



Beyond the scramble: An optimal path to truth in sensitive surveys

Tanveer Ahmad Tarray^a, Gazala Salam^a, Zahoor Ahmad Ganie^b, Eid Sadun Alotaibi^c

^a Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir, India

^b Department of Electrical Engineering, Islamic University of Science and Technology, Kashmir, India

^c Department of Mathematics and Statistics, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

ARTICLE INFO

Keywords:

Sensitive surveys
Randomized response technique
Optimal estimation
Privacy protection
Scrambling variables

ABSTRACT

Randomized response techniques (RRT) are used by survey statisticians to reduce bias arising from surveys with questions on sensitive characteristics. Few models come with built-in optimal techniques for maximal statistical efficiency, even though the more recent models have enhanced privacy protection. The paper introduces a new Optimal Composite Scrambling model, which incorporates an analytically defined optimal weight in a synergistic way with a sophisticated and high-privacy scrambling mechanism. In the name of strong respondent privacy, the suggested model masks the real response with two scrambling variables by a composite of additive and multiplicative noise. More importantly, we derive a scalar weight that ensures higher precision by minimizing the variance of the resulting estimator. We find the minimum variance of the suggested estimator and prove its unbiasedness. Theoretical comparisons and comprehensive numerical simulations reveal that the OCS model regularly outperforms current state-of-the-art methods. The results point to remarkable gains in percentage relative efficiency, attesting that the proposed method provides a more accurate and reliable tool to estimate the mean of sensitive quantitative variables in sample surveys.

1. Introduction

When researching private, stigmatized, or illegal behaviors, obtaining valid estimates is extremely challenging. The self-preservation

as the mean, $\theta = E(S)$. This “Full” RRT model thus accomplished an important extension of Warner’s ideas on privacy protection to continuous data. Its fundamental disadvantage was its relative inefficiency—the addition of the scrambling variable’s variance, σ_S^2 , naturally inflated