2024 HAWAII UNIVERSITY INTERNATIONAL CONFERENCES
SCIENCE, TECHNOLOGY & ENGINEERING, ARTS, MATHEMATICS & EDUCATION JUNE 6 - 8, 2024
PRINCE WAIKIKI RESORT, HONOLULU, HAWAII

STRONG CONSISTENCY OF KERNEL ESTIMATORS OF CUMULATIVE DISTRIBUTION FUNCTIONS

CHENG, FUXIA
MATHEMATICS DEPARTMENT
ILLINOIS STATE UNIVERSITY
NORMAL, ILLINOIS

Dr. Fuxia Cheng
Mathematics Department
Illinois State University
Normal, Illinois

Strong Consistency of Kernel Estimators of Cumulative Distribution Functions

Synopsis:

We consider the strong convergence of Lp-norms ($p \geq 1$) of kernel estimator of cumulative distribution function (CDF). Under some mild conditions, the law of the iterated logarithm (LIL) for Lp-norms of empirical processes is extended to the kernel estimator of the CDF.

Strong Consistency of Kernel Estimators of Cumulative Distribution Functions

Abstract

In this paper we consider the strong convergence of L_p-norms ($p \geq 1$) of kernel estimator of cumulative distribution function (CDF). Under some mild conditions, the law of the iterated logarithm (LIL) for L_p-norms of empirical processes is extended to the kernel estimator of the CDF.

1 Introduction

Kernel density estimation (KDE) is a non-parametric method used to estimate the probability density function of a random variable. It is widely applied in statistical analysis, where kernel density estimates play roles in hypothesis testing, classification, and regression. Moreover, KDE aids data visualization by generating kernel density plots, allowing analysts to visually interpret the distribution of data and identify patterns, peaks, tails, and more.

Here we shall develop the good property of the kernel estimators of cumulative distribution functions. Consider an independent identically distributed random sample $X_1, X_2, \cdots, X_n$ from a population with an unknown cumulative distribution function (CDF). For the empirical distribution function F_n, defined as follows,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \forall x \in R^1,$$

with I denoting the indicator function.

The law of the iterated logarithm (LIL) for $F_n(t)$, i.e.,

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log(\log n)}} \sup_x \Big| F_n(x) - F(x) \Big| = \frac{1}{2} \quad a.s. \tag{1.1}$$

has been proved by Smirnov (1944) and, independently, Chung (1949).

Finkelstein (1971) obtained the L_2-version of the law of iterated logarithm,

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log(\log n)}} [\int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x)]^{1/2} = \frac{1}{\pi} \quad a.s. \tag{1.2}$$

For any $p \geq 1$, setting

$$C(p) = \frac{1}{2}(\frac{p(p+2)}{\pi})^{1/2}(\frac{2}{p+2})^{1/p}\frac{\Gamma(1/p+\frac{1}{2})}{\Gamma(1/p)}, \tag{1.3}$$

the law of the iterated logarithm for L_p-norm of $F_n(x)$,

$$\limsup_{n\to\infty}\sqrt{\frac{n}{2\log(\log n)}}\Big[\int_{-\infty}^{\infty}|F_n(x)-F(x)|^p dF(x)\Big]^{1/p} = C(p) \;\; a.s. \tag{1.4}$$

was developped by Gajek, Kahszka, and Lenic (1996).

Notice that there is one serious discontinuity drawback of F_n, regardless of F being continuous or discrete. To treat this deficiency of F_n, Yamato (1973) proposed the following kernel distribution estimator

$$\hat{F}(x) = \int_{-\infty}^{x} n^{-1}\sum\nolimits_{i=1}^{n} k_h(u-X_i)\,du,\; x\in\mathbb{R}, \tag{1.5}$$

in which $h = h_n > 0$ is called bandwidth, k is a probability density function(PDF) called kernel, and $k_h(u) = k(u/h)/h$.

The aim of this paper is to provide certain conditions to guarantee the LIL of L_p-norm of $\hat{F}$.

2 The main result

Under certain assumptions, for any $p \geq 1$ and the continuous CDF F with bounded second order derivative, we have

$$\limsup_{n\to\infty}\sqrt{\frac{n}{2\log(\log n)}}\Big[\int_{-\infty}^{\infty}|\hat{F}(x)-F(x)|^p dF(x)\Big]^{1/p} = C(p) \qquad a.s., \tag{2.1}$$

where $C(p)$ is defined in (1.3)

References

Cheng,F. (2017). Strong uniform consistency rates of kernel estimators of cumulative distribution functions, *Communications in Statistics—Theory and Methods*, 46, 6803-6807.

Chung, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc., 67:36-50.*

Csörgő, M. and Révész ,P. (1981). *Strong Approximation in Probability and Statistics*, Academic Press, New York, .

Fabian, V. and Hannan, J. (1985). *Introduction to Probability and Mathematical Statistics.*

Finkelstein, H. (1971), The law of the iterated logarithm for empirical distributions, *Ann. Math.Statist. 42, 607-615.*

Gajek, L. Kahszka, M. and Lenic, A. (1996). The law of the iterated logarithm for Lp-norms of empirical processes, *Statist. Probab. Lett., 28, 107–110.*

Smirnov, N. V. (1944). An approximation to the distribution laws of random quantities deter- mined by empirical data, *Uspehi. Mat. Nauk., 10:179-206.*

Wang, J., Cheng, F. and Yang, L. (2013). Smooth simultaneous confidence band for cumulative distribution function, *Journal of Nonparametric Statistics, 25, 395-407.*

Winter, B.B, (1979). Convergence Rate of Perturbed Empirical Distribution Functions, *Journal of Applied Probability, 16, 163-173.*

Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, *Bull. Math. Statist.* 15, 69-78.