Pak. J. Statist.
2009 Vol. 25(3), 235-240

ESTIMATION OF THE SURVIVAL FUNCTION FOR A DISCRETE-TIME STOCHASTIC PROCESS

Hassan Doosti and **Hosein Ali Niroumand**
Department of Statistics, School of Mathematical Sciences,
Ferdowsi University of Mashhad, Mashhad, Iran
Email: doosti@math.um.ac.ir and hniroumand@yahoo.com

ABSTRACT

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with survival function $\bar{F}(x) = P[X_1 > x]$. The empirical survival function $\bar{F}_n(x)$ based on $X_1, X_2, ..., X_n$ is proposed as an estimator for $\bar{F}_n(x)$. We suppose that the process is strongly mixing and we show strong consistency and pointwise as well as uniform of $\bar{F}_n(x)$ are depended on the behavior of a special quadratic characteristic.

KEYWORDS

Survival function; strongly mixing; associated; uniform strong consistency.

1. INTRODUCTION

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with distribution function $F(x)$, or equivalently, survival function $\bar{F}(x) = P[X_1 > x]$. 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

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

We propose $\bar{F}_n(x)$ as an estimator for $\bar{F}(x)$ and study it. 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. Bagai and Prakasa Rao (1991) proposed $\bar{F}(x)$ and studied strong consistency of it for sequence of associated random variables. Doosti and Zarei (2006) extended their results to negatively associated case. Shirazi and Doosti (2008) and Zarei (2009) extended the results for m-dependent sequence of random variables. If we want the survival function estimator (1.1) for a stochastic process to attain the same result as for the associated, negatively associated and m-dependent cases, we have to impose certain *weak dependence* conditions on the considered process $\{X_n,\ n \geq 1\}$ defined on the $(\Omega, \aleph, P)$. Let $\mathbf{N}_k^m$ denote the σ-algebra generated by the events

$$\{X_k \in A_k, \ldots, X_m \in A_m\}.$$

We consider the following classical mixing conditions:

1. s*trong mixing* (s.m.), also called α-mixing,

$$\sup_m \sup_{A \in \mathbf{N}_1^m,\, B \in \mathbf{N}_{m+s}^\infty} \left|P(AB) - P(A)P(B)\right| = \alpha(s) \to 0 \quad as \quad s \to \infty,$$

2. *complete regularity* (c.r.), also called β-mixing,

$$\sup_m E\left\{Var_{B \in \mathbf{N}_{m+s}^\infty} \left|P\left(B \mid \mathbf{N}_1^m\right) - P(B)\right|\right\} = \beta(s) \to 0 \quad as \quad s \to \infty,$$

3. *uniformly strong mixing* (u.s.m.), also called φ - mixing,

$$\sup_m \sup_{A \in \mathbf{N}_1^m,\, P(A)>0,\, B \in \mathbf{N}_{m+s}^\infty} \frac{|P(AB) - P(A)P(B)|}{P(A)} = \varphi(s) \to 0 \quad as \quad s \to \infty,$$

4. ρ*-mixing*

$$\sup_m \sup_{X \in L^2(\mathbf{N}_1^m),\, Y \in L^2(\mathbf{N}_{m+s}^\infty)} \left|corr(X, Y)\right| = \rho \to 0 \quad s \to \infty.$$

Following (Davydov, 1973) we denote $\mathrm{var}_{A \in \mathbf{F}}\, \mu(A)$ the total variation of the restriction of the measure μ defined on some σ-algebra $\mathbf{N}$ to the σ-algebra $\mathbf{F}$. We call the corresponding values $\alpha(s), \beta(s)$ and $\varphi(s)$ the s.m., c.r. and u.s.m, coefficients, respectively.

Moreover, we will show that under certain conditions of weak dependence (more precisely, under strong mixing conditions) the rate of convergence of wavelet estimators is the same (up to a constant) as for the independent case. As we will see, for the estimators to attain the “independent” rates of convergence, we should require the stochastic process to satisfy some local regularity conditions.

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 1, 2 and 3. This is a Rosenthal-type inequality (see Doukhan, 1994; Rio, 1994, for other inequalities of this kind). We suppose that (ξ_i) is a strong mixing sequence of real random variables on the probability space $(\Omega, \aleph, P)$. In Lemma 1, let $\alpha(l)$ denote the strong mixing coefficient associated with (ξ_i).

Lemma 1.

(Leblance (1996)) Let $\infty > p \geq 2$ and $\xi_1, ..., \xi_n$ be a sequence of real-valued random variable such that $E(\xi_i) = 0,\ \| \xi_i \|_\infty < S,$ and $E\left(\xi_i^2\right) \leq \sigma^2$. Then there exists C such that:

$$E\left(| \sum_{i=1}^{n} \xi_i |^p \right) \leq C \left\{ \left(\frac{n}{l} \right)^{p/2} \sigma_l^p + \frac{n}{l} \sigma_l^2 (lS)^{p-2} + S^p n^p \alpha(l) \right\},$$

where $l \in N, 2 \leq l \leq n/2, \sigma_l^2 = \max\left\{ \max_{1 \leq u \leq n} \sigma_u^2(l), \max_{1 \leq u \leq n} \sigma_u^2(l-1) \right\}$ and $\sigma_u^2(l) = E(\sum_{i=u}^{u+l-1} \xi_i)^2$. In what follows, $\alpha(l)$ is the strong mixing coefficient defined in the introduction. We denote by E_f the mathematical expectation w.r.t, the law of the process and

$$\sigma_l^2 = \max_{1 \leq u \leq n-l+1} \max\left(\sigma_u^2(l), \sigma_u^2(l-1) \right), \quad \sigma_u^2(l) = E_f \left(\sum_{i=u}^{u+l-1} \left(Y_i - EY_i \right) \right)^2 .$$

Theorem 1.

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with bounded continuous density for X_1. Suppose that there exist constants $\alpha > 1$ and c_α such that for any l, $\alpha(l) \leq c_\alpha \alpha^{-1}$. Furthermore, suppose that there is a function g with $g(l) \geq G$ (G is a positive constant), such that for any $l = O(\ln(n)),\ \sigma_l^2 \leq \lg(l)$. Then for some $r > 1$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

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

Theorem 2.

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with bounded continuous density for X_1. Suppose that $\alpha(l) \leq c_\alpha l^{-\alpha}$, $\alpha \geq p(1+s)/s$ for any $l \in N, 2 \leq l \leq n/2$. Let us set $\mu = p(s+1)/\left[\alpha(1+2s)\right]$ and suppose that there is a function g with $g(l) \geq G$ (G is a positive constant), such that for any $l = O(n^\mu)$, $\sigma_l^2 \leq \lg(l)$. Then for some $r > 1$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

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

Theorems 1 and 2 are simple corollaries of the following result.

Proposition 1.

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

$$\sup_x P\left[|\bar{F}_n(x)-\bar{F}(x)|>\varepsilon\right]\le C\varepsilon^{-2r}\left\{n^{-r}\sigma_l^{2r}l^{-r}+n^{-2r+1}l^{2r-3}\sigma_l^2+\alpha(l)\right\}.$$

Proof.

By using Markov inequality, we get that for every $\varepsilon>0$,

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

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 $E\xi_i=0$. Hence applying the Lemma 1 we have

$$E|\sum_{i=1}^{n}(Y_i-EY_i)|^{2r}\le C\left\{\left(\frac{n}{l}\right)^r\sigma_l^{2r}+\frac{n}{l}\sigma_l^2l^{2r-2}+n^{2r}\alpha(l)\right\}. \qquad (2.2)$$

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

Proof of Theorem 1 and 2.

To obtain the results it is sufficient to balance the terms in the upper bound (2.1) by choosing the parameters.

Remark.

In the case of independent random variables, $\sigma_l^2=O(1)$. Moreover, in the dependent case a rough bound $\sigma_l^2=O(l^2)$ can be easily obtained. If some additional conditions are imposed on the process (X_i), the bound $\sigma_l^2=O(1)$ can be achieved (see Section 3). Let us consider the following condition:

$$C_\sigma \quad : \ \sigma_l^2=O(1).$$

When the condition C_σ is satisfied, the same rate as for the associated case is attained. We study C_σ in the next section.

Theorem 3.

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with bounded continuous density for X_1. If assumption C_σ is satisfied then for some $r > 1$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

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

Corollary 1.

Under the conditions of Theorem 3 for every x,

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

Proof.

For $r > 1$ observe that

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

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

Next we obtained a version of Glivenko-Cantelli Theorem valid for ρ-mixing random variables. The proof follows along the lines of analogous result for associated of random variables (Bagai and Prakasa Rao 1991) and using the results of last theorems.

Theorem 4.

Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables satisfying the conditions of Theorem 3. Then for any compact subset $J \subset R$,

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

3. DISCUSSION OF CONDITION $\mathbf{C}_\sigma$

We study the condition $\mathbf{C}_\sigma$ for some processes. Consider the following conditions:

M1: The process is ρ *-mixing* and $\sum_{t=1}^{\infty} \rho(t) < R < \infty$.

M2: The process is φ - mixing and $\sum_{t=1}^{\infty} \varphi(t) < \Phi < \infty$.

Comment.

Since the inequality $\rho(t) \leq 2\varphi(t)^{1/2}$ holds (see Doukhan, 1994), M2 implies M1. For Gaussian processes φ-mixing is equivalent to m-dependence (see Ibragimov and Linnik, 1971, Section 1), whereas ρ-*mixing* is equivalent to α-mixing (see Kolmogorov and Rozanov, 1960, Section 2.1). If the process $\{X_n,\ n \geq 1\}$ is ρ- *mixing*, we obtain:

Proposition 2.

Let $\{X_n,\ n \geq 1\}$ be a stochastic process on (R). Suppose that $\{X_n,\ n \geq 1\}$ admits a bounded marginal density which is common for all n. If assumption (M1) is satisfied then there exists a constant G such that for any $l, \sigma_l^2 \leq Gl$.

REFERENCES

1. Bagai, I. and Rao, P.B.L.S. (1991). Estimation of the survival function for stationary associated processes. *Statist. and Prob. Letters*, 12, 385-391.
2. Devydov, Y.A. (1973). Mixing conditions for Markov chains. *Theory Prob. Appl.*, 18, 312-328.
3. Doosti, H. and Zarei, H. (2006). Estimation of the survival function for negatively associated processes. *J. Sci. Islam. Repub. Iran.* 17(3), 249-251.
4. Doukhan, P. (1994). Mixing: properties and examples. *Lecture Notes in Statistics*, Vol. 85 Springer, New York.
5. Ibragimov, I.A. and Y.V. Linnik (1971). *Independent and Stationary Sequences of Random variables*. (Walter Nordhoof Pub., Groningen).
6. Kolmogorov, A.N. and Y.A. Rozanov (1960). On the strong mixing conditions for stationary Gaussian sequences. *Theory Prob. Appl.*, 5, 204-207.
7. Leblanc, F. (1996). Wavelet linear density estimator for a discrete-time stochastic process: Lp-losses. *Statist. and Prob. Letters*, 27, 71-84.
8. Serfling, R.J. (1970). Convergence properties of S_n under moment restriction. *Ann. Math. Statist.* 41, 1235-1248.
9. Shirazi, E. and Doosti, H. (2008). Estimation of the survival function for m-dependent random variables. *Far East Journal of Theoretical Statistics*, 25(1), 135-144.
10. Zarei, H. (2009). Survival Function Estimation for M-Dependent Processes. *Pak. J. Statist.*, 25(1), 1-4.