Journal of the Korean Statistical Society 44 (2015) 530–537

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Frequency polygon estimation of density function for dependent samples

Xin Yang

Department of Mathematics, Guilin University of Aerospace Technology, Guilin 541004, China

ARTICLE INFO

Article history:
Received 27 July 2014
Accepted 28 January 2015
Available online 21 February 2015

AMS 2000 subject classifications:
primary 62G05
secondary 62G20

Keywords:
Frequency polygon
Uniform strong consistency
Strong mixing sample

ABSTRACT

The frequency polygon, which is a density estimator based on histogram technique, has the advantages of computational simplicity and is widely used in many fields. The purpose of this article is to further investigate the uniform strong consistency of frequency polygon under strong mixing samples. Our conclusions make the conditions on mixing coefficient and bin width are more simplified and weaker than those of Carbon et al. (1997).

1. Introduction

In many of the density estimations, histogram estimation has the advantage of computational simplicity. It is widely used in many fields. So, there are many literature about the estimation based on histogram technique. For example, Eidous (2005, 2011) proposed an additive histogram estimator for the wildlife population density from the line transect data when the shoulder condition is not satisfied. Liu and Van Ryzin (1985) proposed a histogram type estimation of hazard rate in the setting of censored data and studied the asymptotic properties of the estimator including strong consistency, strong uniform consistency and asymptotic distribution. Recently, Patil and Bagkavos (2012) investigated its mean square error properties and discussed the choice of bin width.

For more researches, one can refer to Chen and Zhao (1987), Scott (1979), Tran (1994), Zhao, Krishnaiah, and Chen (1990) for histogram density estimation; Chamayou (1980), Fernando, Maier, and Dandy (2009), Scott (1985a, 2010) for average shifted histogram; Dong and Zheng (2001), Jones, Samiuddin, Al-Harbey, and Maatouk (1998) for edge frequency polygon; and so on.

This paper will pay attention to frequency polygon, which is constructed by connecting with straight lines the mid-bin values of a histogram. Let $\{X_i : i \geq 1\}$ be a sequence of identically distributed random variables with common density $f(x)$ on the real line R. Consider a partition $\cdots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 \cdots$ of the real line into equal intervals $I_{n,j} \equiv [(j-1)b_n, jb_n)$ of length b_n, where b_n is the bin width and $j = 0, \pm 1, \pm 2, \ldots$. For $x \in [(j-1/2)b_n, (j+1/2)b_n)$, consider the two adjacent histogram bins $I_{n,j}$ and $I_{n,j+1}$. Denote the number of observations falling in these intervals respectively by $\upsilon_{n,j}$ and $\upsilon_{n,j+1}$. Then the values of the histogram in these previous bins are given by

$$f_{n,j} = \upsilon_{n,j}/(nb_n), \qquad f_{n,j+1} = \upsilon_{n,j+1}/(nb_n). \tag{1.1}$$

E-mail address: xinyang_emily@sina.cn.

http://dx.doi.org/10.1016/j.jkss.2015.01.006
1226-3192/

Thus the frequency polygon estimation of the density function $f(x)$ is defined as follows

$$\widehat{f}_n(x) = \left(\frac{1}{2} + j - \frac{x}{b_n}\right) f_{n,j} + \left(\frac{1}{2} - j + \frac{x}{b_n}\right) f_{n,j+1} \tag{1.2}$$

for $x \in [(j - 1/2)b_n, (j + 1/2)b_n)$. In many literature, j is usually regarded as zero. Here it is reserved for convenience of description later.

$\{X_i\}$ is called to be strong mixing or α-mixing, if it satisfies the strong mixing condition

$$\alpha(n) = \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+n}^\infty} |P(AB) - P(A)P(B)| \downarrow 0, \quad \text{as } n \to \infty, \tag{1.3}$$

where $\mathcal{F}_1^k = \sigma(X_i, 1 \leq i \leq k)$ and $\mathcal{F}_{k+n}^\infty = \sigma(X_i, i > k + n)$.

The strong mixing sequence is an important kind of dependency, it contains many types of mixing sequences, such as ψ-mixing, φ-mixing, ρ-mixing and so on. The important properties about it can be referred to the Bradley (2005).

The frequency polygon was comprehensively researched by Scott (1985b). It is show that the frequency polygon has rates of convergence similar to those of modem non-negative kernel estimators and greater than the rate for a histogram. Later, Carbon, Garel, and Tran (1997) extended the discussion to the case of strong mixing process, showed the integrated mean squared error and uniform strong consistency of the estimator. Recently, Carbon, Francq, and Tran (2010) established the asymptotic normality of frequency polygons for random fields, while Bensaïd and Dabo-Niang (2010) studied the mean square error and uniform strong consistency for the continuous random fields. Moreover, Xing, Yang, and Liang (2014) proved the uniformly strong or weak consistency of the estimator for ψ-mixing samples, and gave the convergence rate which is nearly equal to the one obtained by Carbon et al. (1997).

The purpose of this paper is to further research the uniform strong consistency of frequency polygon under strong mixing samples. Our conclusions make the conditions on mixing coefficient and bin width are more simplified and weaker than those of Carbon et al. (1997).

Our paper is organized as follows: Section 2 presents the main results. Section 3 provides some lemmas for the proofs of theorem. Section 4 is to prove the theorems in Section 2.

Throughout this paper, the letter C will be used to denote positive constants whose values are unimportant and may vary, but not dependent on the sample size n.

2. Main results

We need the following basic assumptions.

(A1) As $n \to \infty$, $b_n \to 0$ and $nb_n/\log n \to \infty$.

(A2) $\{X_i : i \geq 1\}$ is a sequence of strong mixing and identically distributed random variables. X_1 has an uniformly continuous density $f(x)$ on the real line R and the mixing coefficient $\alpha(n) = O(n^{-\rho})$ where $\rho \geq 2$.

(A3) The joint density function $f_{ij}(x, y)$ of (X_i, X_j) exists and satisfies

$$\sup_{(x,y)\in R^2} |f_{ij}(x, y) - f(x)f(y)| \leq M, \quad \forall i \neq j, \tag{2.1}$$

for some positive constant M.

(A4) There exists a constant $C > 0$ such that

$$|f(x) - f(x')| \leq C|x - x'| \quad \text{for } x, x' \in R.$$

(A5) $E|X|^{2/T} < \infty$ for some $T > 0$.

Theorem 2.1. *Suppose that* (A1)–(A3) *are satisfied.* (1) *If for some* $\varepsilon > 0$

$$g(n, 1) \equiv nb_n^{-2}(nb_n/\log n)^{-\rho} \log n(\log\log n)^{1+\varepsilon} \to 0, \tag{2.2}$$

then for any compact subset D of R, we have

$$\sup_{x\in D} |\widehat{f}_n(x) - f(x)| = o(1), \quad a.s. \tag{2.3}$$

(2) *If* (A5) *is satisfied and* $g(n, 1)n^T \to 0$, *then*

$$\sup_{x\in R} |\widehat{f}_n(x) - f(x)| = o(1), \quad a.s. \tag{2.4}$$

Denote $\Psi_n = \max\{b_n, (\log n/(nb_n))^{1/2}\}$.

Theorem 2.2. *Suppose that* (A1)–(A4) *are satisfied.* (1) *If for some* $\varepsilon > 0$

$$g(n, 2) \equiv nb_n^{-2}(nb_n/\log n)^{-\rho/2+1/2} \log n(\log\log n)^{1+\varepsilon} \to 0, \tag{2.5}$$