

**ESTIMATION OF THE DISTRIBUTION
FUNCTION FOR STATIONARY RANDOM
FIELDS OF ASSOCIATED PROCESSES**

TAE-SUNG KIM[†], MI-HWA KO[‡] AND YEON-SUN YOO

ABSTRACT. For a stationary field $\{X_j, j \in \mathbb{Z}_+^d\}$ of associated random variables with distribution function $F(x) = P(X_1 \leq x)$ we study strong consistency and asymptotic normality of the empirical distribution function, which is proposed as an estimator for $F(x)$. We also consider strong consistency and asymptotic normality of the empirical survival function by applying these results.

1. Introduction and notation

A finite family $\{X_i, 1 \leq i \leq n\}$ of random variables is said to be associated if for all coordinatewise nondecreasing functions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$

$$(1) \quad \text{Cov}(f(X_1, \dots, X_n), g(X_1, \dots, X_n)) \geq 0,$$

where the covariance is defined. An infinite family $\{X_i, i \geq 1\}$ is associated if every finite subfamily is associated. The concept of association for random variables was introduced by Esary, Proschan and Walkup ([7]).

Associated families occur frequently in probabilistic models in statistical mechanics, including the percolation model and models for ferromagnetism (see the discussion in [8]). The concept of association is also very useful in reliability situations where the random variables of interest are very often not independent but are associated. The asymptotic

Received May 28, 2002.

2000 Mathematics Subject Classification: 60F05.

Key words and phrases: empirical distribution function, associated, random field, asymptotic normality, estimator, survival function.

[†]This paper was partially supported by Korea Research Foundation Grant (KRF-2002-042).

[‡]This paper was partially supported by Statistical Research Center for Complex Systems, Seoul National University in 2003.

properties of associated random variables have been studied by Newman ([8], [9]) and Birkel ([2], [3]) among others. They observed that in any asymptotic property of associated random variables the covariance structure plays an important role. Bagai and Prakasa Rao ([1]) have studied the estimation of the survival function for stationary associated processes by using these asymptotic properties. In this paper we will extend this estimation to the associated random fields.

For a positive integer d , let \mathbb{Z}_+^d be the lattice of all points in \mathbb{R}^d having positive integer coordinates, and for each $\underline{n} = (n_1, \dots, n_d)$ in \mathbb{Z}_+^d , let $X_{\underline{n}}$ be a real-valued random variable defined on an underlying probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Thus, what we are dealing with here is a random field $\{X_{\underline{n}}, \underline{n} \in \mathbb{Z}_+^d\}$ whose elements take values in \mathbb{R} . For \underline{u} and \underline{v} in \mathbb{R}^d , $\underline{u} \leq \underline{v}$ means that $u_i \leq v_i$, $i = 1, \dots, d$, and these inequalities are all strict for $\underline{u} < \underline{v}$. For $\underline{n} = (n_1, \dots, n_d)$, let $|\underline{n}|$ denote the product $n_1 \times \dots \times n_d$. The norm, $\|\underline{n}\|$ of \underline{n} is taken to be $\|\underline{n}\| = \max\{n_i, i = 1, \dots, d\}$. Let $\underline{0} = (0, 0, \dots, 0)$ and $\underline{1} = (1, 1, \dots, 1)$. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be increasing (that is, nondecreasing), if $f(\underline{u}) \leq f(\underline{v})$ whenever $\underline{u} < \underline{v}$. Finally, for any nonempty subset A of $\{1, \dots, m\}$, let \mathbb{R}^A denote the cartesian product of $|A|$ copies of \mathbb{R} . All random fields in this paper are collections of random variables indexed by \mathbb{Z}_+^d and are assumed centered and stationary i.e. $E[X_{\underline{1}}] = 0$ and the distribution is invariant with respect to translations of the indices by the group \mathbb{Z}_+^d . A random field $\{X_{\underline{n}}, \underline{n} \in \mathbb{Z}_+^d\}$ is said to be associated if for any finite subset $A \subseteq \mathbb{Z}_+^d$ and for all coordinatewise nondecreasing functions $f, g : \mathbb{R}^A \rightarrow \mathbb{R}$ $Cov[f(X_{\underline{k}}, \underline{k} \in A), g(X_{\underline{k}}, \underline{k} \in A)]$ is nonnegative where the covariance is defined. In the following, let $\{X_{\underline{n}}, \underline{n} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with distribution function $F(x) = P(X_{\underline{1}} \leq x)$ or equivalently, survival function $\bar{F}(x) = 1 - F(x)$ and define the empirical distribution function $F_{\underline{n}}(x)$ by

$$\begin{aligned} F_{\underline{n}}(x) &= \frac{1}{|\underline{n}|} \sum_{\underline{1} \leq \underline{j} \leq \underline{n}} Y_{\underline{j}}(x) \\ (2) \quad &= \frac{1}{n_1 \times \dots \times n_d} \sum_{j_1=1}^{n_1} \dots \sum_{j_d=1}^{n_d} Y_{(j_1, \dots, j_d)}(x), \end{aligned}$$

where $\underline{j} = (j_1, \dots, j_d)$, $|\underline{n}| = n_1 \times \dots \times n_d$ for $\underline{n} = (n_1, \dots, n_d)$ and

$$(3) \quad Y_{\underline{j}}(x) = \begin{cases} 1 & \text{if } X_{\underline{j}} \leq x, \\ 0 & \text{otherwise.} \end{cases}$$

We propose $F_{\underline{n}}(x)$ as an estimator for $F(x)$ and derive the strong consistency and asymptotic normality of $F_{\underline{n}}(x)$ by using the moment inequalities of sums of associated random fields in Bulinskii ([4]). We also apply these concepts to the survival distribution for a stationary field of associated random variables and generalize the results of Bagai and Prakasa Rao ([1]) to the random fields.

2. Preliminaries

LEMMA 2.1. (Bagai, Prakasa Rao [1]) Suppose X and Y are associated random variables with bounded continuous densities. Then, there exists a constant $C > 0$ such that for any $T > 0$,

$$(4) \quad \begin{aligned} & \sup_{x,y} \{P[X \leq x, Y \leq y] - P[X \leq x]P[Y \leq y]\} \\ & \leq C\{T^2 Cov(X, Y) + \frac{1}{T}\}. \end{aligned}$$

Let us define for a stationary field $\{X_{\underline{j}}, \underline{j} \in \mathbb{Z}_+^d\}$ of associated random variables and $n \in \mathbb{Z}_+$, the coefficient (see [6])

$$(5) \quad u(n) = 2 \sum_{\underline{j}: \|\underline{j}\| \geq n+1} Cov(X_{\underline{1}}, X_{\underline{j}}),$$

where $\|\underline{j}\| = \max_{1 \leq k \leq d} \|j_k\|$ for $\underline{j} \in \mathbb{Z}_+^d$.

From Corollary 1 of Bulinskii ([4]) we have the following result:

LEMMA 2.2. Let $\{X_{\underline{j}}, \underline{j} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with zero mean. Let

$$(6) \quad E|X_{\underline{1}}|^{r+\delta} < \infty \text{ for some } r > 2, \delta > 0,$$

$$(7) \quad u(n) = O(n^{-\nu}) \text{ for some } \nu \geq 0.$$

If $\frac{r}{2} \geq \eta(1 - \nu\delta(d\eta)^{-1})/(r - \delta - 2)$ and $0 \leq \nu < \eta/\delta$, where $\eta = \delta + (r + \delta)(r - 2)$, then

$$(8) \quad \begin{aligned} \sup E \left| \sum_{\underline{1} < \underline{j} \leq \underline{n}} X_{\underline{j}} \right|^r &= \sup E \left| \sum_{j_1=1}^{n_1} \cdots \sum_{j_d=1}^{n_d} X_{\underline{j}} \right|^r \\ &= O(|\underline{n}|^{\frac{r}{2}}) \\ &= O((n_1 \times n_2 \times \cdots \times n_d)^{\frac{r}{2}}) \text{ as } |\underline{n}| \rightarrow \infty \end{aligned}$$

for $\underline{n} = (n_1, \dots, n_d)$.

REMARK. When $d = 1$ and $d = 2$ we obtain the estimate (8) if $\nu \geq d\nu_0$, where $\nu_0 = (1 + \delta)(r - 2)/(2\delta)$. In case $d \geq 3$ we obtain the estimate (8) too, provided that $\nu_0 < (d - 2)^{-1}$. Thus we have deduced for $d = 1$ the result extending [2] (see Remark 3 of Bulinskii [4]).

LEMMA 2.3. (Newman [8]) Let $\{X_j, j \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with $E[X_1^2] < \infty$. Assume

$$(9) \quad 0 < \sigma^2 = Var(X_1) + 2 \sum_{1 < j} Cov(X_1, X_j) < \infty.$$

Then

$$(10) \quad |\underline{n}|^{-\frac{1}{2}} \sigma^{-1} (S_{\underline{n}} - ES_{\underline{n}}) \xrightarrow{\mathcal{D}} Z \text{ as } |\underline{n}| \rightarrow \infty,$$

where $S_{\underline{n}} = \sum_{1 \leq i \leq \underline{n}} X_i$, $\xrightarrow{\mathcal{D}}$ indicates the convergence in distribution and Z is a standard normal variable.

3. The empirical distribution function

In the following, let $\{X_{\underline{n}}, \underline{n} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with distribution function $F(x) = P(X_1 \leq x)$ or equivalently, survival function $\bar{F}(x) = 1 - F(x)$ and define the empirical distribution function $F_{\underline{n}}(x)$ as in (2).

THEOREM 3.1. Let $\{X_j, j \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with bounded continuous density for X_1 and $E|X_1|^{r+\delta} < \infty$ for some $r > 2, \delta > 0$. Assume that

$$(11) \quad \sum_{j: \|j\| \geq n+1} \{Cov(X_1, X_j)\}^{1/3} = O(n^{-\nu}) \text{ for some } \nu \geq 0,$$

where $\|j\| = \max\{|j_i|, i = 1, \dots, d\}$. If $\frac{r}{2} \geq \eta(1 - \nu\delta(d\eta)^{-1})/(r - \delta - 2)$ and $0 \leq \nu < \eta/\delta$, where $\eta = \delta + (r + \delta)(r - 2)$, then there exists a constant $C > 0$ such that, for every $\epsilon > 0$,

$$(12) \quad \sup_x P[|F_{\underline{n}}(x) - F(x)| > \epsilon] \leq C\epsilon^{-r} |\underline{n}|^{-\frac{r}{2}}.$$

PROOF. Observe that

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

which is nonnegative since $Y_{\underline{1}}(x)$ and $Y_{\underline{j}}(x)$ are associated. Then there exists a constant $C > 0$ such that

$$\begin{aligned}
 (13) \quad & \sum_{\underline{j}: \|\underline{j}\| \geq n+1} Cov(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) \\
 & \leq \sum_{\underline{j}: \|\underline{j}\| \geq n+1} \sup_x \{P[X_{\underline{1}} > x, X_{\underline{j}} > x] - P[X_{\underline{1}} > x]P[X_{\underline{j}} > x]\} \\
 & \leq C \sum_{\underline{j}: \|\underline{j}\| \geq n+1} \{Cov(X_{\underline{1}}, X_{\underline{j}})\}^{1/3},
 \end{aligned}$$

by taking $T = \{Cov(X_{\underline{1}}, X_{\underline{j}})\}^{-1/3}$ in Lemma 2.1 whenever $Cov(X_{\underline{1}}, X_{\underline{j}}) > 0$ and if $Cov(X_{\underline{1}}, X_{\underline{j}}) = 0$ then $X_{\underline{1}}$ and $X_{\underline{j}}$ are independent as they are associated and $Cov(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) = P[X_{\underline{1}} > x, X_{\underline{j}} > x] - P[X_{\underline{1}} > x]P[X_{\underline{j}} > x] = 0 \leq [Cov(X_{\underline{1}}, X_{\underline{j}})]^{1/3}$. Furthermore

$$(14) \quad \sup_x \sup_{\underline{j}} |Y_{\underline{j}}(x) - EY_{\underline{j}}(x)| \leq 2$$

and from (11) and (13)

$$\begin{aligned}
 (15) \quad u(n, x) &= 2 \sum_{\underline{j}: \|\underline{j}\| \geq n+1} Cov(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) \\
 &\leq C \sum_{\underline{j}: \|\underline{j}\| \geq n+1} \{Cov(X_{\underline{1}}, X_{\underline{j}})\}^{1/3} \\
 &= O(n^{-\nu})
 \end{aligned}$$

for all real x , where C is independent of n and x . Hence the conditions (6) and (7) in Lemma 2.2 are satisfied according to (14) and (15), and thus from Lemma 2.2 it follows that, for every $\underline{n} \geq \underline{1}$

$$\begin{aligned}
 (16) \quad & \sup_x E \left| \sum_{1 \leq \underline{j} \leq \underline{n}} (Y_{\underline{j}}(x) - EY_{\underline{j}}(x)) \right|^r \\
 &= \sup_x E \left| \sum_{j_1=1}^{n_1} \cdots \sum_{j_d=1}^{n_d} (Y_{\underline{j}}(x) - EY_{\underline{j}}(x)) \right|^r \\
 &\leq C(n_1 \times \cdots \times n_d)^{\frac{r}{2}}, \quad \underline{j} = (j_1, \dots, j_d)
 \end{aligned}$$

where C is independent of \underline{n} and x . Then, by using Markov inequality, we get that for every $\epsilon > 0$,

$$\begin{aligned}
 (17) \quad & \sup_x P[|F_{\underline{n}}(x) - F(x)| > \epsilon] \\
 &= \sup_x P\left[\frac{1}{n_1 \times \dots \times n_d} \sum_{j_1=1}^{n_1} \dots \sum_{j_d=1}^{n_d} (Y_{\underline{j}}(x) - EY_{\underline{j}}(x)) > \epsilon\right] \\
 &\leq \sup_x \left\{ (n_1 \times \dots \times n_d)^{-r} \epsilon^{-r} E\left[\left|\sum_{j_1=1}^{n_1} \dots \sum_{j_d=1}^{n_d} (Y_{\underline{j}}(x) - EY_{\underline{j}}(x))\right|^r\right]\right\} \\
 &\leq C\epsilon^{-r} (n_1 \times \dots \times n_d)^{-r/2} \\
 &= C\epsilon^{-r} |\underline{n}|^{-r/2}.
 \end{aligned}$$

Thus, the desired result follows.

COROLLARY 3.1. *Let $\{X_{\underline{j}}, \underline{j} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with bounded continuous density for $X_{\underline{1}}$ and $E|X_{\underline{1}}|^{r+\delta} < \infty$ for some $r > 2$, $\delta > 0$. Assume that (11) holds. If $\frac{r}{2} \geq \eta(1 - \nu\delta(d\eta)^{-1})/(r - \delta - 2)$ and $0 \leq \nu < \eta/\delta$, where $\eta = \delta + (r + \delta)(r - 2)$, then for every x ,*

$$(18) \quad F_{\underline{n}}(x) \rightarrow F(x) \text{ a.s.}$$

PROOF. It follows from (17) that

$$\begin{aligned}
 \sum_{\underline{n} \geq \underline{1}} P[|F_{\underline{n}}(x) - F(x)| > \epsilon] &\leq C\epsilon^{-r} \sum_{\underline{n} \geq \underline{1}} |\underline{n}|^{-r/2} \\
 &= C\epsilon^{-r} \sum_{n_1=1}^{\infty} \dots \sum_{n_d=1}^{\infty} (n_1 \times \dots \times n_d)^{-r/2} \\
 &< \infty \text{ for } r > 2.
 \end{aligned}$$

Then the result follows by the Borel-Cantelli Lemma. □

THEOREM 3.2. *Let $\{X_{\underline{j}}, \underline{j} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with bounded continuous density for $X_{\underline{1}}$ and distribution function $F(x)$. Assume that*

$$(19) \quad \sum_{\underline{j} > \underline{1}}^{\infty} \{Cov(X_{\underline{1}}, X_{\underline{j}})\}^{1/3} < \infty.$$

Define

$$(20) \quad \sigma^2(x) = F(x)[1 - F(x)] + 2 \sum_{j>\underline{1}} \{P[X_{\underline{1}} \leq x, X_{\underline{j}} \leq x] - F^2(x)\}.$$

Then, for all x such that $0 < F(x) < 1$,

$$(21) \quad |\underline{n}|^{\frac{1}{2}} [F_{\underline{n}}(x) - F(x)] / \sigma(x) \xrightarrow{\mathcal{D}} Z, \text{ as } |\underline{n}| \rightarrow \infty,$$

where Z is a standard normal variable and $\xrightarrow{\mathcal{D}}$ indicates convergence in distribution.

PROOF. Note that, for $\underline{n} = (n_1, \dots, n_d)$

$$|\underline{n}| F_{\underline{n}}(x) = (n_1 \times n_2 \times \dots \times n_d) F_{\underline{n}}(x)$$

with

$$0 < \text{Var}(Y_{\underline{1}}) = F(x)(1 - F(x)) < 1$$

and

$$\text{Cov}(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) \geq 0 \text{ by association.}$$

Using Lemma 2.1 we get that

$$\begin{aligned} 0 &< \text{Var}(Y_{\underline{1}}) + 2 \sum_{j>\underline{1}} \text{Cov}(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) \\ &\leq F(x)(1 - F(x)) + 2 \sum_{j>\underline{1}} \{\text{Cov}(X_{\underline{1}}, X_{\underline{j}})\}^{1/3} \\ &< \infty \end{aligned}$$

by arguments similar to those given in the proof of Theorem 3.1. □

The result now follows from Lemma 2.3 due to Newman ([8]).

REMARK. Note that

$$\text{Cov}(Y_{\underline{1}}(x), Y_{\underline{j}}(x)) = P(X_{\underline{1}} \leq x, X_{\underline{j}} \leq x) - F^2(x)$$

and

$$|\underline{n}|^{\frac{1}{2}} [F_{\underline{n}}(x) - F(x)] / \sigma(x) = |\underline{n}|^{-\frac{1}{2}} \sum_{j \geq \underline{1}} [Y_{\underline{j}}(x) - EY_{\underline{j}}(x)] / \sigma(x).$$

Let $\{X_{\underline{j}}, \underline{j} \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables with survival function $\bar{F}(x) = P[X_{\underline{1}} > x]$. The empirical survival function $\bar{F}_{\underline{n}}(x)$ based on $X_{\underline{j}}, \underline{1} \leq \underline{j} \leq \underline{n}$, is proposed as an estimator for $\bar{F}(x)$ and is defined by

$$\bar{F}_{\underline{n}}(x) = \frac{1}{|\underline{n}|} \sum_{\underline{1} \leq \underline{j} \leq \underline{n}} \bar{Y}_{\underline{j}}(x)$$

where

$$\bar{Y}_{\underline{j}}(x) = \begin{cases} 1 & \text{if } X_{\underline{j}} > x, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\bar{F}_{\underline{n}}(x) = 1 - F_{\underline{n}}(x)$$

and $\bar{F}(x) = 1 - F(x)$. Thus from Theorems 3.1 and 3.2 and Corollary 3.1 we obtain the following results:

COROLLARY 3.2. *Let $\{X_j, j \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables under conditions of Theorem 3.1. Then, there exists a constant $C > 0$ such that, for every $\epsilon > 0$ and for some $r > 0$*

$$\sup_x P[|\bar{F}_{\underline{n}}(x) - \bar{F}(x)| > \epsilon] \leq C\epsilon^{-r}|\underline{n}|^{-r/2}.$$

COROLLARY 3.3. *Let $\{X_j, j \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables under conditions of Theorem 3.1. Then, for every x ,*

$$\bar{F}_{\underline{n}}(x) \longrightarrow \bar{F}(x) \text{ a.s. as } \underline{n} \rightarrow \infty.$$

COROLLARY 3.4. *Let $\{X_j, j \in \mathbb{Z}_+^d\}$ be a stationary field of associated random variables under conditions of Theorem 3.2. Define*

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

Then

$$|\underline{n}|^{\frac{1}{2}}[\bar{F}_{\underline{n}}(x) - \bar{F}(x)]/\sigma(x) \xrightarrow{\mathcal{D}} Z$$

as $\underline{n} \rightarrow \infty$ where Z is a standard normal distribution.

REMARK. Note that

$$\begin{aligned} \sigma^2(x) &= F(x)[1 - F(x)] + 2 \sum_{j>1} \{P[X_1 \leq x, X_j \leq x] - F^2(x)\} \\ &= \bar{F}(x)[1 - \bar{F}(x)] + 2 \sum_{j>1} \{P[X_1 > x, X_j > x] - \bar{F}^2(x)\}. \end{aligned}$$

ACKNOWLEDGMENTS. The authors wish to thank the referee for a very thorough review of this paper.

References

- [1] I. Bagai and B.L.S. Prakasa Rao, *Estimation of the survival function for stationary associated processes*, Statist. Probab. Letts. **12** (1991), 385–391.
- [2] T. Birkel, *Moment bounds for associated sequences*, Ann. Probab. **16** (1988), 1184–1193.
- [3] ———, *On the convergence rate in the central limit theorem for associated processes*, Ann. Probab. **16** (1988), 1685–1698.
- [4] A. V. Bulinskii, *Inequalities for moments of sums of associated multi-indexed variables*, Theory Probab. Appl. **38** (1993), 342–349.
- [5] R. M. Burton and T. S. Kim, *An invariance principle for associated random fields*, Pacific J. Math. **132** (1988), 11–19.
- [6] J. T. Cox and G. Grimmett, *Central limit theorems for associated random variables and the percolation model*, Ann. Probab. (1984) **12** 514–528.
- [7] J. Esary, F. Proschan, and D. Walkup, *Association of random variables with applications*, Ann. Math. Statist. **38** (1967), 1466–1474.
- [8] C. M. Newman, *Normal fluctuations and the FKG inequalities*, Comm. Math. Phys. **74** (1980), 119–128.
- [9] C. M. Newman, *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) (1984) pp.127–140.
- [10] C. M. Newman and A. L. Wright, *An invariance principle for certain dependent sequences*, Ann. Probab. **9** (1981), 671–675.

Tae-Sung Kim
Division of Mathematics and Informational Statistics
and Institute of Basic Natural Science
WonKwang University
Iksan 570-749, Korea
E-mail: starkim@wonkwang.ac.kr

Mi-Hwa Ko
Statistical Research Center for Complex Systems
Seoul National University
Seoul 151-742, Korea
E-mail: songhack@wonkwang.ac.kr

Yeon-Sun Yoo
Division of Mathematics and Informational Statistics
and Institute of Basic Natural Science
WonKwang University
Iksan 570-749, Korea