Statistics & Probability Letters 12 (1991) 385–391
North-Holland

November 1991

Estimation of the survival function for stationary associated processes

Isha Bagai
Department of Statistics, Panjab University, Chandigarh 160 014, India

B.L.S. Prakasa Rao
Indian Statistical Institute, New Delhi 110 016, India

Abstract: Let $\{X_n, n \geqslant 1\}$ be a stationary sequence of associated random variables with survival function $\bar{F}(x) = P[X_1 > x]$. The empirical survival function $\bar{F}_n(x)$ based on $X_1, X_2, \ldots, X_n$ is proposed as an estimator for $\bar{F}(x)$. Strong consistency, pointwise as well as uniform, and asymptotic normality of $\bar{F}_n(x)$ are discussed.

AMS 1980 Subject Classifications: 62G05, 62M09.

Keywords: Survival function, associated sequences, uniform strong consistency, asymptotic normality.

1. Introduction

Let $\{X_n, n \geqslant 1\}$ be a stationary sequence of associated random variables with distribution function $F(x)$, or equivalently, survival function $\bar{F}(X) = 1 - F(x)$. We say that $X_1, X_2, \ldots, X_n$ are associated if for every pair of functions $h(x)$ and $g(\boldsymbol{x})$ from $\mathbb{R}^n$ to $\mathbb{R}$, which are non-decreasing componentwise, the following holds:

$$\operatorname{Cov}(h(\boldsymbol{X}), g(\boldsymbol{X})) \geqslant 0$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. An infinite family $\{X_n, n \geqslant 1\}$ is associated if every finite sub-family is associated. We assume, of course, that the covariances involved exist.

The concept of association for random variables was introduced by Esary, Proschan and Walkup (1967). It is very useful in reliability situations where the random variables of interest are very often not independent but are associated. The asymptotic properties of associated random variables have been discussed by Newman (1980, 1984) and Birkel (1988a,b) among others. They observed that in any asymptotic study of associated random variables the covariance structure plays an important role. We impose conditions on the covariance structure of the associated sequence of random variables along the lines of Birkel (1988a,b).

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

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

where

$$Y_j(x) = \begin{cases} 1 & \text{if } X_j > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We propose $\bar{F}_n(x)$ as an estimator for $\bar{F}(x)$ and study its asymptotic properties.

Esary, Proschan and Walkup (1967) showed that nondecreasing functions of associated random variables are associated. Therefore, for fixed x, $Y_1, \ldots, Y_n$ are associated.

In what follows 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 (see, for example, Bagai and Prakasa Rao, 1990). Asymptotic normality of $\bar{F}_n(x)$ is also discussed. Some lemmas, useful in proving the results concerning $\bar{F}_n(x)$, are stated and proved in Section 2.

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

2. Some lemmas

Lemma 2.1 (Sadikova, 1966). *Let $F(x, y)$ and $G(x, y)$ be two bivariate distribution functions, with characteristic functions $f(s, t)$ and $g(s, t)$ respectively. Define*

$$\hat{f}(s, t) = f(s, t) - f(s, 0) f(0, t)$$

and

$$\hat{g}(s, t) = g(s, t) - g(s, 0) g(0, t).$$

Suppose that partial derivatives of G with respect to x and y exist. Let

$$A_1 = \sup_{x,y} \frac{\partial G(x, y)}{\partial x} \quad \text{and} \quad A_2 = \sup_{x,y} \frac{\partial G(x, y)}{\partial y}.$$

Suppose A_1 and A_2 are finite. Then, for any $T > 0$,

$$\begin{aligned} \sup_{x,y} |F(x, y) - G(x, y)| \leqslant{} & \frac{2}{(2\pi)^2} \int_{-T}^{T} \int_{-T}^{T} \left| \frac{\hat{f}(s, t) - \hat{g}(s, t)}{st} \right| \mathrm{d}s \, \mathrm{d}t \\ & + 2 \sup_x |F(x, \infty) - G(x, \infty)| + 2 \sup_y |F(\infty, y) - G(\infty, y)| \\ & + 2 \frac{A_1 + A_2}{T} (3\sqrt{2} + 4\sqrt{3}). \qquad \square \end{aligned} \tag{3}$$

Lemma 2.2. *Suppose X and Y are associated random variables with bounded continuous densities. Then, there exists a constant $C > 0$, such that for any $T > 0$,*

$$\sup_{x,y} |P[X \leqslant x, Y \leqslant y] - P[X \leqslant x] P[Y \leqslant y]| \leqslant C \left\{ T^2 \operatorname{Cov}(X, Y) + \frac{1}{T} \right\}. \tag{4}$$

Proof. Let

$$F(x, y) = P[X \leqslant x, Y \leqslant y]$$

and

$$G(x, y) = P[X \leqslant x] P[Y \leqslant y]$$

in (3). It is easy to see that the function $G(x, y)$ satisfies the conditions stated in Lemma 2.1. Then (4) follows from Lemma 2.1 and the following inequality by Newman (1980); for all real s and t,

$$|f(s, t) - f(s, 0)f(0, t)| \leqslant |t||s|\operatorname{Cov}(X, Y). \quad \square$$

Lemma 2.3 (Birkel, 1988a). *Let* $\{X_n, n \geqslant 1\}$ *be a sequence of associated random variables with mean zero and* $\sup_n |X_n| < \infty$. *Let* $S_n = \sum_{j=1}^n X_j$ *and*

$$u(n) = \sup_{k \geqslant 1} \sum_{j:\,|j-k| \geqslant n} \operatorname{Cov}(X_j, X_k).$$

Assume that $u(n) = O(n^{-(r-2)/2})$ *for some* $r > 2$. *Then, there exists a constant* $C > 0$, *not depending on* n, *such that*

$$E\left[|S_n|^r\right] \leqslant C n^{r/2}$$

for all $n \geqslant 1$. $\square$

The above lemma can easily be generalized to obtain the following result by methods in Birkel (1988a)

Lemma 2.4. *For every* $\alpha \in I$, *an index set, let* $\{X_n(\alpha), n \geqslant 1\}$ *be an associated sequence with* $EX_n(\alpha) = 0$ *and*

$$\sup_{\alpha \in I} \sup_{n \geqslant 1} |X_n(\alpha)| \leqslant A < \infty.$$

Let

$$S_n(\alpha) = \sum_{j=1}^n X_j(\alpha)$$

and

$$u(n, \alpha) = \sup_{k \geqslant 1} \sum_{j:\,|j-k| \geqslant n} \operatorname{Cov}\big(X_j(\alpha), X_k(\alpha)\big).$$

Suppose that there exists $b > 0$, *independent of* $\alpha \in I$ *and* $n \geqslant 1$, *such that for some* $r > 2$, *and all* $\alpha \in I$ *and* $n \geqslant 1$,

$$u(n, \alpha) \leqslant b n^{-(r-2)/2}.$$

Then, there exists a constant C, *not depending on n and* α, *such that for all* $n \geqslant 1$,

$$\sup_{\alpha \in I} \sup_{m \geqslant 0} E|S_{n+m}(\alpha) - S_n(\alpha)|^r \leqslant C n^{r/2}. \quad \square$$

Lemma 2.5 (Newman, 1980). *Let* $\{X_n, n \geqslant 1\}$ *be a stationary associated sequence of random variables with* $E[X_1^2] < \infty$ *and* $0 < \sigma^2 = \operatorname{Var}(X_1) + 2\sum_{j=2}^\infty \operatorname{Cov}(X_1, X_j) < \infty$. *Then,* $n^{-1/2}(S_n - E(S_n))$ *converges in distribution to* $N(0, \sigma^2)$ *as* $n \to \infty$. $\square$

3. The empirical survival function

Theorem 3.1. *Let* $\{X_n, n \geqslant 1\}$ *be a stationary sequence of associated random variables with bounded continuous density for* X_1. *Assume that, for some* $r > 1$,

$$\sum_{j=n+1}^{\infty} \{\operatorname{Cov}(X_1, X_j)\}^{1/3} = O(n^{-(r-1)}).$$

Then, there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

$$\sup_x P\left[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon\right] \leqslant C\varepsilon^{-2r} n^{-r} \quad \text{for every } n \geqslant 1.$$

Proof. Observe that,

$$\begin{aligned}\operatorname{Cov}(Y_1(x),\ Y_j(x)) &= P[X_1 > x,\ X_j > x] - P[X_1 > x]P[X_j > x] \\ &= P[X_1 \leqslant x,\ X_j \leqslant x] - P[X_1 \leqslant x]P[X_j \leqslant x],\end{aligned}$$

which is non-negative since $Y_1(x)$ and $Y_j(x)$ are associated. Then there exists a constant $C > 0$ such that

$$\begin{aligned}\sum_{j=n+1}^{\infty} \operatorname{Cov}(Y_1(x),\ Y_j(x)) &\leqslant \sum_{j=n+1}^{\infty} \sup_x \left\{ P[X_1 \leqslant x,\ X_j \leqslant x] - P[X_1 \leqslant x]P[X_j \leqslant x] \right\} \\ &\leqslant C \sum_{j=n+1}^{\infty} \left\{\operatorname{Cov}(X_1,\ X_j)\right\}^{1/3},\end{aligned}$$

by taking $T = \{\operatorname{Cov}(X_1,\ X_j)\}^{-1/3}$ in Lemma 2.2 whenever $\operatorname{Cov}(X_1,\ X_j) > 0$ and if $\operatorname{Cov}(X_1,\ X_j) = 0$ then X_1 and X_j are independent as they are associated and $\operatorname{Cov}(Y_1(x),\ Y_j(x)) = P[X_1 \leqslant x,\ X_j \leqslant x] - P[X_1 \leqslant x]P[X_j \leqslant x] = 0 \leqslant [\operatorname{Cov}(X_1,\ X_j)]^{1/3}$. Furthermore

$$\sup_x \sup_j |Y_j(x) - EY_j(x)| \leqslant 2$$

and

$$u(n,\ x) = 2 \sum_{j=n+1}^{\infty} \operatorname{Cov}(Y_1(x),\ Y_j(x)) \leqslant C \sum_{j=n+1}^{\infty} \left\{\operatorname{Cov}(X_1,\ X_j)\right\}^{1/3}$$

for all real x, where C is independent of n and x. From Lemma 2.4, it follows that, for every $n \geqslant 1$,

$$\sup_x E\left|\sum_{j=1}^{n} (Y_j(x) - EY_j(x))\right|^{2r} \leqslant Cn^r$$

where C is independent of n and x. Then, by using Markov Inequality, we get that for every $\varepsilon > 0$,

$$\begin{aligned}\sup_x P\left[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon\right] &= \sup_x P\left[|\bar{F}_n(x) - \bar{F}(x)|^{2r} > \varepsilon^{2r}\right] \\ &\leqslant \sup_x \left\{ n^{-2r}\varepsilon^{-2r} E\left[\left|\sum_{j=1}^{n} (Y_j(x) - EY_j(x))\right|^{2r}\right]\right\} \\ &\leqslant C\varepsilon^{-2r} n^{-r}. \quad \square\end{aligned}$$

Corollary 3.1. *Under the conditions of Theorem 3.1, for every x,*

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

Proof. Observe that

$$\sum_{n=1}^{\infty} P\left[|\bar{F}_n(x) - \bar{F}(x)| > \varepsilon\right] \leqslant C\varepsilon^{-2r} \sum_{n=1}^{\infty} \frac{1}{n^r} < \infty \quad \text{for } r > 1.$$

The result then follows by using the Borel–Cantelli Lemma. □

Remark 3.1. Corollary 3.1. is valid under the weaker condition

$$\frac{1}{n}\sum_{j=1}^{n} \operatorname{Cov}(X_1, X_j) \to 0 \quad \text{as } n \to \infty$$

for any stationary associated sequence as pointed out by the referee. This result is a consequence of Theorem 7 in Newman (1984).

Next we obtain a version of Glivenko–Cantelli Theorem valid for associated random variables. The proof follows along the lines of the proof of an analogous result for mixing sequences of random variables (Roussas, 1989).

Theorem 3.2. *Let $\{X_n, n \geqslant 1\}$ be a stationary sequence of associated random variables satisfying the conditions of Theorem 3.1. Then for any compact subset $J \subset \mathbb{R}$,*

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

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

$$\sum_n \delta_n^{-1} n^{-r} < \infty. \tag{5}$$

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

$$b_n \leqslant C\delta_n^{-1}. \tag{6}$$

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

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

with

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

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

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

and

$$\bar{F}_n(x_{n,j+1}) \leqslant \bar{F}_n(x) \leqslant \bar{F}_n(x_{n,j}). \tag{7}$$

Hence

$$\begin{aligned} &\left[\bar{F}_n(x_{n,j+1}) - \bar{F}(x_{n,j+1})\right] + \left[\bar{F}(x_{n,j+1}) - \bar{F}(x)\right] \\ &\quad \leqslant \bar{F}_n(x) - \bar{F}(x) \leqslant \left[\bar{F}_n(x_{n,j}) - \bar{F}(x_{n,j})\right] + \left[\bar{F}(x_{n,j}) - \bar{F}(x)\right]. \end{aligned} \tag{8}$$

Therefore

$$\begin{aligned} &\sup\left[|\bar{F}_n(x) - \bar{F}(x)| : x \in J \right] \\ &\quad \leqslant \sup\left[|\bar{F}_n(x) - \bar{F}(x)| : K_1 \leqslant x \leqslant K_2 \right] \\ &\quad \leqslant \max_{1 \leqslant j \leqslant N} \left|\bar{F}_n(x_{n,j}) - \bar{F}(x_{n,j})\right| + \max_{1 \leqslant j \leqslant N} \left|\bar{F}_n(x_{n,j+1}) - \bar{F}(x_{n,j+1})\right| \\ &\qquad + \max_{1 \leqslant j \leqslant N} \sup_{x \in I_{nj}} \left|\bar{F}(x_{n,j}) - \bar{F}(x)\right| + \max_{1 \leqslant j \leqslant N} \sup_{x \in I_{nj}} \left|\bar{F}(x_{n,j+1}) - \bar{F}(x)\right|. \end{aligned} \tag{9}$$

Now

$$\bar{F}(x_{n,j}) - \bar{F}(x) = F(x) - F(x_{n,j})$$
$$= (x - x_{n,j}) f(u^*) \quad \text{for } x_{n,j} < u^* < x \tag{10}$$

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

$$\left|\bar{F}(x_{n,j}) - \bar{F}(x)\right| \leqslant C\delta_n, \qquad \left|\bar{F}(x_{n,j+1}) - \bar{F}(x)\right| \leqslant C\delta_n, \tag{11}$$

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

$$2C\delta_n < \tfrac{1}{3}\varepsilon.$$

From (9) and (10), we get, for $n \geqslant n(\varepsilon)$,

$$P\left[\sup_{x \in J} \left|\bar{F}_n(x) - \bar{F}(x)\right| > \varepsilon\right]$$
$$\leqslant P\left[\max_{1 \leqslant j \leqslant N} \left|\bar{F}_n(x_{n,j}) - \bar{F}(x_{n,j})\right| > \tfrac{1}{3}\varepsilon\right] + P\left[\max_{1 \leqslant j \leqslant N} \left|\bar{F}_n(x_{n,j+1}) - \bar{F}(x_{n,j+1})\right| > \tfrac{1}{3}\varepsilon\right]$$
$$\leqslant \sum_{j=1}^{N} \left\{P\left[\left|\bar{F}_n(x_{n,j+1}) - \bar{F}(x_{n,j})\right| > \tfrac{1}{3}\varepsilon\right] + P\left[\left|\bar{F}_n(x_{n,j+1}) - \bar{F}(x_{n,j})\right| > \tfrac{1}{3}\varepsilon\right]\right\}$$
$$\leqslant CN\varepsilon^{-2r} n^{-r} = C\varepsilon^{-2r} b_n n^{-r} \quad \text{(by Theorem 3.1)}$$
$$\leqslant C\varepsilon^{-2r}\delta_n^{-1} n^{-r}.$$

The result follows by using (5) and the Borel–Cantelli Lemma. □

Theorem 3.3. *Let $\{X_n, n \geqslant 1\}$ be a stationary associated sequence of random variables with bounded continuous density for X_1 and survival function $\bar{F}(x)$. Suppose that*

$$\sum_{j=2}^{\infty} \left\{\operatorname{Cov}(X_1, X_j)\right\}^{1/3} < \infty. \tag{12}$$

Define

$$\sigma^2(x) = \bar{F}(x)\left[1 - \bar{F}(x)\right] + 2\sum_{j=2}^{\infty} \left\{P\left[X_1 > x, X_j > x\right] - \bar{F}^2(x)\right\}.$$

Then, for all x such that $0 < F(x) < 1$, $n^{1/2}\,[\bar{F}_n(x) - \bar{F}(x)]/\sigma(x)$ converges in distribution to a standard normal variable as $n \to \infty$.

Proof. Observe that $n\bar{F}_n(x) = \sum_{i=1}^{n} Y_i(x)$, with $0 < \operatorname{Var} Y_1(x) = \bar{F}(x)(1 - \bar{F}(x)) < 1$, and $\operatorname{Cov}(Y_1(x), Y_j(x)) \geqslant 0$ by association. Using Lemma 2.2, we get that

$$0 < \operatorname{Var} Y_1(x) + 2\sum_{j=1}^{\infty} \operatorname{Cov}(Y_1(x), Y_j(x))$$
$$\leqslant \bar{F}(x)(1 - \bar{F}(x)) + 2\sum_{j=2}^{\infty} \left\{\operatorname{Cov}(X_1, X_j)\right\}^{1/3} < \infty$$

by arguments similar to those given in the proof of Theorem 3.1. The result now follows from Lemma 2.5 due to Newman (1980). □

Remark 3.2. Theorem 3.3 can be extended to an invariance principle by using Theorem 2.2 of Newman (1984).

Acknowledgement

The authors thank the referee for the detailed comments and suggestions.

References

Bagai, I. and B.L.S. Prakasa Rao (1990), Kernel-type density and failure rate estimation for associated sequences, in preparation.

Birkel, T. (1988a), Moment bounds for associated sequences, *Ann. Probab.* **16**, 1184–1193.

Birkel, T. (1988b), On the convergence rate in the central limit theorem for associated processes, *Ann. Probab.* **16**, 1685–1698.

Esary, J., F. Proschan and D. Walkup (1967), Association of random variables with applications, *Ann. Math. Statist.* **38**, 1466–1474.

Newman, C.M. (1980), Normal fluctuations and the FKG inequalities, *Comm. Math. Phys.* **74**, 119–128.

Newman, C.M. (1984), Asymptotic independence and limit theorems for positively and negatively dependent random variables, in: Y.L. Tong, ed., *Inequalities in Statistics and Probability* (IMS, Hayward, CA) pp. 127–140.

Roussas, G.G. (1989), Hazard rate estimation under dependence conditions, *J. Statist. Plann. Inference* **22**, 81–93.

Sadikova, S.M. (1966), Two dimensional analogues of an inequality of Essen with applications to the central limit theorem, *Theory Probab. Appl.* **11**, 325–335.