Journal of Applied Probability and Statistics
Vol. 6, No. 1&2, pp. 73-85
ISOSS Publications 2012

A GENERALIZED REFLECTION METHOD FOR KERNEL DISTRIBUTION AND HAZARD FUNCTIONS ESTIMATION

Jan Koláček

Jan Koláček, Department of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
Email: kolacek@math.muni.cz

Rohana J. Karunamuni

Rohana J. Karunamuni, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada
Email: R.J.Karunamuni@ualberta.ca

SUMMARY

In this paper we focus on kernel estimates of cumulative distribution and hazard functions (rates) when the observed random variables are nonnegative. It is well known that kernel distribution estimators are not consistent when estimating a distribution function near the point $x = 0$. This fact is rather visible in many applications, for example in kernel ROC curve estimation [10]. In order to avoid this problem we propose a bias reducing technique that is a kind of generalized reflection method. Our method is based on ideas of [8] and [19] developed for boundary correction in kernel density estimation. The proposed estimators are compared with the traditional kernel estimator and with the estimator based on "classical" reflection method using simulation studies.

Keywords and phrases: kernel estimation, reflection, distribution function, hazard function.

AMS Classification: 30C40, 62G30.

1 Introduction

The most commonly used nonparametric estimate of a cumulative distribution function F is the empirical distribution function F_n, where $F_n(x) = n^{-1} \sum_{i=1}^{n} I[X_i \leq x]$ with X_1, ..., X_n being the observations. But F_n is a step function even in the case that F is a continuous function. Another type of nonparametric estimator for F is derived from kernel smoothing methods. Kernel smoothing is most widely used because it is easy to apply and produce estimators which have good small and asymptotic properties. Kernel smoothing has received a lot of attention in density estimation. Good references in this area are [3], [16] and [17].