

Measures of privacy in randomized response surveys for quantitative stigmatizing variables

Mausumi Bose*

Indian Statistical Institute, Kolkata 700108, India mausumi@isical.ac.in

In many socio-economic surveys, the variable of interest is sensitive or stigmatizing. Examples include tax evasion, criminal conviction, alcohol expenses, induced abortion, and drug abuse. In such situations, the technique of randomized response is very useful as it does not require the respondent to reveal his/her true value. The issue of privacy protection is important in this context and a few researchers have studied this problem for surveys on dichotomous populations, where the objective is to estimate the proportion of persons bearing the sensitive trait. There is a considerable literature on various randomized response techniques. However, not much is known as yet about the extent of privacy protection when the variable under study is quantitative in nature. In this article we study the issue of privacy protection when the randomized response technique is used for a quantitative variable which could be either discrete or continuous. We propose a measure of privacy for each case. Taking cognizance of a conflict between protection of privacy and enhancing estimation efficiency, we discuss how, given a stipulated level of our privacy measure, the parameters of the randomization device can be determined so as to maximize the efficiency of estimation. Numerical examples are provided to illustrate our results.

Key Words: Continuous sensitive variable, discrete sensitive variable, revealing distribution, distance measure.