Estimation and Testing with Interval-Censored Data

Talk to be given at The University of Auckland
Department of Statistics
September 4, 2006

by

Jon A. Wellner
Department of Statistics, University of Washington, Seattle WA
(Currently visiting Victoria University, Wellington)

Abstract: Suppose that X is a random variable (a "survival time") with distribution function F and Y is an independent random variable (an "observation time") with distribution function G. Suppose that we can only observe (Y, Δ) where $\Delta = 1_{[X \le Y]}$, and our goal is to estimate the distribution function F of the random variable X. The Nonparametric Maximum Likelihood Estimator (NPMLE) $\widehat{F}_n$ of F was described in 1955 in papers by H. D. Brunk (and four co-authors), and by C. van Eeden.

A further problem involves inference about the function F at a fixed point, say t_0. If we consider testing $H : F(t_0) = \theta_0$, then one interesting test statistic is the likelihood ratio statistic

$$\lambda_n = \frac{\sup_F L_n(F)}{\sup_{F : F(t_0) = \theta_0} L_n(F)} = \frac{L_n(\widehat{F}_n)}{L_n(\widehat{F}_n^0)} .$$

This involves the additional problem of *constrained* estimation: we need to find the NPMLE $\widehat{F}_n^0$ of F subject to the constraint $F(t_0) = \theta_0$. Inversion of the likelihood ratio tests leads to natural confidence intervals for $F(t_0)$.

Even though the problem of estimating F is non-regular, with associated rate of convergence $n^{-1/3}$ rather than the usual $n^{-1/2}$, the likelihood ratio statistic λ_n has a limiting distribution analogous to the usual χ_1^2 distribution for regular problems which is free of all nuisance parameters in the problem, and this leads to especially appealing tests and confidence intervals for $F(t_0)$.

In this talk I will describe the estimator $\widehat{F}_n$ and its constrained counterpart $\widehat{F}_n^0$, discuss the asymptotic behavior of these estimators and the log-likelihood ratio statistic $2 \log \lambda_n$, and briefly describe the inversion of the tests to obtain confidence intervals. Some open problems concerned with generalizations in several directions will also be mentioned.