9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Incentive-Compatible Sampling Estimators

Keywords: surveys, sampling, game theory, mechanism design, Horvitz–Thompson estimator

Extended Abstract

Surveys used to study social phenomena often yield inaccurate and poor results due to study participants deliberately lying or concealing their true opinions. Treating this problem as a form of measurement error results in the researcher applying *post hoc* corrections to account for this form of response bias. However, a line of research in theoretical computer science on 'incentive-compatible learning' (e.g., Cai et al., 2015; Dekel et al., 2010; Hardt et al., 2016) has found estimators which can account for subjects who prefer not to reveal their true opinion *ex ante*. This work formalizes the incentive structures of study respondents using game theory to find estimators with theoretical guarantees concerning their incentive-compatibility. This paper extends this approach to sampling estimators.

We primarily consider the case of the Horvitz-Thompson estimator (Horvitz & Thompson, 1952) although our results extend to the Hansen-Hurwitz estimator (Hansen & Hurwitz, 1943). Prior work by (Caragiannis et al., 2016) finds 'truthful' univariate estimators of a population mean when samples are supplied by strategic agents who wish to pull the estimate as close as possible to their own value. The value of this estimator is most easily seen if we consider the example of surveying the population of a building to find a common temperature to set the building to. For participants who value extreme temperatures, the sample mean is easily manipulated: by lying about their preferred temperature participants can pull the sample mean closer to their desired temperature. This estimator is not 'incentive-compatible' in the sense that participants may gain by lying.

In the setting of sampling estimators considered here, survey panelists can manipulate the information researchers use to determine the probability that they are included the sample. Platforms like Qualtrics and SurveyMonkey rely on panelist information to create strata from which to sample. Panelists who deliberately misreport information which is used to formulate the inclusion probabilities (e.g., age, gender, race, etc) can pull estimates of the population mean of a survey questionoh towards their own question responses. More specifically, we consider the case where a stratified sample of n respondents is drawn from a panel of size N with K distinct strata. Each respondent $i = 1,\ldots,n$ provides a survey question response X_i and their inclusion probability is given by π_k, where $k = s(i) = 1,\ldots,K$ is the corresponding stratum they are drawn from. The Horvitz-Thompson estimate $\hat{\mu}_{HT}$ of the population mean μ is

$$\hat{\mu}_{HT} = \frac{1}{N}\sum_{i=1}^{n}\frac{X_i}{\pi_{s(i)}} \tag{1}$$

We consider the possibility that individuals can provide false information which is used to determine their inclusion probability $\pi_{s(i)}$. Furthermore, we assume that people cannot lie indiscriminately but can only do so by pretending to be in slightly more or less probable strata. If we arrange the strata from least to most probable:

$$\pi_1 \leq \ldots \leq \pi_{s(i)-1} \leq \pi_{s(i)} \leq \pi_{s(i)+1} \leq \ldots \leq \pi_K \tag{2}$$

a panelist can choose to report any of $\pi_{s(i)-1}, \pi_{s(i)}, \pi_{s(i)+1}$[1]. Panelists' utility is given by $-|\hat{\mu}_{HT} - X_i|$, the distance between their question response and the Horvitz-Thompson estimate of the population mean. Notice that when panelists can lie about their strata to influence their inclusion probability, reporting being in strata $s(i) - 1$ will increase the weight of their survey question response X_i on the estimator $\hat{\mu}_{HT}$.

How can sample estimators be made incentive-compatible? To elicit truthful responses, we consider two approaches. If the surveyor can impose a cost on survey respondents, panelists who are surveyed can be penalized proportional to the difference between their question response X_i and the within-strata average. If this cost is large enough, panelists cannot gain by lying about information which determines their inclusion probability. In the case where surveyors cannot impose costs on participants, we explore two ways of modifying the estimator $\hat{\mu}_{HT}$ to ensure incentive compatibility. In the first case, we find a probability ε_k for randomly combining less probable and more probable strata which removes the incentive to lie. In the second case, each survey respondent's question response X_i is replaced by another randomly selected (with replacement) X_i' drawn from the same strata. The effect of outlier responses can be mitigated by a careful choice of re-sampling responses within strata. These modified sampling estimators remove any benefit panelists can derive from lying about information used to determine their inclusion probabilities. We explore under what conditions these estimators are dominant-strategy incentive-compatible and demonstrate their performance relative to the baseline Horvitz-Thompson estimator using simulations.

We take a first step introducing and formalizing considerations of self-interest into the design of sampling estimators. The account presented here adapts an approach in theoretical computer science to explicitly model the incentive structures of study participants to better estimate quantities of interest in the social world.

References

Cai, Y., Daskalakis, C., & Papadimitriou, C. (2015). Optimum statistical estimation with strategic data sources. In P. Grünwald, E. Hazan, & S. Kale (Eds.), *Proceedings of the 28th conference on learning theory* (pp. 280–296). PMLR.

Caragiannis, I., Procaccia, A., & Shah, N. (2016). Truthful univariate estimators. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of the 33rd international conference on machine learning* (pp. 127–135). PMLR.

Dekel, O., Fischer, F., & Procaccia, A. D. (2010). Incentive compatible regression learning. *Journal of Computer and System Sciences*, *76*(8), 759–777.

Hansen, M. H., & Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *The Annals of Mathematical Statistics*, *14*(4), 333–362.

Hardt, M., Megiddo, N., Papadimitriou, C., & Wootters, M. (2016). Strategic classification. *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, 111–122.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

[1] We consider the general form of this problem with $s(i) \pm j$ ordered strata, where $j \leq \lfloor K/2 \rfloor$.