambda_1, \dots, \lambda_t$ (with $\sum \lambda_i = 1$) remain constant for each experiment, then the joint distribution of the random variables n_1, \dots, n_t , representing the number of occurrences of the events $Y = 1, \dots, Y = t$ (with $\sum n_j = n$) is Multinomial with parameters n and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_t)'$. Therefore, for $i = 1, \dots, t$

$$E(n_i) = n\lambda_i, \quad Var(n_i) = n\lambda_i(1 - \lambda_i) \text{ and } Cov(n_i, n_j) = -n\lambda_i\lambda_j \text{ if } i \neq j. \quad (1)$$

We now define the conditional probabilities

$$p_{ij} = P(Y = i | X = j) \text{ if } i \neq j \text{ and } p_{ii} = 0, \quad (2)$$

so that

$$\lambda_i = \sum_{j=1}^t P(Y = i | X = j) \pi_j = \sum_{j=1}^t p_{ij} \pi_j \text{ for } i = 1, \dots, t \quad (3)$$

and, in matrix notation

$$\boldsymbol{\lambda} = P\boldsymbol{\pi}, \quad (4)$$

where P is a $(t \times t)$ -dimensional design matrix with $(i, j)^{th}$ element given by p_{ij} and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_t)'$. It is convenient to recall that the maximum likelihood estimator (mle) of λ_i , for $i = 1, \dots, t$, is given by $\hat{\lambda}_i = \frac{n_i}{n}$ with

$$E(\hat{\lambda}_i) = \lambda_i, \quad Var(\hat{\lambda}_i) = \frac{1}{n}\lambda_i(1 - \lambda_i) \text{ and } Cov(\hat{\lambda}_i, \hat{\lambda}_j) = -\frac{1}{n}\lambda_i\lambda_j. \quad (5)$$

The following result presents the estimation formulas to be employed here, it follows from eq. 5 and is basically the same result appearing in Chaudhuri and Mukerjee (1988), p. 37, thus a proof can be found there.

Proposition 1. *If P is nonsingular, an unbiased estimator of $\boldsymbol{\pi}$ and its corresponding variance are given by*

$$\hat{\boldsymbol{\pi}} = P^{-1}\hat{\boldsymbol{\lambda}} \text{ and } Var(\hat{\boldsymbol{\pi}}) = \frac{1}{n}P^{-1}[Diag(\boldsymbol{\lambda}) - \boldsymbol{\lambda}\boldsymbol{\lambda}']P'^{-1} \quad (6)$$

where $\hat{\boldsymbol{\lambda}} = \frac{1}{n}(n_1, \dots, n_t)'$ and $Diag(\boldsymbol{\lambda})$ is a diagonal matrix with elements $\lambda_1, \dots, \lambda_t$.

The validity of the estimates $\hat{\pi}_1, \dots, \hat{\pi}_t$ depends on an adequate sampling to produce reliable values of n_1, \dots, n_t , as would also be the case in a positive survey, and on having a good choice of the p_{ij} 's.

Knowing how individuals choose an option is the extra information needed to estimate the desired proportions while keeping personal data concealed. One possibility is to instruct respondents on how to choose among available options. For instance a simple, straightforward design is to give each category an equal chance of being selected, that is, for $i = 1, \dots, t$,

$$p_{ij} = \frac{1}{t-1} \text{ if } j \neq i \text{ and } p_{ii} = 0. \quad (7)$$

In this scenario the probability of choosing option i is

$$\lambda_i = \frac{1}{t-1}(1 - \pi_i), \quad (8)$$

which follows from eq. 3 and noting that $\sum_{j \neq i}^t \pi_j = 1 - \pi_i$. Therefore, using equations 8 and 5 we establish the following result.

Proposition 2. *Let us assume that the equal-chance design implied by expression 7 is valid. Then an unbiased estimator of the population proportion π_i is given by*

$$\hat{\pi}_i = 1 - (t-1)\hat{\lambda}_i, \text{ with } \hat{\lambda}_i = \frac{n_i}{n} \text{ for } i = 1, \dots, t. \quad (9)$$

The variance of $\hat{\pi}_i$ and the covariance of $\hat{\pi}_i$ and $\hat{\pi}_j$ for $j \neq i$ are

$$\text{Var}(\hat{\pi}_i) = \frac{(t-1)^2}{n} \lambda_i(1-\lambda_i) \text{ and } \text{Cov}(\hat{\pi}_i, \hat{\pi}_j) = -\frac{(t-1)^2}{n} \lambda_i \lambda_j. \quad (10)$$

In particular, from eqs. 8 and 10 it follows that

$$\text{Var}(\hat{\pi}_i) = \frac{\pi_i(1-\pi_i)}{n} \left(1 + \frac{t-2}{\pi_i} \right), \quad (11)$$

where the first factor is the variance of the mle of π_i based on a direct survey and the second factor is the variance inflation due to randomization. Thus, the number of options should be kept as low as possible in order not to over inflate the variance of $\hat{\pi}_i$. Note that designs in which a respondent randomly eliminates more than one option allow incrementing the total number of categories without excessively inflating the variance. The details and analysis of such designs, while not significantly different from the present one, are left for future work.

Finally, we stress that we focus on one particular proportion, corresponding to the sensitive characteristic under study. Thus, to make inference from the sample results we rely on the fact that any random variable n_i is marginally Binomially distributed with parameters n and λ_i , for $i = 1, \dots, t$. Thus, a confidence interval for λ_i can be constructed with the following adjusted Wald expression suggested by Agresti and Coull (1998)

$$\tilde{p}_i \pm z_{\alpha/2} \sqrt{\tilde{p}_i(1-\tilde{p}_i)/\tilde{n}}, \quad (12)$$

where $\tilde{p}_i = \tilde{n}_i/\tilde{n}$, $\tilde{n}_i = n_i + z_{\alpha/2}^2/2$, and $\tilde{n} = n + z_{\alpha/2}^2$, with $z_{\alpha/2}$ the upper $\alpha/2$ -percentage point of the $N(0,1)$ distribution. Hence, a $100(1-\alpha)\%$ confidence interval for π_i is obtained by multiplying eq. 12 by $(t-1)$ and subtracting these bounds from unity.

With the equal chance setup, respondents provided with a fair $(t-1)$ -sided die can select an answer by privately obtaining a value m and choosing, for

instance, the m^{th} true option from the top, skipping over the false category if needed. One difficulty of the approach is, in fact, having a $(t - 1)$ -sided die readily available, as during a phone survey. This is compounded when asking several questions, each with a different number of categories. This is addressed in the next section.

2.2. Two-option survey scheme

Direct response surveys, as well as RRTs, require presenting all categories to the interviewee—for only one choice is true. Conversely, in a negative question survey all options except one are true; consequently, only a subset needs to be evaluated by the respondent. With this in mind, we can preselect, uniformly at random, a subset of the categories (two or more) for each of the individuals questioned. Therefore, the questionnaire is not necessarily the same for all respondents. This fact allows us to simplify noticeably the second part of the randomization technique (the one to be carried out by the interviewee) while assigning greater responsibility to the survey administrator. Moreover, this feature of the proposed two-option scheme resembles the paired-alternative method of RRTs (see Horvitz et al. (1967); Chaudhuri and Mukerjee (1988)) and does away with using a $(t - 1)$ -sided die.

Consider the case where each subject is presented with a question and two options; if both options prove true, one should be selected at random—the interviewee privately tosses a fair coin prior to reading the question and picks the first category if heads, or the second if tails—otherwise, when only one option is true, the true option must be selected. Note that this setup only makes sense when the original survey does not include categories of the type “None of the above”.

The probability of choosing i in this setup, is the probability of it being presented in the questionnaire times the probability that it is selected by the interviewee, that is

$$\lambda_i = P(i \in \text{questionnaire}) \times P(Y = i) = \frac{2}{t} P(Y = i). \quad (13)$$

We analyze this as the sum of two terms: when it is presented alongside the only false option (j), in which case it will be selected, and when it is paired with another true alternative and a choice must be made between them, so that

$$\begin{aligned} P(Y = i) &= P(Y = i|X \neq j)P(j \in \textit{questionnaire}) \\ &+ \sum_{k \neq i, j}^t P(Y = i|X = k)P(k \in \textit{questionnaire}). \end{aligned} \quad (14)$$

The probability of i being selected if it does not answer the question truthfully is considered to be zero and therefore omitted, hence

$$\lambda_i = \frac{2}{t(t-1)} + \sum_{k \neq i, j}^t \frac{2}{t(t-1)} P(Y = i|X = k). \quad (15)$$

Next, we obtain bounds for selecting i when it is a true alternative by setting $P(Y = i|X = k)$ to zero and one respectively,

$$\lambda_i \in \left[\frac{2}{t(t-1)}, \frac{2(t-2)}{t(t-1)} \right] \text{ for } t \geq 3. \quad (16)$$

Consider the case in which all categories are presented to each respondent and respondents ignore the outcome of the randomizing device. Then, there might be a particular alternative that never gets chosen or an option that is always chosen (except when it proves false to the interviewee). Using the two-option survey scheme, this possible error is greatly reduced.

Finally, if no additional information is available regarding how subjects choose between true alternatives, we assume $P(Y = i|X = k) = \frac{1}{2}$ for all $i \neq k$ and, therefore, that $p_{ij} = \frac{1}{t-1}$ for all $i \neq j$. It should be said that, even though the λ_i 's so obtained have different interpretation from those corresponding to the equal-chance design proposed in the previous section, they are formally equivalent. The sought after proportions, π_i 's, can be estimated as indicated by expression eq. 9 and their variances and covariances are given by eq. 10.

Three interesting features of this approach are worth noting. First, the use of a single randomizing device for every question, independently of the number of categories (for the above scenario, the use of a coin instead of a $(t-1)$ -sided

die). Second, the possibility of conducting a survey without disclosing all of the options to individual respondents. Third, that the error in estimating the p_{ij} 's is bounded, even if the coin were used improperly.

We close this section by stressing the importance of a proper design in the application of a negative questionnaire: how respondents are approached; how questions are posed; how procedures are explained; and so forth, and point out that such issues are beyond the scope of this work.

3. A simulation study of bias

We created an agent-based computer simulation in order to study the precision of a negative question survey *vis a vis* its direct counterpart under different scenarios. The simulation consists of a group of respondents, each belonging to a particular class and equipped to answer both the direct and the negative survey. Respondents make use of a pseudo-random number generator to choose among available options. An unbiased respondent outputs his true class when answering a direct question or chooses an option uniformly at random when answering the negative question.

The purpose of the experiment is to assess the impact of different types of biases on both the positive and the negative question models; we consider errors due to lack of participation and to faulty answers. Our measure of accuracy is a Chi-square test of fit between the real population distribution and the proportions estimated by the negative and direct question surveys.

An archetypical population distribution is created artificially by selecting t integers between 0 and 100, which is then sampled to construct the pool of respondents. Additionally, the sensitivity of a question is modeled by assigning to each of its categories an integer between 1 and 10 (10 being the most sensitive). We assume that a question has the same sensitivity in both its positive and negative version.

Non-response bias is simulated by systematically biasing the pool of respondents towards those with the least sensitive answers. This is achieved by includ-

ing an agent in the survey’s sample with a probability inversely proportional to how sensitive the question is to him/her.

Response bias follows one of these two models:

- The respondent chooses an option, less (more) sensitive than his/her own with a probability inversely (directly) proportional to its sensitivity
- The respondent chooses the least (most) sensitive option. Ties are broken at random.

A respondent changes his/her answer only when it isn’t the least stigmatizing.

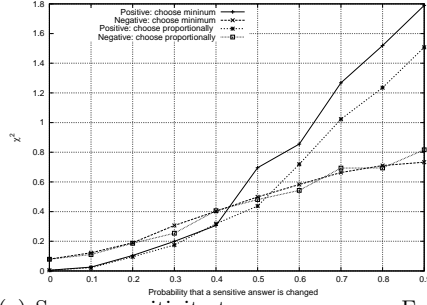
After a simulated negative question survey is run, the number of responses for each category i , n_i , are gathered and used to compute $\hat{\lambda}$ as in equation 5. We estimate the positive proportion of each class π_i following the design of Sect. 2.1, in which respondents presumably use an unbiased device to make their choices, i.e., $p_{ij} = \frac{1}{t-1}$ if $i \neq j$ and $p_{ii} = 0$ in the P matrix. Each data point required 2.5 million runs and represents the average Chi-square discrepancy of 1000 randomly selected population and sensitivity distributions. Each distribution is sampled 50 times and 50 surveys are simulated for each sample. The Chi-square discrepancy was calculated with

$$\chi_{Neg}^2 = \sum_{i=1}^t \frac{(\hat{\pi}_i - \pi_i)^2}{\pi_i} \text{ and } \chi_{Pos}^2 = \sum_{i=1}^t \frac{(\tilde{\pi}_i - \pi_i)^2}{\pi_i} \quad (17)$$

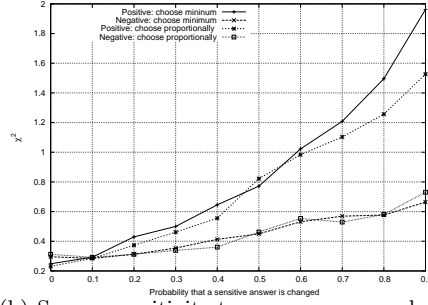
for the negative and positive cases, respectively.

Figure 1(a) shows the behavior of the techniques as a function of the probability that a participant with a sensitive answer changes his/her response. The effect of adding non-response bias is illustrated in Figure 1(b).

The outcome of these simulations suggest that our technique can yield better results than positive question surveys when the nature of the subject matter being studied causes biases in the sample, even if the bias for both kinds of surveys has the same magnitude (although we expect them to be lower for the negative question survey). For instance, considering the non-response and response error models suggested above, we can see from Figure 1(b) that the



(a) Survey sensitivity to response error. Four Classes ($t = 4$), 600 respondents ($n=600$)



(b) Survey sensitivity to non-response and response error. Four Classes ($t = 4$), 600 respondents ($n=600$)

average discrepancy of the negative question survey is lower than its counterpart whenever the response bias exceeds 10%.

4. Conclusions

Survey accuracy depends on minimizing the incidence of nonrespondents and on the honest participation of respondents. In studies where interviewees are required to answer sensitive questions, care must be taken to avert these difficulties by ensuring respondent privacy. In this paper we presented a method for administering a questionnaire mindful of these requirements. The survey in question seeks to estimate the population frequencies of a polychotomous variable: it consists of a single (potentially sensitive) question and t options from which to choose an answer. Its privacy preserving properties do not rely on anonymity, cryptography or on any legal contracts, but rather on participants not revealing their true answer to the survey's query. Respondents are only required to discard, with a known probability distribution, some of the categories that do not answer the question for them. This information is enough to estimate the population proportions of the variable under study; yet, insufficient to ascribe a sensitive datum to a particular individual. We call the method negative question surveys.

Negative question surveys are closely related to RRTs in that both aim at conducting private surveys and both rely on the (secret) use of a randomizing

device to answer questionnaires. One key distinction, however, is that in RRTs participants use the device to choose among questions, at least one of which is sensitive; while in negative question surveys, they use it to choose among answers, avoiding the problem of selecting the proper alternative question altogether. Also, with RRTs some subjects will be selected by the randomizing device, to answer the potentially stigmatizing question. It might still be problematic for them to participate, as the question remains sensitive and answering it demands a measure of trust on the surveying scheme and on the surveyors. A study in which the randomizing device was rigged in order to study respondent behavior is cited in Fox and Tracy (1986); a similar practice could be used to subvert their privacy. Negative question surveys never prompt respondents to answer a sensitive query directly. We also presented a special setup for negative question surveys that reduces the complexity of the randomizing device to a simple fair coin, revealing an important characteristic of our method: an interviewee does not need to contemplate all of the question’s potential answers to pick his/her own, furnishing a level of secrecy to the survey itself and providing robustness against the non-observance of questionnaire instructions (Ref. Ambainis et al. (2004) discussed cryptographic techniques to avoid cheating in RRTs).

We conducted a series of simulations in order to study the effects of response and non-response bias in a negative question survey. Its results indicate that our technique is more robust than direct question surveys to these kinds of errors and that it can have superior accuracy if the probability of such errors is high.

We expect that the privacy of a negative question survey with its comprehensible guarantee and robustness will increase the level of cooperation and accuracy in topic-sensitive studies.

Acknowledgements

The authors gratefully acknowledge the support provided by Asociación Mexicana de Cultura, A. C. The first author would also like to thank Lu-

cinda Gatchel for her valuable insights and the support of the PORTIA project (NSF grant 0331548) for partially funding this research. We would also like to acknowledge the anonymous reviewers whose comments helped improve the presentation of this paper.

References

- Abul-El, A., Greenberg, B., Horvitz, D., 1967. A multiproportions randomized response model. *Journal of the American Statistical Association* 62, 990–1008.
- Agresti, A., Coull, B., 1998. Approximate is better than exact for interval estimation of Binomial proportions. *The American Statistician* 52, 119–126.
- Ambainis, A., Jakobsson, M., Lipmaa, H., 2004. Cryptographic randomized response techniques. In: *Public Key Cryptography PKC 2004*. Springer-Verlag, London, UK, pp. 425–438.
- Bourke, P. D., Dalenius, T., 1973. Multi-proportions randomized response using a single sample. Tech. Rep. 68, University of Stockholm, Institute of Statistics.
- Chaudhuri, A., Mukerjee, R., 1988. *Randomized Response: Theory and Techniques*. Marcel Dekker Inc.
- Fox, J. A., Tracy, P. E., 1986. *Randomized Response: A Method for Sensitive Surveys (Quantitative Applications in the Social Sciences)*. Sage Publications.
- Gjestvang, C., Singh, S., 2006. A new randomized response model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (3), 523–530.
- Horvitz, D., Shah, B., Simmons, W., 1967. The unrelated question randomized response model. In: *Proceedings of the Social Statistics Section, American Statistical Association*.
- Kim, J., Warde, W., 2005. Some new results on the multinomial randomized response model. *Communications in Statistics—Theory and Methods* 34, 847–856.

Warner, S. L., 1965. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60 (309), 63-69.