Statistical Inferences for Outcome-Dependent Sampling Design with Multivariate Outcomes

by

Haibo Zhou Department of Biostatistics University of North Carolina at Chapel Hill U.S.A. zhou@bios.unc.edu

Abstract

An outcome-dependent sampling (ODS) (Zhou et al. 2002; Weaver and Zhou, 2005) design has been shown to be a cost-effective sampling scheme. In the ODS design with a continuous outcome variable, one observes the exposure with a probability, maybe unknown, depending on the outcome. In practice, multivariate data arise in many contexts, such as longitudinal data or data collected on the basis of participating cluster units. While the ODS design has seen an increased interest in both statistical and applied literature in the recent years, the statistical design and inference procedures for such design with multivariate cases still remain undeveloped. In this talk, we propose three commonly occurring Multivariate-ODS sampling strategies in practice and propose new semiparametric methods for estimating the regression parameters. The first design, the Multivariate-ODS with a maximum selection criterion, selects the supplemental sample based on whether the maximum values of the out! comes that belong to an individual exceed a known cutpoint; the second design, the Multivariate-ODS with a summation criterion, draws the supplemental sample based on whether the sums of the outcome values are above a given cutpoint: the third design, the Multivariate-ODS with a general criterion, select the supplemental samples based on each individual's outcome values, instead of on the aggregate of the outcomes.

The proposed estimators are semiparametric in the sense that all the underlying distributions of covariates are left unspecified and modeled nonparametrically using the empirical likelihood methods. We develop the asymptotic properties for the proposed estimators. Simulation studies illustrate that the proposed estimators are more efficient than the estimator obtained using only the simple-random-sample portion of the Multivariate-ODS and the estimator from a simple random sample of the same sample size. The Multivariate-ODS design and the proposed inference procedures provide an efficient alternative to conduct and analyze multivariate studies. We illustrate the proposed methods with a real data example.

This is joint work with Tsui-shan Lu.