ON LOCAL POLYNOMIAL ESTIMATION OF HAZARD RATES AND THEIR DERIVATIVES UNDER RANDOM CENSORING

JIANCHENG JIANG
Peking University

KJELL DOKSUM
University of California at Berkeley

The Dirac function is used to define an empirical hazard rate $\lambda_n(\cdot)$ whose integral up to time t equals to the Nelson-Aalen estimator. This empirical hazard rate exists only in space of Schwartz distributions, so we introduce a local polynomial approximation to $\lambda_n(\cdot)$ which provides estimators of the hazard rate and its derivatives. Consistency and joint asymptotic normality of the local polynomial estimators are established. The estimators have favorable properties similar to those of local polynomial regression estimators, that is, the hazard rate estimator is boundary adaptive and under certain smoothness conditions the rate of convergence can be made arbitrary close to root n. The estimator is boundary corrected even if a local constant smoother is employed. Asymptotic expressions for the mean squared errors (MSE's) are obtained and used in bandwidth selection. A data-driven local bandwidth selection rule is proposed and is illustrated on the Stanford heart transplant data. We use Monte Carlo methods to show that the proposed estimator compares favorably with the Müller–Wang estimator.

1. Introduction

Assume that $T_1, \ldots, T_n$ are i.i.d. lifetimes (that is, nonnegative random variables) with distribution function F, and that $C_1, \ldots, C_n$ are i.i.d. censoring times with distribution function G. The C_i, T_i are assumed to be independent, and the actual observations are (X_i, δ_i), for $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(X_i = T_i)$ is an indicator of the censoring status of X_i.

Let L denote the distribution function of X_i, then $\bar{L} = \overline{F}\,\overline{G}$, where for any distribution function E, $\overline{E} = 1 - E$ is the corresponding survival function. Let $\Lambda(x) = -\log(\overline{F}(x))$ be the cumulative hazard function. We consider the problem of estimating $\lambda(x) = \Lambda'(x) = f(x)/\overline{F}(x)$ and $\lambda^{(k)}(x)$ for $k \geq 0$ on the interval $[0, T]$, where $T < T^* \equiv \inf\{x : L(x) = 1\}$.

Ramlau-Hausen (1983), Tanner and Wong (1983), and Yandell (1983) studied the asymptotic properties of kernel estimators of hazard functions based on the idea of convolution. Müller and Wang (1990) considered local bandwidth choice for convolution-type kernel estimators with fixed higher order kernels, and Müller and Wang (1994) proposed to estimate hazard functions with varying kernels and data-adaptive bandwidths in order to remove boundary effects. Hess et al. (1999) reviewed various kernel-based

Keywords and phrases: Censoring; data-driven local bandwidth selection; Dirac function; hazard rate; local polynomial..

AMS subject classifications: Primary 62G07; Secondary 62G20, 62E25.