Journal of Sciences, Islamic Republic of Iran 17(3): 249-251 (2006) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Estimation of the Survival Function for Negatively Dependent Random Variables

H. Doosti[1,*] and H. Zarei[2]

[1] *Department of Statistics, School of Mathematical Sciences, Ferdowsi University, P.O. Box 1159-91775, Mashhad, Islamic Republic of Iran*
[2] *Department of Mathematics, Faculty of Sciences, Zabol University, Zabol, Islamic Republic of Iran*

Abstract

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of pair wise negative quadrant dependent random variables with survival function $\bar{F}(x) = P[X > x]$. The empirical survival function $\bar{F}_n(x)$ based on $X_1, X_2, ..., X_n$ is proposed as an estimator for $\bar{F}_n(x)$. Strong consistency and point wise as well as uniform of $\bar{F}_n(x)$ are discussed.

Keywords: Survival function; Pair wise negative quadrant dependent random variables; Uniform strong consistency

1. Introduction

Suppose that $\{X_n, n \geq 1\}$ is a sequence of random variables with distribution function $F(x)$, or equivalently, survival function $\bar{F}(x) = P[X > x]$. An estimator of $\bar{F}(x)$ been studied by Bagai and Prakasa Rao [1] for the case where $\{X_n, n \geq 1\}$ is for associated random variables.

Consider the estimator $\bar{F}_n(x)$ defined by

$$\bar{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} Y_j(x) \tag{1.1}$$

where

$$Y_i(x) = \begin{cases} 1 & , X_i > x, \\ 0 & , \textit{otherwise}. \end{cases} \tag{1.2}$$

Definition 1. The random variables $X_1, ..., X_n$ are said to be negatively dependent (ND) if we have

$$P[\bigcap_{j=1}^{n} (X_j \leq x_j)] \leq \prod_{j=1}^{n} P(X_j \leq x_j),$$

and

$$P[\bigcap_{j=1}^{n} (X_j > x_j)] \leq \prod_{j=1}^{n} P(X_j > x_j),$$

for all $x_1, ..., x_n \in R$. An infinite sequence $\{X_n, n \geq 1\}$ is said to be ND if every finite subset $X_{i_1}, ..., X_{i_n}$ is ND.

We propose $\{X_n, n \geq 1\}$ is a sequence of negatively dependent random variables and $\bar{F}_n(x)$ as an estimator for $\bar{F}(x)$ and study it. In this paper we discuss the strong consistency, point wise and uniform of $\bar{F}_n(x)$. These results are useful in the study of kernel-type

*E-mail: *doosti@math.um.ac.ir*

density and failure rate estimators of the unknown density and failure rate function. In fact we extend Bagai and Prakasa Rao [1] to negatively dependent case.

The following lemma was proved by Bozorgnia, Patterson and Taylor [2]. We use it for obtaining the main result in the next section.

Lemma 1.1. ([1]) Let $\{X_n, n \geq 1\}$ be a sequence of ND random variables and $\{f_n, n \geq 1\}$ be a sequence of Borel functions all of which are monotone increasing (or all are monotone decreasing). Then $\{f_n(X_n), n \geq 1\}$ is a sequence of ND random variables.

Throughout the paper C will denote a positive constant not necessarily the same from one step to another.

2. The Empirical Survival Function

First, we present a bound for the moment of order p of the sum of N random variables which depends on the second moment and mixing coefficients. This bound constitutes the basis of the main results of this paper– Theorems 2.1 and 2.2. This is a Rosenthal-type inequality. The following lemma was proved by Rivaz [4] theorem 7.2.2. page 32.

Lemma 2.1. Let $\xi_1, ..., \xi_n$ be a sequence of ND identically random variables such that $\mathbf{E}(\xi_i) = 0$, $\| \xi_i \|_\infty < M$. Then there exist $C(p)$ such that

$$\mathbf{E}(|\sum_{i=1}^{n} \xi_i|^p) \leq C(p)\{M^{p-2} \sum_{i=1}^{n} \mathbf{E}(\xi_i^2)$$

$$+ (\sum_{i=1}^{n} \mathbf{E}(\xi_i^2))^{p/2}\}, \quad p > 2$$

Theorem 2.1. Let $\{X_n, n \geq 1\}$ be a stationary sequence of ND random variables with bounded continuous density for X_1. Then for some $r > 1$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

$$\sup_x P[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon] \leq C\varepsilon^{-2r} n^{-r}$$

for every $n \geq 1$.

Proof. Using Markov inequality, we get that for every $\varepsilon > 0$,

$$\sup_x P[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon] =$$

$$\sup_x P[|\bar{F}_n(x) - \bar{F}(x)|^{2r} > \varepsilon^{2r}] \tag{2.1}$$

$$\leq \sup_x \{(n\varepsilon)^{-2r} E\, |\sum_{i=1}^{n} (Y_i - EY_i)|^{2r}\}$$

to complete the proof, it is sufficient to estimate $E\, |\sum_{i=1}^{n} (Y_i - EY_i)|^{2r}$. Denote $\xi_i = Y_i - EY_i$. Note that $\| \xi_i \|_\infty < 2$ and $E\,\xi_i = 0$. In view of ND property of the sequence $\{X_n,\ n \geq 1\}$ and the monotonicity of the function Y_i, Lemma 1.1 follows that the sequence $\{\xi_n,\ n \geq 1\}$ is also sequence of ND random variables. Hence applying the Lemma 2.1 we have

$$E\, |\sum_{i=1}^{n} (Y_i - EY_i)|^{2r} \leq Cn^{-r}. \tag{2.2}$$

By substituting (2.2) in (2.1), we obtain the desired result.

Corollary 2.1. Under the conditions of Theorem 2.1 for every x,

$$\bar{F}_n(x) \to \bar{F}(x) \quad a.s. \qquad as \qquad n \to \infty.$$

Proof. For $r > 1$ observe that

$$\sum_{n=1}^{\infty} P[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon] \leq C\varepsilon^{-2r} \sum_{n=1}^{\infty} n^{-r} < \infty$$

The result then follows by using the Borel-Contelli Lemma. Next we obtain a version of Glivenko-Cantelli Threorem valid for *ND* random variables. The proof follows along the lines of analogous result for associated of random variables (Bagai and Prakasa Rao [1]).

Theorem 2.2. Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of ND random variables satisfying the conditions of Theorem 2.1. Then for any compact subset $J \subset R$,

$$\sup[|\bar{F}_n(x) - \bar{F}(x)|:\ x \in J] \to 0 \ \ a.s \ \ as \ \ n \to \infty.$$

Proof. Let K_1 and K_2 be chosen such that $J \subset [K_1, K_2]$ into b_n sub-intervals of length $\delta_n \to 0$ where $\{\delta_n\}$ is chosen such that

$$\sum_n \delta_n^{-1} n^{-r} < \infty. \tag{2.3}$$

such a choice of $\{\delta_n\}$ is possible. For instance, choose $\delta_n = n^{-\theta}$ where $0<\theta<r-1$. Note that $b_n \le C\,\delta_n^{-1}$.

Let $I_{nj} = (x_{n,j}, x_{n,j+1})$, $j=1,\ldots,b_n = N$, where

$$K_1 = x_{n,1} < x_{n,2} < \ldots < x_{n,N+1} = K_2,$$

with

$$x_{n,j+1} - x_{n,j} \le \delta_n \quad \text{for } 1\le j\le N.$$

Then for $x\in I_{nj}$, $j=1,2,\ldots,N$ we have

$$\bar F(x_{n,j+1}) \le \bar F(x) \le \bar F(x_{n,j}),$$

and

$$\bar F_n(x_{n,j+1}) \le \bar F_n(x) \le \bar F_n(x_{n,j}).$$

Hence

$$[\bar F_n(x_{n,j+1}) - \bar F(x_{n,j+1})] + [\bar F(x_{n,j+1}) - \bar F(x)]$$
$$\le \bar F_n(x) - \bar F(x)$$
$$\le [\bar F_n(x_{n,j}) - \bar F(x_{n,j})] + [\bar F(x_{n,j}) - \bar F(x)].$$

Therefore

$$\begin{aligned}
&\sup[\,|\bar F_n(x) - \bar F(x)|:\ x\in J\,]\\
&\le \sup[\,|\bar F_n(x) - \bar F(x)|:\ K_1\le x\le K_2\,]\\
&\le \max_{1\le j\le N} |\bar F_n(x_{n,j}) - \bar F(x_{n,j})|\\
&+\max_{1\le j\le N} |\bar F(x_{n,j+1}) - \bar F(x_{n,j+1})| \qquad (2.4)\\
&+\max_{1\le j\le N}\sup_{x\in I_{nj}} |\bar F_n(x_{n,j}) - \bar F(x)|\\
&+\max_{1\le j\le N}\sup_{x\in I_{nj}} |\bar F(x_{n,j+1}) - \bar F(x)|.
\end{aligned}$$

Now by the mean value theorem for $x_{n,j} < u^* < x$ we have

$$\begin{aligned}
\bar F(x_{n,j}) - \bar F(x) &= F(x) - F(x_{n,j})\\
&= (x - x_{n,j}) f(u^*).
\end{aligned} \qquad (2.5)$$

Since f, the density of X_1 is bounded by the hypothesis, it follows that there exists a constant $C>0$ such that

$$|\bar F(x_{n,j}) - \bar F(x)| \le C\,\delta_n, \quad |\bar F(x_{n,j+1}) - \bar F(x)| \le C\,\delta_n,$$

for $1\le j\le N$ and $x\in I_{nj}$. Then for $\varepsilon>0$, choose $n=n(\varepsilon)$ such that

$$2C\,\delta_n \le \frac{1}{3}\varepsilon.$$

From (2.3) and (2.4), we get, for $n\le n(\varepsilon)$,

$$\begin{aligned}
&P[\sup_{x\in J} |\bar F_n(x) - \bar F(x)| > \varepsilon]\\
&\le P[\max_{1\le j\le N} |\bar F_n(x_{n,j}) - \bar F(x_{n,j})| > \frac{1}{3}\varepsilon]\\
&+P[\max_{1\le j\le N} |\bar F(x_{n,j+1}) - \bar F(x_{n,j+1})| > \frac{1}{3}\varepsilon]\\
&\le \sum_{j=1}^{N} P\ |\bar F_n(x_{n,j+1}) - \bar F(x_{n,j})| > \frac{1}{3}\varepsilon]\\
&+\sum_{j=1}^{N} P[|\bar F_n(x_{n,j+1}) - \bar F(x_{n,j})| > \frac{1}{3}\varepsilon]\\
&\le CN\,\varepsilon^{-2r} n^{-r}\\
&= C\,\varepsilon^{-2r} b_n n^{-r} \quad \text{(by Theorem 2.1)}\\
&\le C\,\varepsilon^{-2r}\delta_n^{-1} n^{-r}
\end{aligned}$$

The result follows by using (2.5) and Borel-Cantelli Lemma.

References

1. Bagai I. and Prakasa Rao B.L.S. Estimation of the survival function for stationary associated processes. *Statistics and Probability Letters*, **12**: 385-391 (1991).
2. Bozorgnia A., Patterson R.F., and Taylor R. Limit theorems for dependent random variables. *World Congress Nonlinear Analysts*, **92**: 1639-1650 (1996).
3. Lehman E.L. Some concepts of dependence. *Annals of Mathematics and Statistics*, **37**: 1137-1153 (1966).
4. Rivaz F. Negative association of random variables and complete of weighted sum of NA. M.Sc. *Dissertation* (2002).