SOME ASYMPTOTIC RESULTS OF KERNEL DENSITY ESTIMATORS UNDER RANDOM LEFT-TRUNCATION AND DEPENDENT DATA

VAHID FAKOOR SARAH JOMHOORI

Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Iran. P. O. Box: 1159-91775.
Department of Statistics, Faculty of Sciences, Birjand University, Iran.

ABSTRACT. Problems with truncated data arise frequently in survival analyses and reliability applications. The estimation of the density function of the lifetimes is often of interest. In this paper, the estimation of density function by the kernel method is considered, when truncated data are showing some kind of dependence. We apply the strong Gaussian approximation technique to study the strong uniform consistency for kernel estimators of the density function under a truncated dependent model. We also apply the strong approximation results to study the integrated square error properties of the kernel density estimators under the truncated dependent scheme.

1. Introduction and Preliminaries

In medical follow-up or in engineering life testing studies, one may not be able to observe the variable of interest, referred to hereafter as the lifetime. Among the different forms in which incomplete data appear, right censoring and left-truncation are two common ones. Left truncation may occur if the time origin of the lifetime precedes the time origin of the study. Only subjects that fail after the start of the study are being followed, otherwise they are left truncated. Woodroofe (1985) reviews examples from astronomy and economy where such data may occur.

Let $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_N}$ be a sequence of the lifetime variables which may not be mutually independent, but have a common unknown distribution function (d.f.) F with a density function f. Let $\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_N}$ be a sequence of independent and identically distributed random variables with continuous d.f. G, they are also assumed to be independent of the random variables $\mathbf{X_i}$'s. In the left-truncation model, $(\mathbf{X_i}, \mathbf{T_i})$ is observed only when $\mathbf{X_i} \geq \mathbf{T_i}$. Let $(X_1, T_1), \ldots, (X_n, T_n)$ be the actually observed sample (i.e., $X_i \geq T_i$), and put $\gamma := \mathbf{P}(\mathbf{T_1} \leq \mathbf{X_1}) > 0$, where $\mathbf{P}$ is the absolute probability (related to the N-sample). Note that n itself is a random variable and that γ can be estimated by n/N (although this estimator cannot be calculated since N is unknown). Assume, without loss of generality, that $\mathbf{X_i}$ and $\mathbf{T_i}$ are nonnegative random variables, $i = 1, \ldots, N$. For any d.f. L denote the left and right endpoints of its support by $a_L = \inf\{x : L(x) > 0\}$ and $b_L = \sup\{x : L(x) < 1\}$, respectively. Then under the current model, as discussed by Woodroofe (1985),

Key words and phrases. Integrated square error, Kiefer process, Strong Gaussian approximation, Strong mixing, Strong uniform consistency, Truncated data.

E-mail addresses: fakoor@math.um.ac.ir, Sara_jomhoori@yahoo.com.

133