Pak. J. Statist.
2009 Vol. 25(1), 1-4

SURVIVAL FUNCTION ESTIMATION FOR M-DEPENDENT PROCESSES

Hassan Zarei
Zabol University, Zabol, Iran.
Email: hzarei51@gmail.com

ABSTRACT

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of m-dependent random variables with survival function $\bar{F}(x) = P\left[X_1 > x\right]$. 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 pointwise as well as uniform of $\bar{F}_n(x)$ are discussed.

KEY WORDS

Survival function; m-dependent random variables; uniform strong consistency

1. INTRODUCTION

Let $\{X_n,\ n \geq 1\}$ be a sequence of m-dependent random variables with distribution function $F(x)$, or equivalently, survival function $\bar{F}(x) = P\left[X_1 > x\right]$. We impose conditions on the covariance structure of survival function $\bar{F}(x)$ based on the observations $\{X_1, X_2, ..., X_n\}$. For a sequence of associated random variables, Bagai and Prakasa Rao (1991) proposed an estimator. Doosti and Zarei (2006) extended Bagai's (1991) result in the case of negatively associated random variables. The purpose of this note is to extend these results for estimating the survival function of *m-dependent sequences* of random variables. Consider the estimator $\bar{F}_n(x)$ defined by

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

where

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

We propose $\bar{F}_n(x)$ as an estimator for $\bar{F}(x)$. In this paper we discuss the strong consistency, pointwise and uniform of $\bar{F}_n(x)$. These results are useful in the study of kernel-type density and failure rate estimators of the unknown density and failure rate function. In fact we extend Bagai and Rao (1991) to m-dependent case. Some lemmas, useful in proving the result concerning $\bar{F}_n(x)$, are stated and proved in Section 2.

2. MAIN RESULTS

The following Lemma was proved by Romano and Wolf (2000) (Corollary A.1. P. 121).

Lemma 2.1

Let $\{X_i, i \geq 1\}$ be an m-dependent sequence of mean zero random variables. Assume $\mathbf{E}\left(|X_i|^q\right) \leq \Delta$ for some $q \geq 2$ and all i. Then, for all $n \geq 2m$

$$\mathbf{E}\left(\left|\sum_{i=1}^{n} X_i\right|^q\right) \leq C_q \Delta \left(4mn\right)^{q/2},$$

where C_q is a positive constant depending only upon q.

Theorem 2.1

Let $\left\{X_n,\ n \geq 1\right\}$ be an m-dependent sequence of random variables with bounded continuous density for X_1. Then

i) for some $r > 2$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

$$\sup_x P\left[\left|\bar{F}_n(x) - \bar{F}(x)\right| > \varepsilon\right] \leq C\varepsilon^{-2r}\left(\frac{m}{n}\right)^r \quad \textit{for every } n \geq 1.$$

ii) for any compact subset $J \subset R$,

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

Remark 1. If one considers m as a fix integer, then it can be shown that the upper bound in Theorem 2.1 (i) is $C\varepsilon^{-2r} n^{-r}$ which is an analogue bound in Theorem 2.1 in Bagai and Rao (1991).

Remark 2. In Theorem 2.1, (ii) is a version of Glivenko-Cantelli Theorem valid for m-dependent random variables. The proof follows along the lines of analogous result for associated of random variables (Bagai and Rao 1991).

Proof. By using Markov inequality, for every $\varepsilon > 0$ we have

$$\begin{aligned} \sup_x P\left[\left|\bar{F}_n(x) - \bar{F}(x)\right| > \varepsilon\right] &= \sup_x P\left[\left|\bar{F}_n(x) - \bar{F}(x)\right|^{2r} > \varepsilon^{2r}\right] \\ &\leq \sup_x \left\{(n\varepsilon)^{-2r} E\left|\sum_{i=1}^{n}(Y_i - EY_i)\right|^{2r}\right\}. \end{aligned} \tag{2.1}$$

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

$$E\left|\sum_{i=1}^{n}\left(Y_i - EY_i\right)\right|^{2r} \leq C(mn)^r. \tag{2.2}$$

By substituting (2.2) in (2.1), we obtain the desired result in (i). For proof of second part of Theorem, 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 $b_n \leq C\delta_n^{-1}$. Let $I_{nj} = \left(x_{n,j},\ x_{n,j+1}\right),\ j = 1, \ldots,\ b_n = N,$ where $K_1 = x_{n,1} < x_{n,2} < \ldots < x_{n,N+1} = K_2$, with $x_{n,j+1} - x_{n,j} \leq \delta_n$ for $1 \leq j \leq N$. Then for $x \in I_{nj}$, $j = 1, 2, \ldots, N$ we have $\bar{F}\left(x_{n,j+1}\right) \leq \bar{F}(x) \leq \bar{F}\left(x_{n,j}\right)$, and $\bar{F}_n\left(x_{n,j+1}\right) \leq \bar{F}_n(x) \leq \bar{F}_n\left(x_{n,j}\right)$. Hence

$$\begin{aligned}&\left[\bar{F}_n\left(x_{n,j+1}\right) - \bar{F}\left(x_{n,j+1}\right)\right] + \left[\bar{F}\left(x_{n,j+1}\right) - \bar{F}(x)\right] \\ &\quad \leq \bar{F}_n(x) - \bar{F}(x) \leq \left[\bar{F}_n\left(x_{n,j}\right) - \bar{F}\left(x_{n,j}\right)\right] + \left[\bar{F}\left(x_{n,j}\right) - \bar{F}(x)\right].\end{aligned}$$

Therefore

$$\begin{aligned}\sup_x&\left[\left|\bar{F}_n(x) - \bar{F}(x)\right| : \quad x \in J\right] \leq \sup_x\left[\left|\bar{F}_n(x) - \bar{F}(x)\right| : \quad K_1 \leq x \leq K_2\right] \\ &\leq \max_{1\leq j\leq N}\left[\left|\bar{F}_n\left(x_{n,j}\right) - \bar{F}\left(x_{n,j}\right)\right|\right] + \max_{1\leq j\leq N}\left|\bar{F}\left(x_{n,j+1}\right) - \bar{F}\left(x_{n,j+1}\right)\right| \\ &+ \max_{1\leq j\leq N}\sup_{x\in I_{nj}}\left|\bar{F}_n\left(x_{n,j}\right) - \bar{F}(x)\right| + \max_{1\leq j\leq N}\sup_{x\in I_{nj}}\left|\bar{F}\left(x_{n,j+1}\right) - \bar{F}(x)\right|.\end{aligned} \tag{2.4}$$

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

$$\bar{F}\left(x_{n,j}\right) - \bar{F}(x) = F(x) - F\left(x_{n,j}\right) = \left(x - x_{n,j}\right) f\left(u^*\right). \tag{2.5}$$

Since f, the density of X_1 is bounded by the hypothesis, it follows that there exists a constant $C > 0$ such that $\left|\bar{F}\left(x_{n,j}\right) - \bar{F}(x)\right| \leq C\delta_n$, and $\left|\bar{F}\left(x_{n,j+1}\right) - \bar{F}(x)\right| \leq C\delta_n$, for $1 \leq j \leq N$ and $x \in I_{nj}$. Then for $\varepsilon > 0$, choose $n = n(\varepsilon)$ such that $2C\delta_n \leq \frac{1}{3}\varepsilon.$

From (2.4) and (2.5), we get, for $n \le n(\varepsilon)$,

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

The result follows by using (2.3) and Borel-Cantelli Lemma.

Corollary 2.1

Under the conditions of Theorem 2.1, if $m = O\left(n^{1-\frac{p}{r}}\right)$ where $p > 1$, 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

$$
\begin{aligned}
\sum_{n=1}^{\infty} P\left[\left|\bar{F}_n(x)-\bar{F}(x)\right|>\varepsilon\right] &\le C\varepsilon^{-2r}\sum_{n=1}^{\infty}\left(\frac{m}{n}\right)^r \\
&\le C\varepsilon^{-2r}\sum_{n=1}^{\infty} n^{-p} < \infty
\end{aligned}
$$

The result then follows by using the Borel-Contelli Lemma.

REFERENCES

1. Bagai, I. and Rao, Prakasa B.L.S. (1991). Estimation of the survival function for stationary associated processes. *Statist. Prob. Letters*, 12, 385-391.
2. Doosti, H. and Zarei, H. (2006). Estimation of the survival function for negatively associated processes. *J. Sci. Islam. Repub. Iran*. 17(3), 249-251.
3. Romano, J.P. and Wolf, M. (2000). A more general central limit theorem for m-dependent random variables with unbounded. *J. Theoretical. Prob.* 13, 343-356.
4. Sadikova, S.M. (1966). Two dimensional analogues of an inequality of Essen with applications to the central limit theorem, *Theory Prob.. Appl.* 11, 325-335.