COLLOQUIUM

Department of Statistics and Probability
Michigan State University

Dr. Qin Jing

NIAID/NIH

Empirical likelihood, selection biased sampling, semiparametric generalized linear model: connections and applications

Tuesday November 05, 2013
Wells Hall, Room C405
10:20 a.m. - 11:10 a.m.

Abstract

Since the seminal work by Owen (1988), the empirical likelihood method has become popular in many research areas. The generalized linear model is a standard statistical method since the landmark work by Nelder and Wedderburn (1972). Selection biased sampling occurs when an investigator records an observation by nature according to a certain stochastic model; the recorded observation will not have the original distribution. It can be thought as if the sampled items were drawn from a population whose cumulative distribution function (CDF) is F, but the sampling mechanism is such that the probability of any individual to be included in the sample is proportional to $w(y)$. For different choices of $w(y)$, many applications have been found. For example if $w(y) = y$, this is the well known length biased sampling problem which has wide applications in renewal process (Cox 1969 and Vardi 1982), economic label duration study (Lancaster 1992), epidemiology prevalence sampling (Wolfson et al. 2001), and genetic study (Terwilliger et al. 1997) etc. If $w(y)$ is a probability distribution or a survival function, then this is corresponding to the missing data problem where each data set is observable if the indicator variable $D = 1$ with probability $P(D = 1|y) = w(y)$. Another application of selection biased sampling problem is the case and control study where the most popular choice is $w(y) = \exp(\alpha + \beta y)$ (Anderson 1979).

In this talk we will illustrate the connections between generalized linear model when the baseline carrier density is not specified and selection biased sampling problems. The profile empirical likelihood method is used to find semiparametric maximum likelihood estimates. Applications in genetic mixture models, case control study with secondary outcome and missing data problems will be highlighted.

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.