The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 229-235

On the Estimation of the Empirical Distribution Function for Negatively Associated Processes[1]

Tae Sung Kim[2], Seung Woo Lee[3] and Mi Hwa Ko[4]

Abstract

Let $\{X_n,\ n\geq 1\}$ be a stationary sequence of negatively associated random variables with distribution function $F(x)=P(X_1\leq x)$. The empirical distribution function $F_n(x)$ based on $X_1, X_2, \ldots, X_n$ is proposed as an estimator for $F_n(x)$. Strong consistency and asymptotic normality of $F_n(x)$ are studied. We also apply these ideas to estimation of the survival function.

Keywords : Empirical distribution function, negatively associated sequences, strong consistency, asymptotic normality, survival function.

1. Introduction

A finite family $\{X_1,\ X_2,\ \ldots\ ,\ X_n\}$ of random variables is said to be associated if

$$Cov(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \leq 0, \tag{1}$$

for all real coordinatewise nondecreasing functions f and g on R^n, such that the covariance exists. It is said to be negatively associated if for every pair of disjoint subsets A and B of {1, 2, ... ,n}

$$Cov(f(X_i, i\in A), g(X_j, j\in B)) \geq 0, \tag{2}$$

whenever f and g are coordinatewise nondecreasing and such that the covariance exists. An infinite family of random vriables is associated(negatively associated) if every finite subfamily

1) This paper was supported by WonKwang University in 2000
2) Professor, Division of Mathematics and Informational Statistics, WonKwang University, Iksan, Chonbuk 570-749, Korea.
E-mail : starkim@wonkwang.ac.kr
3) Professor, Division of Mathematics and Informational Statistics, WonKwang University, Iksan, Chonbuk 570-749, Korea
4) Lecturer, Division of Mathematics and Informational Statistics, WonKwang University, Iksan, Chonbuk 570-749, Korea

is associated(negatively associated). The concept of association was introduced by Esary, Proschan and Walkup(1967) and the definition of negative association is introduced by Alam and Saxena(1981) and carefully studied by Joag-Dev and Proschan(1983) and Block et al.(1982). As pointed out and proved by Joag-Dev and Proschan(1983), a number of well known multivariate distributions possess the negative association property, such as (a) multinomial, (b) convolution of unlike multinomials, (c) multivariate hypergeometric, (d) Dirichlet, (e) Dirichlet compound multinomial, (f) negatively correlated normal distribution, (g) permutation distribution, (h) random sampling without replacement, and (i) joint distribution of ranks. Because of its wide applications in multivariate statistical analysis and reliability, the notion of negative association have received considerable attenion recently. We refer to Joag-Dev and Proschan(1983) for fundamental properties, Newman(1984) and Su and Chi(1988) for the central limit theorem, Matula(1992) for three series theorem, Su et al.(1997) for a moment inequality, a weak invariance principle and an example to show that there exists infinite family of non-degenerate non-independent strictly stationary negative association random variables, Shao(1998) for convergence rates of law of the iterated logarithm, Roussas(1994) for the central limit theorem of random fields, some examples and applications.

Let $F_n(x)$ be defined by

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} Y_j(x) \tag{3}$$

where

$$Y_j(x) = \begin{cases} 1 & \text{if } X_j \le x, \\ 0 & \textit{otherwise}, \end{cases} \tag{4}$$

and proposed as an estimator for distribution function $F(x) = P(X_1 \le x)$.

In this note we discuss, for a stationary sequence $\{X_n, n \ge 1\}$ of negatively associated random variables with distribution function $F(x) = P(X_1 \le x)$, the strong consistency, pointwise and uniform of $F_n(x)$. Asymptotic normality of $F_n(x)$ is discussed. We also apply these ideas to the survival function.

2. Preliminaries

Lemma 2.1 Let $r \ge 2$ and let $\{X_n, n \ge 1\}$ be a stationary negatively associated sequence of random variables with $EX_1 = 0$ and $E|X_1|^r < \infty$. Then, there exist constant $A_r > 0$ and $B_r > 0$ such that

$$E|S_n|^r \le A_r n^{\frac{r}{2}}, \tag{5}$$

$$E(\max_{1 \le k \le n} |S_k|^r) \le B_r n^{\frac{r}{2}}, \tag{6}$$

where $S_n = X_1 + \ldots + X_n$.

Proof. From results of Su et al. (1997) and stationarity we obtain, for $r \ge 2$,

$$E|S_n|^r \le C_r\{(\sum_{i=1}^{n} E X_i^2)^{\frac{r}{2}} + \sum_{i=1}^{n} E|X_i|^r\}$$

$$\le C_r\{(nE X_1^2)^{\frac{r}{2}} + nE|X_1|^r\}$$

by Hollder's inequality

$$\le C_r\{(n^{\frac{r}{2}} E|X_1|^r) + n^{\frac{r}{2}} E|X_1|^r\}$$

$$\le A_r n^{\frac{r}{2}}$$

and

$$E(\max_{1\le k\le n}|S_k|^r) \le D_r\{(\sum_{i=1}^{n} E X_i^2)^{\frac{r}{2}} + \sum_{i=1}^{n} E|X_i|^r\}$$

$$\le D_r\{(\sum_{i=1}^{n} E X_1^2)^{\frac{r}{2}} + nE|X_1|^r\}$$

$$\le D_r\{(n^{\frac{r}{2}} E|X_1|^r) + n^{\frac{r}{2}} E|X_1|^r\}$$

$$\le B_r n^{\frac{r}{2}} \quad \textit{by Hollder's inequality.}$$

The above lemma can be easily be generalized to obtain the following result by methods in Su et al.(1997).

Lemma 2.2 For every $\alpha \in I$, an index set, let $\{X_n(\alpha), n \ge 1\}$ be a stationary sequence of negatively associated random variables with $EX_1(\alpha) = 0$ and $E|X_1(\alpha)|^r < \infty$ for $r \ge 2$. Then, there exists a constant C such that, for all $n \ge 1$,

$$sup_{\alpha \in I} sup_{n \ge 1} E|S_n(\alpha)|^r \le C n^{\frac{r}{2}} \tag{7}$$

for some $r \ge 2$.

Lemma 2.3 (Newman, 1984) Let $\{X_n, n \ge 1\}$ be a stationary sequence of negatively associated random variables with $EX_1 = 0, E(X_1^2) < \infty$ and

$$0 < \sigma^2 = Var(X_1) + 2\sum_{j=2}^{\infty} Cov(X_1, X_j). \tag{8}$$

Then, $n^{-\frac{1}{2}} S_n$ converges in distribution to $N(0, \sigma^2)$ as $n \to \infty$, that is,

$$n^{-\frac{1}{2}} S_n \to \sigma Z, \ as \ \ n \to \infty, \tag{9}$$

where Z is a standard normal random variable.
Proof. See the proof of Theorem 12 of Newman (1984).

3. The empirical distribution function

Theorem 3.1 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables. Then, there exists a constant C>0 such stat, for every $\varepsilon > 0$,

$$supP[|F_n(x) - F(x)| > \varepsilon] \leq C\varepsilon^{-2r} n^{-r} \quad \text{for } some \ \ r \geq 1.$$

Proof. First note that $[Y_j(x) - EY_j(x)]'s$ are negatively associated according to Joag-Dev and Proschan(1983). Clearly,

$$E(Y_j(x) - EY_j(x)) = 0,$$

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

and

$$E|Y_j(x) - EY_j(x)|^{2r} < \infty \quad \text{for} \ \ r > 1.$$

Hence, from Lemma 2.2 it follows that for some $r > 1$

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

Then, by using Markov Inequality, we get that for every $\varepsilon > 0$ and some $r > 1$

$$\begin{aligned} &sup_x P[|F_n(x) - F(x)| > \varepsilon] \\ &= sup_x P\frac{1}{n}\left|\sum_{j=1}^{n} (Y_j(x) - EY_j(x))\right| > \varepsilon \\ &\leq sup_x n^{-2r} \varepsilon^{-2r} E\left|\sum_{j=1}^{n} (Y_j(x) - EY_j(x))\right|^{2r} \\ &\leq C\varepsilon^{-2r} n^{-r}. \end{aligned} \tag{10}$$

Corollary 3.2 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables. Then, for every x,

$$F_n(x) \to F(x) \quad a.s. \quad as \ n \to \infty.$$

Proof. According to (10) we observe that, for $r > 1$,

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

Thus the desired result follows by Borel-Cantelli lemma.

Remark Corollary 3.2 is valid under the weaker condition finite second moments for any stationary negatively associated sequence. This result is a consequence of Corollary of

Matula(1992).

Theorem 3.3 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables with $Var(X_1) < \infty$ and distribution function $F(x)$. Define

$$\sigma^2(x) = F(x)[1-F(x)] + 2\sum_{j=2}^{\infty}\{P[X_1 \leq x, X_j \leq x] - F^2(x)\}. \tag{11}$$

If $\sigma^2(x) > 0$ then, for all x, such that $0 < F(x) < 1$,

$$n^{\frac{1}{2}}[F_n(x) - F(x)]/\sigma(x) \rightarrow Z, \tag{12}$$

as $n \rightarrow \infty$, where Z is a standard normal variable.

Proof. First note that

$$nF_n(x) = \sum_{i=1}^{n} Y_i(x),$$

$$0 < Var(Y_1) = F(x)(1-F(x)) < 1,$$

and

$$Cov(Y_1(x), Y_j(x)) = P[X_1 \leq x, X_j \leq x] - P[X_1 \leq x]P[X_j \leq x].$$

Then it follows from Lemma 8 of Newman(1984) that

$$\begin{aligned} \sigma^2(x) &= F(x)[1-F(x)] + 2\sum_{j=2}^{\infty}\{P[X_1 \leq x, X_j \leq x] - F^2(x)\} \\ &= VarY_1(x) + 2\sum_{j=2}^{\infty} Cov(Y_1(x), Y_j(x)) \\ &\leq VarY_1(x) = F(x)(1-F(x)) < 1. \end{aligned} \tag{13}$$

Note that

$$\begin{aligned} & n^{\frac{1}{2}}[F_n(x) - F(x)]/\sigma(x) \\ &= n^{\frac{1}{2}}[\frac{1}{n}\sum_{i=1}^{n} Y_i(x) - EY_1(x)]/\sigma(x) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}(Y_i(x) - EY_i(x))/\sigma(x). \end{aligned} \tag{14}$$

Thus, from (13), (14) and Lemma 2.3 the desired result (12) follows.

Remark. Theorem 3.3 can be extended to an invariance principle by using the results of Su, Zhao, and Wang(1997).

4. Applications

We apply the ideas in section 3 to the empirical survival function.
Let $\overline{F_n}(x)$ be defined by

$$\overline{F_n}(x) = \frac{1}{n} \sum_{j=1}^{n} Z_j(x), \tag{15}$$

where

$$Z_j(x) = \begin{cases} 1 & \text{if} \quad X_j > x, \\ 0 & \textit{otherwise}, \end{cases} \tag{16}$$

and proposed as an estimator for $\overline{F}(x) = P(X_1 > x)$.
Note that $Z_j(x) = 1 - Y_j(x)$. Hence,

$$\begin{aligned} \overline{F_n}(x) = \frac{1}{n} \sum_{j=1}^{n} Z_j(x) &= \frac{1}{n} \sum_{j=1}^{n} (1 - Y_j(x)) \\ &= 1 - F_n(x). \end{aligned}$$

Moreover, $\overline{F}(x) = 1 - F(x)$.
Thus Theorem 3.1, Corollary 3.2 and Theorem 3.3 imply Theorem 4.1, Corollary 4.2 and Theorem 4.3, respectively.

Theorem 4.1 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables. Then, there exist a constant $C > 0$ such stat, for every $\varepsilon > 0$,

$$supP[|\overline{F_n}(x) - \overline{F}(x)| > \varepsilon] \leq C\varepsilon^{-2r} n^{-r} \quad \text{for } \textit{some} \quad r \geq 1.$$

Corollary 4.2 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables. Then, for every x,

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

Theorem 4.3 Let $\{X_n, n \geq 1\}$ be a stationary sequence of negatively associated random variables with $Var(X_1) < \infty$ and survival function $\overline{F}(x)$. Define

$$\sigma^2(x) = \overline{F}(x)[1 - \overline{F}(x)] + 2 \sum_{j=2}^{\infty} \{P[X_1 > x,\ X_j > x] - \overline{F^2}(x)\}. \tag{17}$$

If $\sigma_2{}^2(x) > 0$ then, for all x, such that $0 < F(x) < 1$,

$$n^{\frac{1}{2}}[\overline{F_n}(x) - \overline{F}(x)]/\sigma(x) \to Z, \tag{18}$$

as $n \to \infty$, where Z is a standard normal variable.

Remark Theorem 3.1, Corollary 3.2, and Theorem 3.3 hold if and only if Theorem 4.1, Corollary 4.2, and Theorem 4.3 hold, respectively.

References

[1] Alam, K., Saxena, K. M. L. (1981) Positive dependence in multivariate distributions, Communication in Statistics Theory and Method A 10 1183-1196.

[2] Block, H. W., Savits, T. H., Shaked, M. (1982) Some concepts of negative dependence, Annals of Probability 10 765-772.

[3] Esary, J., Proschan F. and Walkup, D. (1967) Association of random variables with applications , Annals of Mathematical Statistics 38 1466-1474.

[4] Joag-Dev, K., Proschan, F. (1982) Negative association of random variables, Annals of Statistics 11 286-295.

[5] Matula, P. (1992) A note on the almost sure convergence of sums of negatively dependent random variables, Statistics & Probability Letters 15 209-213.

[6] Newman, C. M. (1984) Asymptotic independence and limit theorems for positively and negatively dependent random variables, In : Tong, Y. L.(Ed.) Institute of Mathematical Statistics Hayward, CA. 127-140.

[7] Roussas, G. G. (1994) Asymptotic normality of random fields of positively or negatively associated processes, Journal of Multivariate Analysis 50 152-173.

[8] Su, C., Chi, X. (1998) Some results on CLT for nonstationary NA sequences, Acta. Mathematica Applicandae Sinica 21, 9-21(in Chinese)

[9] Su, C., Zhao, L. C., Wang, Y. B. (1997) Moment inequalties and weak convergence for negatively associated sequences, (Ser. A). Sciences in China 40 172-182.