Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

LIKELIHOOD AND PSEUDO LIKELIHOOD ESTIMATION BASED ON RESPONSE-BIASED OBSERVATION

J.F. Lawless
University of Waterloo

ABSTRACT

Response-biased observation refers to situations where the probability a unit is observed depends on the value of a response associated with that unit. We discuss the construction of estimating equations for parametric regression models through likelihood and pseudo likelihoods, for situations in which responses are stratified and sampling is stratum-specific. Properties of the resulting estimators are reviewed and an illustration involving field reliability data is presented.

1 Introduction

In many observational studies the probability that a specific individual or unit is observed or selected in a sample depends upon responses or covariates associated with that unit. That is, if units in some population have associated response variables y and covariates x, then the probability unit i is selected depends upon the values (y_i, x_i) for that unit. When the probability of selection depends upon y_i, we call the observation scheme response-selective, or response-biased.

For simplicity of exposition I will focus mainly on situations where the probability of selection depends solely on y_i. However, as described at the end of Section 2 and in Section 4, situations where the probability of selection depends on both y_i and x_i may also be handled using the methods considered here.

Examples of response-selective observation are abundant. In socio-economic studies based on samples drawn from administrative records, selection is often response-related (e.g. Hausman and Wise 1981). Similarly, in a study of factors affecting low birth weight of humans, one might select newborns over a period of time and measure covariates so as to over-sample babies with