Теорія Ймовір. та Матем. Статист.
Вип. 69, 2003

Theor. Probability and Math. Statist.
No. 69, 2004, Pages 167–174
S 0094-9000(05)00623-X
Article electronically published on February 9, 2005

ESTIMATES OF DISTRIBUTIONS OF COMPONENTS IN A MIXTURE FROM CENSORING DATA

UDC 519.21

A. YU. RYZHOV

Abstract. The problem of estimation of the distribution functions of components in a mixture in the case of censored observations is considered. Optimal estimators are found in the class of linear estimators. Since the optimal estimators depend on unknown distribution functions of components, an adaptive estimation scheme is used. The asymptotic normality is proved for adaptive estimators and it is shown that their concentration coefficient coincides with that of the optimal linear estimator.

1. Introduction

Statistical methods are often used when processing survival data in medicine as well as in engineering. A distinctive feature of survival data is that some observations may be *censored*. Parametric methods of the estimation of censored observations are known in the literature (see, for example, [1]). Nevertheless the nonparametric methods are the most useful for the estimation of the cumulative function of intensity $\Lambda(t)$ and survivor function $S(t)$. Other useful methods are the methods of comparison of these functions for different groups.

In practice, it does often happen that the objects observed during an experiment belong to several populations, and the available data are obtained from a mixture of populations. Thus a natural problem is to estimate probability characteristics for every sample from a mixture of populations.

Here is the precise setting of the problem. Let m populations be given and let the distribution functions of the life time be

$$H_1(t), H_2(t), \dots, H_m(t), \qquad t \ge 0.$$

Assume that $N \ge m$ sample are given and every observation may belong to one of the populations. Let the observations in every sample be right-censored. For an arbitrary observation in the sample j, we also assume that the probability that the observation is censored depends on the sample but does *not* depend on the population containing the observation. Denote by $w_l^{(j)}$ the probabilities (*concentrations* of components in a mixture) that an object of the sample j belongs to the population l. We assume that the probabilities $\{w_l^{(j)} : l = 1, \dots, m\}_{j=1}^N$ are known for every sample. Thus the distribution of the life time of an object in the sample j is given by

$$F_j(t) = w_1^{(j)} H_1(t) + w_2^{(j)} H_2(t) + \dots + w_m^{(j)} H_m(t), \qquad j = 1, \dots, N, \tag{1}$$

2000 *Mathematics Subject Classification.* Primary 62N02; Secondary 62G05.
Key words and phrases. Survival analysis, mixtures with varying concentrations, censoring, Kaplan–Meier estimators, concentration coefficient.

where

$$\sum_{l=1}^{m} w_l^{(j)} = 1, \qquad 0 \le w_l^{(j)} \le 1,$$

for all j. Mixtures of this type are called *mixtures with varying concentrations*.

The distribution functions $F_j(t)$, $j = 1, \dots, N$, can be estimated from any sample using the Kaplan–Meier method. Denote the Kaplan–Meier estimators by $\hat{F}_j(t)$. Now the problem is to find estimators $\hat{H}_1(t), \hat{H}_2(t), \dots, \hat{H}_m(t)$ of the distribution functions of components in the mixture. First we search for estimators $\hat{H}_l(t)$ in the class of linear estimators, that is, in the class of estimators of the following form:

$$\hat{H}(t, \overline{a}^{(l)}(t)) = \hat{H}_l(t) = \sum_{j=1}^{N} a_j^{(l)}(t) \hat{F}_j(t) \tag{2}$$

where $\overline{a}^{(l)}(t) = (a_j^{(l)}(t), j = 1, \dots, N)$ is a nonrandom vector of weight coefficients.

We show in Section 3 that, under certain conditions, estimators $\hat{H}(t, \overline{a}^{(l)}(t))$ are asymptotically normal with some concentration coefficient $\sigma^2(t, \overline{a}^{(l)}(t))$. We also find the optimal $\overline{a}^{(l)}(t)$ for which $\sigma^2(t, \overline{a}^{(l)}(t))$ is minimal. It will turn out, however, that the coefficient $\overline{a}^{(l)}(t)$ depends on unknown values $H_l(t)$. Thus we need to use an adaptive estimation procedure. At the first step of the procedure, we obtain rough estimators of $H_l(t)$ using nonoptimal weight coefficients. The estimators obtained at the first step are used further to obtain optimal weight coefficients. Using the optimal weight coefficients we obtain adaptive estimators at the final step of the procedure.

We show in Section 4 that adaptive estimators are asymptotically normal, and their concentration coefficient coincides with that of the optimal linear estimator.

2. The Kaplan–Meier estimator and its properties

The majority of the methods of survival analysis are based on the *model of random censoring*. In what follows I_A denotes the indicator of a set A.

Definition 1. In the model of random censoring, ordered pairs

$$(T_{jk}, U_{jk}), \qquad k = 1, \dots, n_j,\ j = 1, \dots, N,$$

denote n ($n \equiv \sum_{j=1}^{N} n_j$) independent random variables called *failure* and *censoring* times, respectively. Observed are the variables

$$X_{jk} = \min(T_{jk}, X_{jk}) \equiv T_{jk} \wedge U_{jk}$$

and

$$\eta_{jk} = I_{\{X_{jk} = T_{jk}\}}.$$

The corresponding distribution functions are denoted by

$$S_j(t) := \mathsf{P}\{T_{jk} > t\},$$
$$F_j(t) := 1 - S_j(t),$$
$$C_{jk}(t) := \mathsf{P}\{U_{jk} > t\},$$
$$L_{jk}(t) := 1 - C_{jk}(t).$$

Definition 2. The function $S_j(t)$ is called the *survivor function* of the sample j, while

$$\Lambda_j(t) := \int_0^t \{1 - F_j(s-)\}^{-1} \, dF_j(s)$$

is called the *cumulative intensity function*.

In the general case, the probability of censoring U_{jk} in any of N samples can be different for different objects. In what follows we assume that the distributions $C_{jk}(t)$ and $L_{jk}(t)$ do not depend on k.

Consider the following stochastic processes:

$$\overline{N}_j(t) := \sum_{k=1}^{n_j} N_{jk}(t) = \sum_{k=1}^{n_j} I_{\{X_{jk}\le t,\eta_{jk}=1\}},$$

$$\overline{Y}_j(t) := \sum_{k=1}^{n_j} Y_{jk}(t) = \sum_{k=1}^{n_j} I_{\{X_{jk}\ge t\}}.$$

Denote by L_j the number of failures in a sample of n_j elements (thus, $L_j \le n_j$), $L_j = \sum_{k=1}^{n_j} \eta_{jk}$, and let $T_1^j < T_2^j < \cdots < T_{L_j}^j$ be the failure times placed in ascending order. The *Kaplan–Meier estimator* [2] of the survivor function is defined by

$$\hat{S}_j(t) = \prod_{k:T_k^j\le t} \left(1 - \Delta\overline{N}_j\left(T_k^j\right)/\overline{Y}_j\left(T_k^j\right)\right)$$

where $\Delta\overline{N}_j(t) = \overline{N}_j(t) - \overline{N}_j(t-)$.

The Kaplan–Meier estimate $\hat{S}_j(t)$ has the following properties:

1) If $S_j(t) > 0$ and $T_j = \inf\{s\colon \overline{Y}_j(s) = 0\}$, then

$$\mathsf{E}\left\{\hat{S}_j(t) - S_j(t)\right\} = \mathsf{E}\left[I_{\{T_j<t\}}\frac{\hat{S}_j(T)\{S_j(T) - S_j(t)\}}{S_j(T)}\right].$$

2) If $F_j(t)$ is continuous and $t > 0$ is such that $\overline{Y}_j(t) \to \infty$ in probability, then the estimator is *consistent* on $[0,t]$, that is,

$$\sup_{0\le s\le t}\left|\hat{F}_j(s) - F_j(s)\right| \to 0, \qquad n_j \to \infty,$$

in probability where $\hat{F}_j(s) = 1 - \hat{S}_j(s)$.

(See [3] for the proof of 1) and 2).)

The following result on the limit properties of the Kaplan–Meier estimator is a version of Theorem 6.1.3 of [3].

Theorem 1. *Assume that there are a continuous function* $\pi_j(t)$ *and a constant* $b_j \in [0,1]$ *such that*

$$\sup_{0\le t<\infty}\left|\frac{\overline{Y}_j(t)}{n_j} - \pi_j(t)\right| \to 0, \qquad n \to \infty,$$

in probability and

$$\frac{n_j}{n} \to b_j, \qquad n \to \infty.$$

Let the distribution function $F_j(t)$ *be continuous. Put* $X = \{t\colon \pi_j(t) > 0\}$ *and let* W_j *be the Wiener process. Then for all* $t \in X$

1) $\sqrt{n}(\hat{F}_j(\cdot) - F_j(\cdot)) \Longrightarrow (1 - F_j(\cdot))W_j(v_j(\cdot))$ *in the space* $D[0,t]$ *where*

$$v_j(t) \equiv \int_0^t \pi_j^{-1}(s)\,d\Lambda_j(s).$$

2) *If* $\hat{v}_j(t) := n_j\int_0^t [\{\overline{Y}_j(s) - \Delta\overline{N}_j(s)\}\overline{Y}_j(s)]^{-1}\,d\overline{N}_j(s)$, *then*

$$\sup_{0\le s\le t}|\hat{v}_j(s) - v_j(s)| \to 0, \qquad n \to \infty,$$

in probability.

3. Linear estimators of the distribution functions of components in a mixture

In what follows, for the model of random censoring under consideration, we assume that T_{jk} and U_{jk} are independent for all $k = 1, \dots, n_j$ and $j = 1, \dots, N$ and that the functions

$$S_j(t) = \mathsf{P}\{T_{jk} > t\}, \qquad C_j(t) = \mathsf{P}\{U_{jk} > t\}$$

are continuous. We also assume that $n_j \to \infty$ for all j, and there exist constants $b_j \in [0,1]$ such that $n_j/n \to b_j$ as $n \to \infty$ where $n \equiv \sum_{j=1}^N n_j$. Then one can apply Theorem 1 for any of the N Kaplan–Meier estimators $\hat{F}_j(t)$ by putting

$$\pi_j(t) := \mathsf{P}\{X_{jk} > t\} = S_j(t) \cdot C_j(t)$$

where $X_{jk} = \min(T_{jk}, U_{jk})$. We use the following notation in the proof of Theorem 1:

$$X := \left\| w_i^{(j)} \right\|_{j=1,i=1}^{N \quad m},$$

$$\gamma_{pi}(t) := \sum_{j=1}^N \frac{w_i^{(j)} w_p^{(j)}}{S_j^2(t) v_j(t)},$$

$$\Gamma(t) := \|\gamma_{pi}(t)\|_{m\times m},$$

$$V(t) := \left\| \frac{w_i^{(j)}}{S_j^2(t) v_j(t)} \right\|_{j=1,i=1}^{N \quad m}.$$

Let $\bar{e}_l := (\delta_{l1}, \dots, \delta_{lm})^T$ be the column l of the unit matrix where δ_{ij} is the Kronecker symbol.

Theorem 2. *Assume that the columns of the matrix X are linearly independent and put*

$$X_N = \left\{ t\colon \prod_{j=1}^N \pi_j(t) > 0 \right\}.$$

Let $t \in X_N$, $t < \infty$, and

$$\sum_{j=1}^N a_j^{(l)}(t) w_i^{(j)} = \delta_{li}, \qquad i = 1, \dots, m, \tag{3}$$

for some bounded functions $a_j^{(l)}(\cdot)$. Then

1) *estimators $\hat{H}(t, \overline{a}^{(l)}(t))$ defined by* 2 *are asymptotically unbiased estimators for the distribution functions $H_l(t)$;*
2) *for $0 \le s \le t$,*

$$\sqrt{n}\left(\hat{H}_l(s) - H_l(s)\right) \Longrightarrow W\left(\sigma^2\left(s, \overline{a}^{(l)}(s)\right)\right), \qquad n \to \infty,$$

where W is the Wiener process. The concentration coefficient $\sigma^2(s, \overline{a}^{(l)}(s))$ is minimal in the class of estimators satisfying condition 3 *if*

$$\overline{a}^{(l)}(s) = \left(V(s)\Gamma^{-1}(s)\right)\bar{e}_l.$$

Proof. It follows from (3) and representation (1) that

$$H_l(t) = \sum_{j=1}^N a_j^{(l)}(t) F_j(t) \tag{4}$$

for any fixed $1 \le l \le m$. This implies that

$$\mathsf{E}\left(H_l(t) - \hat{H}_l(t)\right) = \mathsf{E}\left(\sum_{j=1}^{N} a_j^{(l)}(t) F_j(t) - \sum_{j=1}^{N} a_j^{(l)}(t) \hat{F}_j(t)\right)$$

$$= \sum_{j=1}^{N} a_j^{(l)}(t)\,\mathsf{E}\left(F_j(t) - \hat{F}_j(t)\right) = \sum_{j=1}^{N} a_j^{(l)}(t)\,\mathsf{E}\left(\hat{S}_j(t) - S_j(t)\right).$$

Since $t \in X_N$, the definition of the function $\pi_j(t)$ yields that $S_j(t) > 0$ for all j. Let $T_j = \inf\{s\colon \overline{Y}_j(s) = 0\}$. Then

$$0 \le \mathsf{E}\left\{\hat{S}_j(t) - S_j(t)\right\} = \mathsf{E}\left[I_{\{T_j<t\}} \frac{\hat{S}_j(T)\{S_j(T) - S_j(t)\}}{S_j(T)}\right] \le \mathsf{E}\left[I_{\{T_j<t\}}\left(1 - \frac{S_j(t)}{S_j(T)}\right)\right]$$

$$\le \mathsf{E}\, I_{\{T_j<t\}} \le \mathsf{P}\{\overline{Y}_j(t) = 0\}$$

according to properties of the Kaplan–Meier estimator. Since

$$\sup_{0 \le s < \infty} \left| \frac{\overline{Y}_j(s)}{n_j} - \pi_j(s) \right| \to 0, \qquad n \to \infty,$$

in probability and $\pi_j(t) > 0$, we have $\overline{Y}_j(t) \to \infty$ as $n \to \infty$ in probability. Hence

$$0 \le \mathsf{E}\left\{\hat{S}_j(t) - S_j(t)\right\} \le \mathsf{P}\{\overline{Y}_j(t) = 0\} \to 0, \qquad n \to \infty.$$

By assumption $a_j^{(l)}(t)$ are bounded, so that

$$\mathsf{E}\left(H_l(t) - \hat{H}_l(t)\right) \to 0, \qquad n \to \infty.$$

Note that the estimator $\hat{H}_l(t)$ is unbiased if $\overline{Y}_j(t) > 0$ at the point t for all j, since any estimator $\hat{S}_j(t)$ is unbiased in this case.

To prove the second statement of Theorem 1 we consider $\sqrt{n}(\hat{H}_l(t) - H_l(t))$. First we study the case of $N > m$. Using representation (4), we get

$$\sqrt{n}(\hat{H}_l(t) - H_l(t)) = \sqrt{n}\left(\sum_{j=1}^{N} a_j^{(l)}(t) \hat{F}_j(t) - \sum_{j=1}^{N} a_j^{(l)}(t) F_j(t)\right)$$

$$= \sqrt{n}\left(\sum_{j=1}^{N} a_j^{(l)}(t)\left(\hat{F}_j(t) - F_j(t)\right)\right) = \sum_{j=1}^{N} a_j^{(l)}(t)\sqrt{n}\left(\hat{F}_j(t) - F_j(t)\right).$$

Thus one expects that

$$\sqrt{n}\left(\hat{H}_l(t) - H_l(t)\right) \Longrightarrow \sum_{j=1}^{N} a_j^{(l)}(t)(1 - F_j(t)) W_j(v_j(t))$$

$$= W\left(\sum_{j=1}^{N}\left(a_j^{(l)}(t)(1 - F_j(t))\right)^2 v_j(t)\right). \tag{5}$$

Assume that (5) indeed holds. Then the concentration coefficient of estimators $\hat{H}_l(t)$ is minimal if (3) holds and

$$\sigma^2\left(t, \overline{a}^{(l)}(t)\right) = \sum_{j=1}^{N}\left(a_j^{(l)}(t)(1 - F_j(t))\right)^2 v_j(t) \to \inf$$

for all $l = 1, \dots, m$. The latter is a conditional extremum problem. Solving it by the Lagrange method we obtain

$$A(t) = V(t)\Gamma^{-1}(t)$$

where $A(t)$ is the matrix whose columns determine the optimal weight coefficients for the estimators of the distribution functions; the matrices $V(t)$ and $\Gamma(t)$ are defined above. Note that $\sigma^2(t, \overline{a}^{(l)}(t))$, $l = 1, \dots, m$, are minimal for the functions $a_j^{(l)}(t)$ since the sufficient condition for the existence of the conditional extremum is satisfied.

Consider conditions under which $\det \Gamma(t) = 0$. Since $v_k(t) > 0$ and $S_k(t) > 0$ for all k and $t \in X_N$, we have

$$\gamma_{pi} = \sum_{j=1}^{N} \frac{w_i^{(j)}}{S_j(t)\sqrt{v_j(t)}} \frac{w_p^{(j)}}{S_j(t)\sqrt{v_j(t)}},$$

that is, γ_{pi} is the scalar product of the pth and ith columns of the matrix

$$V_1(t) = \begin{pmatrix} \frac{w_1^{(1)}}{S_1(t)\sqrt{v_1(t)}} & \cdots & \frac{w_m^{(1)}}{S_1(t)\sqrt{v_1(t)}} \\ \vdots & \ddots & \vdots \\ \frac{w_1^{(N)}}{S_N(t)\sqrt{v_N(t)}} & \cdots & \frac{w_m^{(N)}}{S_N(t)\sqrt{v_N(t)}} \end{pmatrix},$$

and $\Gamma(t)$ is the Gram matrix constructed from these columns. Thus $\det \Gamma(t) = 0$ if and only if columns of the matrix $V_1(t)$ are linearly dependent. The functions

$$\left\{ S_k(t)\sqrt{v_k(t)} \right\}^{-1}$$

do not affect linear dependence, thus $\det \Gamma(t) \neq 0$ if columns of the matrix X are independent.

Let us show that relation (5) holds for the functions $a_j^{(l)}(t)$ found above. The functions $S_j(\cdot)$ are continuous by assumption. It is easy to show that all the functions $v_j(\cdot)$ also are continuous at the point t. Since $S_j(t) > 0$ and $v_j(t) > 0$, all the entries of matrices $\Gamma^{-1}(t)$, $V(t)$, and $A(t)$ are continuous at this point, too. We have

$$\sqrt{n}\left(\hat{H}_l(\cdot) - H_l(\cdot)\right) = \sum_{j=1}^{N} a_j^{(l)}(\cdot)\sqrt{n}\left(\hat{F}_j(\cdot) - F_j(\cdot)\right)$$
$$\Longrightarrow W\left(\sum_{j=1}^{N} \left(a_j^{(l)}(\cdot)(1 - F_j(\cdot))\right)^2 v_j(\cdot)\right)$$

on $[0, t]$, since

$$\sqrt{n}\left(\hat{F}_j(\cdot) - F_j(\cdot)\right) \Longrightarrow (1 - F_j(\cdot))W_j(v_j(\cdot))$$

on $[0, t]$.

Now let $N = m$. The functions $a_j^{(l)}(t)$ do not depend on t and are uniquely determined by conditions (3) and $\Delta \neq 0$ where

$$\Delta = \det \begin{vmatrix} w_1^{(1)} & \cdots & w_m^{(1)} \\ \vdots & \ddots & \vdots \\ w_1^{(m)} & \cdots & w_m^{(m)} \end{vmatrix}. \qquad \square$$

Remark. If the columns of the matrix X are linearly dependent, then $\det \Gamma(t) = 0$ for all t and there exist numbers $\{\alpha_1, \dots, \alpha_m\} \subset \mathbb{R}$ such that $\sum |\alpha_i| \neq 0$ and

$$\alpha_1 w_1^{(j)} + \alpha_2 w_2^{(j)} + \dots + \alpha_m w_m^{(j)} = 0$$

for all $j = 1, \dots, N$. If $\alpha_1 \neq 0$, then

$$w_1^{(j)} = \beta_2 w_2^{(j)} + \dots + \beta_m w_m^{(j)},$$

whence

$$F_j(t) = (H_2(t) + \beta_2 H_1(t)) w_2^{(j)} + (H_3(t) + \beta_3 H_1(t)) w_3^{(j)} + \dots + (H_m(t) + \beta_m H_1(t)) w_m^{(j)}.$$

Thus consistent estimators of components in the mixture do not exist in this case.

4. Adaptive estimators of the distribution functions of components in a mixture

Column vectors of the matrix $A(t)$ determine the best estimators of unknown distribution functions $H_l(t)$ in the sense of the minimum of the concentration coefficient. However they are of no practical importance, since all of them depend on the unknown functions $v_j(t)$ and $S_j(t)$, $j = 1, \dots, N$. Put

$$Q(t) = \begin{Vmatrix} \frac{1}{S_1^2(t) v_1(t)} & 0 & \cdots & 0 \\ 0 & \frac{1}{S_2^2(t) v_2(t)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{S_N^2(t) v_N(t)} \end{Vmatrix}.$$

Then the matrices $\Gamma(t)$ and $V(t)$ can be rewritten as follows:

$$\Gamma(t) = X^T Q(t) X,$$
$$V(t) = Q(t) X.$$

Consider the matrices $\hat{\Gamma}(t) = X^T \hat{Q}(t) X$ and $\hat{V}(t) = \hat{Q}(t) X$ for t such that $0 < \hat{v}_j(t) < \infty$ and $\hat{S}_j(t) > 0$ for all j. If the columns of the matrix X are linearly independent, then $\det \hat{\Gamma}(t) \neq 0$. In this case the matrix

$$\hat{A}(t) = \hat{V}(t) \hat{\Gamma}^{-1}(t) \tag{6}$$

determines new, *adaptive*, estimators for $H_l(t)$, namely

$$\tilde{H}_l(t) = \sum_{j=1}^{N} \hat{a}_j^{(l)}(t) \hat{F}_j(t), \qquad l = 1, \dots, m,$$

where $\hat{a}_j^{(l)}(t)$ are entries of the column l in the matrix $\hat{A}(t)$.

Now we show that for all $t \in X_N$, $\sqrt{n}(\tilde{H}_l(t) - H_l(t))$ and $\sqrt{n}(\hat{H}_l(t) - H_l(t))$ have the same concentration coefficient

$$\sigma^2 \left(t, \overline{a}^{(l)}(t) \right) = \sum_{j=1}^{N} \left(a_j^{(l)}(t)(1 - F_j(t)) \right)^2 v_j(t).$$

Theorem 3. *Assume that all the assumptions of Theorem 2 hold for functions*

$$\left\{ a_j^{(l)}(t), l = 1, \dots, m, j = 1, \dots, N \right\}, \qquad t \in X_N,\ t < \infty,$$

and let

$$\left\{ \hat{a}_j^{(l)}(t), l = 1, \dots, m, j = 1, \dots, N \right\}$$

be their estimators defined by 6. *Then*

$$\sqrt{n} \left(\tilde{H}_l(t) - H_l(t) \right) \Longrightarrow W \left(\sigma^2 \left(t, \overline{a}^{(l)}(t) \right) \right).$$

Proof. First we show that condition (3) holds for functions $\hat{a}_j^{(l)}(t)$. Indeed, rewriting (3) in the matrix form we get

$$X^T \hat{A}(t) = E_{m\times m}$$

where $E_{m\times m}$ is the unit matrix of size $m \times m$. Since

$$\hat{A}(t) = \hat{V}(t)\hat{\Gamma}^{-1}(t) = \left(\hat{Q}(t)X\right)\left(X^T\hat{Q}(t)X\right)^{-1},$$

we have

$$X^T\hat{A}(t) = \left(X^T\hat{Q}(t)X\right)\left(X^T\hat{Q}(t)X\right)^{-1} = E_{m\times m}.$$

This means that

$$\sqrt{n}\left(\tilde{H}_l(t) - H_l(t)\right) = \sum_{j=1}^{N} \hat{a}_j^{(l)}(t)\sqrt{n}\left(\hat{F}_j(t) - F_j(t)\right)$$

for all l. Since $v_j(t) > 0$ and $\hat{v}_j(t) \to v_j(t)$ as $n \to \infty$ in probability, we obtain

$$\frac{1}{\hat{v}_j(t)} \to \frac{1}{v_j(t)}$$

in probability. Similarly

$$\frac{1}{\hat{S}_j^2(t)} \to \frac{1}{S_j^2(t)}$$

in probability. This implies that the functions $\hat{a}_j^{(l)}(t)$ converge in probability to $a_j^{(l)}(t)$ as $n \to \infty$. Since $a_j^{(l)}(t)$ are constant for a fixed t, $\hat{a}_j^{(l)}(t) \to a_j^{(l)}(t)$ in distribution. Now Theorem 3 follows from the convergence in distribution of sums and products of random variables if the latter converge in distribution. □

Bibliography

1. D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman & Hall, New York, 1984. MR0751780 (86f:62170)
2. E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457–481. MR0093867 (20:387)
3. D. P. Harrington and T. R. Fleming, *Counting Processes and Survival Analysis*, Wiley, New York, 1991. MR1100924 (92i:60091)

Department of Mathematics, Faculty of Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine

E-mail address: tosha@ucr.kiev.ua

Received 14/FEB/2003

Translated by V. SEMENOV