On Kernel Density Estimation for Sum of Two Independent Random Variables

Yu-Sheng Hsu

National Central University, Department of Mathematics

Chungi-Li, Taiwan, R.O.C.

Sy-Mien Chen

Fu Jen University, Department of Mathematics

Taipei, Taiwan, R.O.C.

E-Mail Address: math1013@mails.fju.edu.tw

1. Introduction

Kernel estimation is currently the most popular technique applied in nonparametric statistical inferences. When there are two samples involved, we point out some interesting phenomena in this article.

Let X and Y denote two independent random variables with probability density functions $f(x)$ and $g(y)$ respectively, and $Z = X + Y$. Given two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, there are two ways to estimate

$$f \circ g(z) = \int f(z-t)g(t)dt,$$

the density function of Z. The direct kernel estimator is defined by

$$\hat{f}_{DZ}(z) = \frac{1}{nmh_1} \sum_{i=1}^{n} \sum_{j=1}^{m} K_Z(\frac{z - Z_{ij}}{h_1}),$$

where $Z_{ij} = X_i + Y_j$, K_Z is a kernel function and bandwidth $h_1 = h_1(nm)$ depends on nm (See Prakasa Rao (1983) for details). The indirect kernel estimator is defined as

$$\hat{f}_{IZ}(z) = \int \hat{f}(z-t)\hat{g}(t)dt,$$

where

$$\hat{f}(z-t) = \frac{1}{nh_2} \sum_{i=1}^{n} K_X(\frac{z - t - X_i}{h_2}),$$

and

$$\hat{g}(t) = \frac{1}{mh_3} \sum_{j=1}^{m} K_Y(\frac{t - Y_j}{h_3}),$$

are the ordinary kernel estimators, K_X and K_Y are kernel functions and the bandwidths $h_2 = h_2(n)$ and $h_3 = h_3(m)$ are functions of n and m respectively.

The main purpose of this paper is to compare the large sample performances of $\hat{f}_{DZ}(z)$ and $\hat{f}_{IZ}(z)$.

REFERENCES

Prakasa Rao, B. L. S. (1983). "Nonparametric Functional Estimation," Academic Press.

RESUME

Un maximum de deux variables aléatoires indépendantes a une densité qui peut être mesurée par trois différents estimateurs nucléaires type. Dans cet article nous examinerons les performances asymptotiques de ces trois estimateurs. On fera des comparaisons fondées sur léurs théorèmes sur la limitation et les erreurs du carré moyen.