Estimation of the survival function for a discrete-time stochastic process

HASSAN DOOSTI
Ferdowsi University of Mashhad, Mashhad, Iran.

April 8, 2008

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.

[1]

Key Words: Survival function, Strongly mixing, 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

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

[1]Mathematics Subject Classification (2000):60F15

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 consistensy of it for sequence of associated random variables. Doosti and Zarei (2006) extended their results to negatively associated case.

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 $\mathcal{N}_k^m$ denote the σ-algebra generated by the events

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

We consider the following classical mixing conditions:

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

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

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

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

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

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

4. *ρ-mixing*

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

Following (Davydov, 1973) we denote $var_{A \in \mathcal{F}} \mu(A)$ the total variation of the restriction of the measure μ defined on some σ-algebra $\mathcal{N}$ to the σ-algebra $\mathcal{F}$. We call the corresponding values $\alpha(s), \beta(s)$ and $\phi(s)$ 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. 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 $E(\xi_i) = 0, \quad \|\xi_i\|_\infty < S$, and $E(\xi_i^2) \leq \sigma^2$. Then there exists C such that:

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

where

$$l \in N, 2 \leq l \leq n/2, \sigma_l^2 = \max\{\max_{1\leq u\leq n} \sigma_u^2(l), \max_{1\leq u\leq n} \sigma_u^2(l-1)\} \ \ and \ \ \sigma_u^2(l) = E(\sum_{i=u}^{u+l-1} \xi_i)^2.$$

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

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

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 $\varepsilon > 0$,

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

Theorem 2. Let $\{X_n, \ n \geq 1\}$ be a stationary sequence of random variables with bounded continuous density for X_1. Suppose that $\alpha(l) \leq c_\alpha l^{-\alpha}, \ \alpha \geq r$ for any $l \in N, 2 \leq l \leq n/2$. Let us set $0 < \mu < 1$ 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 $\varepsilon > 0$,

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

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

Proposition 1. Let $\{X_n,\ n \geq 1\}$ 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 $\varepsilon > 0$,

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

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

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

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

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

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, $\sigma_l^2 = O(l)$. 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 $\sigma_l^2 = O(l)$ can be achieved (see Proposition 2). Let us consider the following condition:

$\mathbf{C}_\sigma$: $\sigma_l^2 = O(l)$.

When the condition $\mathbf{C}_\sigma$ is satisfied, the same rate as for the associated case is attained.

Theorem 3. Let $\{X_n,\ n \geq 1\}$ be a stationary sequence of random variables with bounded continuous density for X_1. If assumption $\mathbf{C}_\sigma$ is satisfied then for some $r > 1$, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

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

Corollary 1. Under the conditions of Theorem 3 for every x,

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

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. $\square$

Next we ontaiend a version of Glivenko-Cantelli Threorem 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).

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 $J \subset R$,

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

Proof. Let K_1 and K_2 be chosen such that $J \subset [K_1, K_2]$ into b_n sub-intervals of length $\delta_n \longrightarrow 0$ 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 $I_{nj} = (x_{n,j}, x_{n,j+1}), \quad j = 1, ..., b_n = N$, where

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

with

$x_{n,j+1} - x_{n,j} \leq \delta_n \qquad$ for $1 \leq j \leq N$.

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

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

and

$$\bar{F}_n(x_{n,j+1}) \leq \bar{F}_n(x) \leq \bar{F}_n(x_{n,j}).$$

Hence

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

$$\leq \bar{F}_n(x) - \bar{F}(x) \leq [\bar{F}_n(x_{n,j}) - \bar{F}(x_{n,j})] + [\bar{F}(x_{n,j}) - \bar{F}(x)].$$

Therefore

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

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

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

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

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

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 $n \leq n(\varepsilon)$,

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

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

2.1 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 $\sum_{t=1}^{\infty} \rho(t) \leq R < \infty$.
M2 : The process is ϕ-mixing and $\sum_{t=1}^{\infty} \phi(t) \leq \Phi < \infty$.
Comment. Since the inequality $\rho(t) \leq 2(\phi(t))^{1/2}$ 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) is ρ-mixing, we obtain:
Proposition 2. Let $(X_n, n \geq 1)$ be a stochastic process on (R). Suppose that X_n 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, $\sigma_l^2 \leq Gl$.

Proof. We use the decomposition

$$\begin{aligned}
\sigma_u^2(l) &= E(\sum_{i=u}^{u+l-1} (Y_i - EY_i))^2 \\
&\leq \sum_{i=u}^{u+l-1} E(Y_i - EY_i)^2 + \sum_{u \leq m < t < l+u-1} |Cov(Y_m, Y_t)| \\
&= T_1 + T_2. \qquad (2.6)
\end{aligned}$$

The first term in above can be estimated as follows

$$T_1 \leq 4l. \qquad (2.7)$$

To bound the term T_2 we apply a ρ-mixing covariance inequality (see Doukhan (1994), section 1.2.2.),*i.e.*,

$$|Cov(Y_m, Y_t)| \leq 2\rho(t-m)(EY_m^2)^{1/2}(EY_t^2)^{1/2} \leq 2\rho(t-m).$$

We obtain

$$T_2 \leq 2 \sum_{m=u}^{l+u-2} \sum_{t=m+1}^{l+u-1} \rho(t-m) \leq 2.Rl \qquad (2.8)$$

By substituting the bounds (2.7) and (2.8) in (2.6), proposition will be proved.

References

[1] Bagai, I. and Prakasa Rao, B.L.S. (1991). Estimation of the survival function for stationary associated processes *Statistics and Probability Letters,***12**, 385-391.

[2] Devydov, Y.A. (1973). Mixing conditions for Markov chains, *Theory Probab. 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 Probab. Appl.*, **5** 204-207.

[7] Leblanc, F. (1996). Wavelet linear density estimator for a discrete-time stochastic process:L_p-losses. *Statistics and Probability Letters,* **27**, 71-84.