Теорія Ймовір. та Матем. Статист.
Вип. 96, 2017

Theor. Probability and Math. Statist.
No. 96, 2018, Pages 133–144
https://doi.org/10.1090/tpms/1039
Article electronically published on October 5, 2018

ASYMPTOTIC NORMALITY OF KAPLAN–MEIER ESTIMATORS FOR MIXTURES WITH VARYING CONCENTRATIONS

UDC 519.21

R. E. MAĬBORODA

Abstract. We consider a modified Kaplan–Meier estimator for the distribution of components in a mixture with varying concentrations in the case of censored data. The asymptotic normality of this estimator is proved in the uniform norm.

1. Introduction

The classical Kaplan–Meier estimator (KM estimator) is widely used in the analysis of data similar to the failure times as a non-parametric estimator of the distribution function constructed from censored data. A modification of this estimator (mKM estimator) is introduced in the paper [11] for the case where the observations are obtained from a mixture of several populations (components) with varying concentrations (MVC model). The consistency of the mKM estimators and estimates of rate of convergence are obtained in [11].

In the current paper, the asymptotic normality of the mKM estimator is proved in the uniform norm on a finite interval. This result is a generalization of the classical theorem on the asymptotic normality of the KM estimator [5] and allows us to obtain an analogue of the Greenwood formula for the asymptotic variance of mKM estimator. The proof is based on the asymptotic theory of weighted empirical functions [9, 10], the theory of product integrals [5], and the classical results of the weak convergence of probability measures in functional spaces [2].

We recall the definition of the KM estimator for homogeneous samples and provide some results concerning its asymptotic normality in Section 2.1. The estimation of the distribution function in the MVC model without censoring is considered in Section 2.2. The definition of the mKM estimator and the main result of the paper on the asymptotic normality of the mKM estimator is contained in Section 3. The proof is given in Section 4.

2. Preliminaries

2.1. Censoring and the Kaplan–Meier estimator for homogeneous samples.

We start with the description of a standard random right-censoring model and that of the construction of the KM estimator.

Let ξ_j, $j = 1, \dots, n$, be the failure times of certain objects which are assumed to be independent and identically distributed random variables. The failure times are observed if they do not exceed some censoring moments C_j for the corresponding objects (that is, they are observed if an object fails before it was censored). If censoring precedes the

2010 *Mathematics Subject Classification.* Primary 62N05, 62G05.

Key words and phrases. Kaplan–Meier estimator, model of mixtures with varying concentrations, asymptotic normality, censoring.

failure, then the censoring time is observed for this object. It is also known for every object whether or not the censoring happens for it.

Therefore, the observations are $\mathbf{X} = (\xi_j^*, \delta_j, j = 1\dots, n)$, where

$$\xi_j^* = \min(\xi_j, C_j)$$

are censored failure times and $\delta_j = \mathbb{1}\{\xi_j \le C_j\}$ are indicators of whether or not the censoring happens.

The distribution function F of random variables ξ_j is assumed to be unknown. The empirical maximal likelihood estimator of F constructed from $\mathbf{X}$ coincides with the KM estimator [13]. The latter is defined as follows.

Let

$$\widehat{Y}_n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left\{\xi_j^* \ge t\right\},$$

$$\widehat{N}_n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left\{\xi_j^* \le t, \delta_j = 1\right\}.$$

Assume that all ξ_j^* in a sample are different. Then the classical Kaplan–Meier estimator for $F(x)$ is defined by

$$\widehat{F}(t) = 1 - \prod_{j:\xi_j^*\le t}\left(1 - \frac{\Delta\widehat{N}_n(\xi_j^*)}{\widehat{Y}_n(\xi_j^*)}\right) = 1 - \prod_{j:\xi_j^*\le t}\left(1 - \frac{\delta_j}{n - \sum\limits_{i:\xi_i^*<\xi_j^*} 1}\right), \tag{1}$$

where $\Delta\widehat{N}_n(t) = \widehat{N}_n(t) - \widehat{N}_n(t-)$ means the jump of the function $\widehat{N}_n$ at a point t. The limit from the left of the function F at a point t is denoted by $F(t-) = \lim_{s<t, s\to t} F(s)$.

Assume that the censoring moments C_j are independent identically distributed random variables with the distribution function G. If F and G are continuous and $F(t) < 1$ and $G(t) < 1$ for some $t > 0$, then

$$\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) \xrightarrow{\mathrm{w}} N\left(0, \sigma^2(t)\right), \tag{2}$$

where

$$\sigma^2(t) = \big(1 - F(t)\big)^2 \int_0^t \frac{F(du)}{\big(1 - G(u)\big)\big(1 - F(u)\big)^2} \tag{3}$$

(see [5]). Here and in what follows the symbol $\xrightarrow{\mathrm{w}}$ denotes the weak convergence. Equality (3) is an asymptotic version of the classical Greenwood formula for the KM estimator [6, equality (3.2.31)].

2.2. **Mixtures with varying concentrations.** It is assumed in the MVC model that every object O belongs to one of M different populations. The true number $\operatorname{ind}(O)$ of the population to which an object O belongs is unknown. Instead, one observes a certain characteristic $\xi = \xi(O)$ assumed to be a random variable with the distribution function F_m that depends on a population, that is,

$$F_m(t) = \mathsf{P}\{\xi(O) \le t \mid \operatorname{ind}(O) = m\}.$$

Therefore the distribution function of $\xi(O)$ is a mixture of F_m. The distribution functions F_m are unknown, but the concentrations of components in the mixture are known and different for different observations. Various problems related to this model are considered in [1, 3, 10].

If n independent objects O_j are observed, the statistical data are $(\xi_{j;n},\ j=1,\dots,n)$, where the distribution function of $\xi_{j;n}=\xi(O_j)$ is

$$\mathsf{P}\{\xi_{j;n}\le t\}=\sum_{m=1}^{M}p_j^m F_m(t). \tag{4}$$

Here

$$p_j^m=\mathsf{P}\{\operatorname{ind}(O_j)=m\}$$

is the concentration (mixing probability) of the mth component (population) in the mixture when the jth object is observed.

The set of all concentrations $\{p_{j;n}^m, j=1,\dots,n;\ m=1,\dots,M;\ n=1,2,\dots\}$ is denoted by $\mathbf{p}$. Let $\mathbf{p}_{;n}^m=(p_{1,n}^m,\dots,p_{n;n}^m)^T$ be the vector column of the concentrations of the mth component, $\mathbf{p}_{j;n}=(p_{j,n}^1,\dots,p_{j,n}^M)^T$ be the vector column of the concentrations at the moment when the jth object is observed, and let $\mathbf{p}_{;n}=(p_{j,n}^m, j=1,\dots,n;\ m=1,\dots,M)$ be the matrix of concentrations for a sample of n elements with n columns and M rows.

Analogous notation is used for the set of weight coefficients

$$\mathbf{a}=\{a_{j;n}^m, j=1,\dots,n;\ m=1,\dots,M;\ n=1,2,\dots\}$$

to be introduced below.

The averaging for the whole sample (that is, the averaging with respect to the index j) is denoted by $\langle\mathbf{p}^m\mathbf{a}^k\rangle_n$:

$$\langle\mathbf{p}^m\mathbf{a}^k\rangle_n=\frac{1}{n}\sum_{j=1}^{n}p_{j;n}^m a_{j;n}^k,\qquad \langle(\mathbf{a}^k)^2\rangle_n=\frac{1}{n}\sum_{j=1}^{n}\left(a_{j;n}^k\right)^2,$$

and so on. The addition, multiplication, and raising to the power within brackets is performed coordinatewise. The angle brackets without a subscript mean the limit

$$\langle\mathbf{p}^m\mathbf{a}^k\rangle=\lim_{n\to\infty}\langle\mathbf{p}^m\mathbf{a}^k\rangle_n$$

if it exists. Put

$$\mathbf{\Gamma}_n=\frac{1}{n}\mathbf{p}_{;n}\mathbf{p}_{;n}^T=\left(\langle\mathbf{p}^m\mathbf{p}^k\rangle_n\right)_{m,k=1}^M,\qquad \mathbf{\Gamma}=\lim_{n\to\infty}\mathbf{\Gamma}_n=\left(\langle\mathbf{p}^m\mathbf{p}^k\rangle\right)_{m,k=1}^M.$$

It is proposed in [8, 9] (also see [10]) to use

$$\widehat{F}_{\mathbf{a}}(t)=\frac{1}{n}\sum_{j=1}^{n}a_{j;n}\mathbb{1}\{\xi_{j;n}\le t\}$$

as an estimator of the distribution function $F_m(t)$ constructed from a sample $\mathbf{X}$.

It is shown in [10] that $\widehat{F}_m(t)=\widehat{F}_{\mathbf{a}^m}(t)$ is the minimax estimator of F_m in the class of all unbiased estimators if

$$\mathbf{a}_{;n}^m=\mathbf{\Gamma}^{-1}\mathbf{p}_{;n}. \tag{5}$$

The weight coefficients $\mathbf{a}_{;n}^m$ are called minimax for the component m in a mixture.

3. Main results

Now we consider the censored data obtained from a mixture with varying concentrations.

Let n objects $O_{j;n}$ be observed and let each of them belong to one of M components (subpopulations). Denote by $\xi_{j;n}=\xi(O_{j;n})>0$ the failure time (the variable under consideration) of the object $O_{j;n}$ and by $C_{j;n}=C(O_{j:n})$ the censoring time for $O_{j;n}$. We assume that $\xi(O)$ and $C(O)$ are independent for a fixed component whom the object O belongs to and that $(\xi_{j;n},C_{j;n})$, $j=1,\dots,n$, are independent if n is fixed. In what follows we consider the asymptotic theory in the scheme of series as $n \to \infty$. We do not assume that there exists a relation between observations for different n.

A censored sample $\mathbf{X}_n = \left(\xi^*_{j;n}, \delta_{j;n}, j = 1, \dots, n\right)$ is observed where

$$\xi^*_{j;n} = \min\left(\xi_{j;n}, C_{j;n}\right)$$

is the censored failure time of the observation j and where

$$\delta_{j;n} = \mathbb{1}\left\{\xi_{j;n} < C_{j;n}\right\}$$

is the indicator of censoring ($\delta_{j;n}$ equals 1 if and only if the observation j is not censored).

Note that the distribution of the censoring time in this model is the same for all observations from the same component but may vary from a component to another component. A different censoring model for the data belonging to a mixture with varying concentrations is considered in [12].

Let $\kappa_{j;n} = \operatorname{ind}(O_{j;n})$ be the index of a component containing the object $O_{j;n}$. We assume that the concentrations of components $p^m_{j;n} = \mathsf{P}\{\kappa_{j;n} = m\}$ are known.

Let

$$F_m(x) = \mathsf{P}\{\xi(O) \le x \mid \operatorname{ind}(O) = m\}$$

be the distribution function of the failure time of an mth component and let

$$G_m(x) = \mathsf{P}\{C(O) \le x \mid \operatorname{ind}(O) = m\}$$

be the distribution function of the censoring time for the mth component.

The survival function constructed from a given distribution function F is denoted by $\bar{F}(t) = 1 - F(t)$.

Let $F(t)$ be a function of bounded variation defined in a measurable subset A of the real line. Put

$$F(A) = \int_A F(dt).$$

This definition for $A =]t_1, t_2]$ reduces to $F(A) = F(t_2) - F(t_1)$. In what follows, we keep this notation for interals A even if the variation of F is unbounded.

The distribution functions F_m and G_m, $m = 1, \dots, M$, are unknown. We introduce the modified Kaplan–Meier estimator for F_k, $1 \le k \le M$. Let

$$\widehat{Y}_{m;n}(t) = \frac{1}{n} \sum_{j=1}^{n} a^m_j \mathbb{1}\{\xi^*_{j:n} \ge t\}$$

be the weighted empirical distribution function of the censored data with weight coefficients $a^k_{j;n}$, and let

$$\widehat{N}_{m;n}(t) = \frac{1}{n} \sum_{j=1}^{n} a^m_j \mathbb{1}\{\xi^*_{j:n} \le t, \delta_{j;n} = 1\}$$

be the weighted empirical distribution function of the uncensored data. Now the modified Kaplan–Meier estimator for $F_k(t)$ is defined by

$$\widehat{F}_{k,n}(t) = 1 - \prod_{j:\xi^*_{j;n} \le t} \left(1 - \frac{\Delta \widehat{N}_{k;n}(\xi^*_{j;n})}{\widehat{Y}_{k;n}(\xi^*_{j;n})}\right) = 1 - \prod_{j:\xi^*_{j;n} \le t} \left(1 - \frac{a^k_j \delta_j}{n - \sum\limits_{i:\xi^*_{i;n} < \xi^*_{j;n}} a^k_i}\right). \tag{6}$$

Note that the consistency of this estimator is proved in [11].

The main result of the current paper is the theorem on the asymptotic normality of the empirical process

$$U_{k;n} = \sqrt{n}\left(\widehat{F}_{k;n}(t) - F_k(t)\right) \tag{7}$$

as an element of the space $D[0,T]$ of functions without discontinuities of the second kind in the interval $[0,T]$, where T is a number in $]0,+\infty[$ such that $F_m(T) < 1$ and $G_m(T) < 1$ for all $m = 1,\dots,M$.

We need some additional notation in order to describe the limit Gaussian process in the consistency theorem.

Let $\xi_{(m)}$ and $C_{(m)}$ be independent random variables whose distributions are F_m and G_m, respectively. These variables are treated as the failure times and censoring moment, respectively, of an object sampled randomly in the component with index m. Then $\xi^*_{(m)} = \min\big(\xi_{(m)}, C_{(m)}\big)$ is the censored failure time and $\delta^*_{(m)} = \mathbb{1}\left\{\xi_{(m)} < C_{(m)}\right\}$ is the indicator of whether or not a randomly sampled object of the component m is censored.

The survival function for the censored failure time of objects of the component m is denoted by

$$Y_m(t) = \mathsf{P}\left\{\xi^*_{(m)} \ge t\right\} = \bar{G}_m(t-)\bar{F}_m(t-).$$

Further let

$$N_m(t) = \mathsf{P}\left\{\xi^*_{(m)} \le t, \delta_{(m)} = 1\right\} = \int_{]0,t]} \bar{G}_m(s-)\, F_m(ds).$$

Then

$$\Lambda_m(t) = \int_{]0,t]} \frac{N_m(dt)}{Y_m(t)}$$

is the integral intensity of failure for objects belonging to the component m. In what follows, we use the notation

$$R_m(A) = N_m(A) - \int_A Y_m(t)\, \frac{N_k(dt)}{Y_k(t)}.$$

Note that $R_m(A)$ depends on the index k of a component for which the modified Kaplan–Meier estimator is constructed. However we omit this index to simplify the notation.

Next we introduce the function ρ that describes the covariance of increments of the Gaussian process Z_k to be used in the construction of the limit of $U_{k;n}$. Given four arbitrary numbers,

$$0 \le u_1 < u_2 \le u_3 < u_4,$$

we put $A_1 =]u_1, u_2]$ and $A_2 =]u_3, u_4]$.

The function ρ is defined by

$$\begin{aligned} \rho(A_1, A_2) = \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle \left[\int_{A_2} Y_m(t_2)\Lambda_k(dt_2) \int_{A_1} \Lambda_k(dt_1) - N_m(A) \int_{A_1} \Lambda_k(dt) \right] \\ - \sum_{m_1,m_2=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^{m_1}\mathbf{p}^{m_2} \right\rangle R_{m_1}(A_1) R_{m_2}(A_2) \end{aligned} \tag{8}$$

and

$$\begin{aligned} \rho(A_1, A_1) = \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle \Big[N_m(A_1) - 2\int_{A_1} N_m(]t, u_2])\Lambda_k(dt) \\ + \int_{A_1\times A_1} Y_m\big(\max(t_1,t_2)\big)\Lambda_k(dt_1)\Lambda_k(dt_2) \Big] \\ - \sum_{m_1,m_2=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^{m_1}\mathbf{p}^{m_2} \right\rangle R_{m_1}(A_1) R_{m_2}(A_1). \end{aligned} \tag{9}$$

Then $Z_k(t)$ is defined as a Gaussian stochastic process with the trajectories belonging to the space $D[0,T]$, with zero expectation, and such that

$$\mathsf{E}\, Z_k(A_1)Z_k(A_2) = \rho(A_1, A_2), \qquad \mathsf{E}\big(Z_k(A_1)\big)^2 = \rho(A_1, A_1).$$

Without loss of generality we assume that $Z_k(0) = 0$. Thus, for $u_1 < u_2$,

$$\begin{aligned} \operatorname{Cov}\big(Z_k(u_1), Z_k(u_2)\big) &= C_Z(u_1, u_2) = \mathsf{E}\, Z_k(]0, u_1])Z_k(]0, u_2]) \\ &= \mathsf{E}\big(Z_k(]0, u_1])\big)^2 + \mathsf{E}\, Z_k(]0, u_1])Z_k(]u_1, u_2]) \\ &= \rho(]0, u_1],]0, u_1]) + \rho(]0, u_1],]u_1, u_2]). \end{aligned} \tag{10}$$

Note that Z_k can also be described as a Gaussian process with zero mean and covariance function (10). It is clear that this description is meaningful only if the right hand side of (10) is a covariance function. We show that this is the case under the assumptions of Theorem 3.1.

Theorem 3.1. *Assume that*

(1) $\det \mathbf{\Gamma} \neq 0$.
(2) $\big\langle (\mathbf{a}^k)\mathbf{p}^{m_1}\mathbf{p}^{m_2}\big\rangle$ *exist for all* $m_1, m_2 = 1, \dots, M$.
(3) $F_m(T) < 1$ *and* $G_m(T) < 1$ *for all* $m = 1, \dots, M$.
(4) *For all* $m = 1, \dots, M$, *the functions* F_m *and* G_m *are continuously differentiable in the interval* $[0,T]$.

Then the empirical processes $U_{k;n}$ *defined by equality* (7) *weakly converge in the space* $D[0,T]$ *equipped with the uniform metric as* $n \to \infty$ *to the limit process* U_k *defined by*

$$U_k(t) = \big(1 - F_k(t)\big) \int_{]0,t]} \frac{Z_k(du)}{Y_k(u)}. \tag{11}$$

Remark 3.1. The process $Z_k(t)$ is continuous almost surely in the interval $[0,T]$, and $1/Y_k(u)$ is a function of bounded variation under the assumptions of Theorem 3.1. Thus the integral in (11) can be viewed as a pathwise Riemann–Stieltjes integral. On the other hand, it also can be defined as the mean square limit of integral sums. This integral is sometimes called the quadratic mean (QM) integral. Both interpretations lead to the same distribution of $U_k(t)$.

QM-integral interpretation (11) allows us to evaluate the variance and covariance function of the process $U_k(t)$. For example, an analogue of the asymptotic version of the Greenwood formula is obtained in Corollary 3.1.

Put

$$\begin{aligned} \sigma_t^2 &= \big(1 - F_k(t)\big)^2 \\ &\times \left[\sum_{m=1}^{M} \left(\big\langle (\mathbf{a}^k)^2 \mathbf{p}^m \big\rangle \int_0^t \frac{N_m'(u)\, du}{(Y_k(u))^2} \right) + \iint_{S_t} \frac{\partial^2 \rho(u_1, u_2)}{\partial u_1 \partial u_2} \frac{du_1\, du_2}{Y_k(u_1) Y_k(u_2)} \right], \end{aligned} \tag{12}$$

where

$$S_t = \{(u_1, u_2) \in [0,t],\ u_1 \neq u_2\}, \qquad N_m'(u) = dN_m(u)/du.$$

Corollary 3.1. *Let assumptions of Theorem* 3.1 *hold. Then, for all* $t \in [0,T]$, *the distribution of* $U_{k;n}(t)$ *weakly converges as* $n \to \infty$ *to the normal distribution with zero mean and variance* σ_t^2.

4. Proofs

The proof of Theorem 3.1 is based on the ideas of the paper [5], where the asymptotic normality of the Kaplan–Meier estimator is obtained for the case of homogeneous censored data. We start with two auxiliary results.

Let T be a fixed positive number. Given a function $f\colon [0,T] \to \mathbb{R}$, we denote the uniform metric by $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$ and the variation of f in $[0,T]$ by $\|f\|_V$.

Lemma 4.1. *Let G_n, G, F_n, and F be some functions mapping the interval $[0,T]$ to $\mathbb{R}$. Assume that these functions have bounded variations and*

(1) *there exists a constant $K_F < \infty$ such that $\|F_n\|_V < K_F$ for all $n = 1, 2, \dots$;*
(2) *there exist two continuous functions f and g mapping $[0,T]$ to $\mathbb{R}$ and for which $g_n = \sqrt{n}(G_n - G) \to g$ and $f_n = \sqrt{n}(F_n - F) \to f$ as $n \to \infty$ in the norm $\|\cdot\|_\infty$.*

Then $\|I_n - I\|_\infty \to 0$ as $n \to \infty$, where

$$I_n(t) = \sqrt{n}\left(\int_0^t G_n(u)\,F_n(du) - \int_0^t G(u)\,F(du)\right),$$

$$I(t) = \int_0^t G(u)\,f(du) + \int_0^t g(u)\,F(du).$$

Remark 4.1. We do not assume that the function f in Lemma 4.1 is of bounded variation. On the other hand, we do assume that f is continuous and $\|G\|_V < \infty$. Thus the integral $\int_0^t G(u)\,f(du)$ exists as the limit of Riemann–Stieltjes integral sums and is equal to

$$G(u)f(u)\big|_0^t - \int_0^t f(u)\,G(du).$$

The existence of integrals $\int_0^T G_n(u)\,F_n(du)$ as the unique limits of Riemann–Stieltjes integral sums does not follow from assumptions of Lemma 4.1, but the statement of the lemma is valid for any pair of partial limits of these sums.

Proof. First we consider

$$J_n(t) = \sqrt{n}\int_0^t \big(G_n(u) - G(u)\big)\big(F_n(du) - F(du)\big)$$

and show that

$$\|J_n\|_\infty \to 0 \quad \text{as } n \to \infty. \tag{13}$$

For all m,

$$\begin{aligned}
|J_n(t)| &= \left|\int_0^t g_n(u)\big(F_n(du) - F(du)\big)\right| \\
&\le \left|\int_0^t \big(g_n(u) - g_m(u)\big)\big(F_n(du) - F(du)\big)\right| + \left|\int_0^t g_m(u)\big(F_n(du) - F(du)\big)\right| \\
&\le \|g_n - g_m\|_\infty \|F_n - F\|_V + \big(2\|g_m\|_\infty + \|g_m\|_V\big)\|F_n - F\|_\infty.
\end{aligned}$$

We used the integration by parts

$$\int_0^t g_m(u)\big(F_n(du) - F(du)\big) = g_m(u)\big(F_n(u) - F(u)\big)\big|_0^t - \int_0^t \big(F_n(u) - F(u)\big)\,g_m(du)$$

when estimating the second term in the latter inequality. Passing to the limit as $n \to \infty$ we obtain

$$\limsup_{n\to\infty} \|J_n\|_\infty \le (K_F + \|F\|_V)\limsup_{n\to\infty} \|g_n - g_m\|_\infty,$$

since $\|g_m\|_\infty < \infty$ and $\|g_m\|_V < \infty$ for a fixed m, whence

$$\|F_n - F\|_\infty = \frac{1}{\sqrt{n}}\,\|f_n\|_\infty \to 0 \quad \text{as } n \to \infty$$

by the second assumption of the lemma.

Taking into account the inequality $\|g_n - g_m\|_\infty \le \|g_n - g\|_\infty + \|g_m - g\|_\infty$, we get

$$\limsup_{n\to\infty} \|J_n\|_\infty \le (K_F + \|F\|_V)\,\|g_m - g\|_\infty$$

for all positive integer numbers m. Passing to the limit as $m \to \infty$ we prove relation (13).

Note that

$$I_n(t) = I_n^1(t) + I_n^2(t), \tag{14}$$

where

$$I_n^1(t) = \sqrt{n}\int_0^t \big(G_n(u) - G(u)\big)\,F_n(du),$$
$$I_n^2(t) = \sqrt{n}\int_0^t G(u)\big(F_n(du) - F(du)\big).$$

Thus

$$\begin{aligned} I_n^2 &= \int_0^t G(u)\,f_n(du) = G(u)f_n(u)\big|_0^t - \int_0^t f_n(u)\,G(du) \\ &\to G(u)f(u)\big|_0^t - \int_0^t f(u)\,G(du) \qquad (15)\\ &= \int_0^t G(u)\,f(du) \end{aligned}$$

as $n \to \infty$ uniformly in $t \in [0,T]$. Relation (13) implies that

$$I_n^1(t) = J_n(t) + \int_0^t g_n(u)\,F(du) \to \int_0^t g(u)\,F(du). \tag{16}$$

Considering (14)–(16) we complete the proof of Lemma 4.1. □

Put

$$\zeta_j(A) = \mathbb{1}\left\{\xi_{j;n}^* \in A, \delta_{j;n} = 1\right\} - \int_A \mathbb{1}\left\{\xi_{j;n}^* > t\right\}\,\Lambda_k(dt),$$
$$Z_{k;n}(A) = \frac{1}{\sqrt{n}}\sum_{j=1} a_{j;n}^k \zeta_j(A)$$

and let $Z_{k;n}(t) = Z_{k;n}(]0,t])$ for $t > 0$.

Lemma 4.2. *Let assumptions of Theorem* 3.1 *hold. Then the process $Z_k(t)$ exists, its trajectories are continuous in $[0,T]$, and the processes $Z_{k;n}(t)$ weakly converge to $Z_k(t)$ as $n \to \infty$ in the space $D[0,T]$ equipped with the norm $\|\cdot\|_\infty$.*

Proof. Consider the random vector $\big(\xi_{(m)}, \delta_{(m)}, \xi_{(m)}^*\big)$ whose distribution coincides with the conditional distribution of $\big(\xi(O), \delta(O), \xi^*(O)\big)$ given $\operatorname{ind}(O) = m$.

Note that

$$\begin{aligned} \mathsf{E}\,Z_{k;n}(A) &= \frac{1}{n}\sum_{j=1}^n\sum_{m=1}^M a_{j;n}^k \left(\mathsf{P}\left\{\xi_{(m)}^* \in A, \delta_{(m)} = 1\right\} - \int_A \mathsf{P}\left\{\xi_{(m)}^* > t\right\}\Lambda_k(dt)\right) \\ &= \sum_{m=1}^M \left\langle \mathbf{a}^k\mathbf{p}\right\rangle_n \left(N_m(A) - \int_A Y_m(t)\frac{N_m(dt)}{Y_k(t)}\right) = 0, \end{aligned}$$

since $\left\langle \mathbf{a}^k \mathbf{p} \right\rangle_n = \mathbb{1}\{k = m\}$.

It is easy to see that

$$\mathsf{E}\, Z_{k;n}(A_1) Z_{k;n}(A_2) = \rho_n(A_1, A_2), \qquad \mathsf{E}(Z_{k;n}(A_1))^2 = \rho_n(A_1, A_1),$$

where

$$\begin{aligned} \rho_n(A_1, A_2) = {} & \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle_n \\ & \quad \times \left[\int_{A_2} Y_m(t_2) \Lambda_k(dt_2) \int_{A_1} \Lambda_k(dt_1) - N_m(A) \int_{A_1} \Lambda_k(dt) \right] \\ & - \sum_{m_1, m_2 = 1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle_n R_{m_1}(A_1) R_{m_2}(A_2) \end{aligned} \tag{17}$$

and

$$\begin{aligned} \rho_n(A_1, A_1) = {} & \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle_n \\ & \quad \times \Bigg[N_m(A_1) - 2 \int_{A_1} N_m(]t, u_2])\, \Lambda_k(dt) \\ & \qquad + \int_{A_1 \times A_1} Y_m\big(\max(t_1, t_2)\big)\, \Lambda_k(dt_1)\, \Lambda_k(dt_2) \Bigg] \\ & - \sum_{m_1, m_2 = 1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle_n R_{m_1}(A_1) R_{m_2}(A_1). \end{aligned} \tag{18}$$

For example, consider

$$\begin{aligned} \mathsf{E}(Z_{k;n}(A_1))^2 & = \frac{1}{n} \sum_{j=1}^{n} \left(a_{j;n}^k \right)^2 \mathsf{E}\big(\zeta_j(A_1) - \mathsf{E}\, \zeta_j(A_1) \big)^2 \\ & = \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle_n \mathsf{E} \left(\zeta_{(m)}(A_1) \right)^2 \\ & \quad - \sum_{m_1, m_2 = 1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^{m_1} \mathbf{p}^{m_2} \right\rangle_n \mathsf{E}\, \zeta_{(m_1)}(A_1)\, \mathsf{E}\, \zeta_{(m_2)}(A_1) \end{aligned}$$

for the proof of (18), where

$$\zeta_{(m)}(A) = \mathbb{1} \left\{ \xi_{(m)}^* \in A, \delta_{(m)} = 1 \right\} - \int_A \mathbb{1} \left\{ \xi_{(m)}^* > t \right\} \Lambda_k(dt).$$

It is obvious that

$$\mathsf{E}\, \zeta_{(m)}(A_1) = R_m(A_1),$$

$$\begin{aligned}
\mathsf{E}(\zeta_{(m)}(A_1))^2 = {} & \mathsf{E}\,\mathbb{1}\left\{\xi^*_{(m)} \in A_1, \delta_{(m)} = 1\right\} \\
& - 2\,\mathsf{E}\int_{A_1} \mathbb{1}\left\{\xi^*_{(m)} \in A_1, \delta_{(m)} = 1\right\} \mathbb{1}\left\{\xi^*_{(m)} > t\right\} \Lambda_k(dt) \\
& + \mathsf{E}\int_{A_1}\int_{A_1} \mathbb{1}\left\{\xi^*_{(m)} > t_1\right\} \mathbb{1}\left\{\xi^*_{(m)} > t_2\right\} \Lambda_k(dt_1)\,\Lambda_k(dt_2) \\
= {} & N_m(A_1) - 2\int_{A_1} N_m(]t, u_2])\,\Lambda_k(dt) \\
& + \int_{A_1\times A_1} Y_m\big(\max(t_1,t_2)\big)\,\Lambda_k(dt_1)\,\Lambda_k(dt_2).
\end{aligned}$$

This implies equality (18). Equality (17) is proved analogously.

This, in particular, means that assumptions of Theorem 3.1 imply the convergence of the covariance function of $Z_{k;n}$ to the function $C_Z(u_1, u_2)$ defined by relation (10). Thus $C_Z(u_1, u_2)$ is the covariance function of some Gaussian stochastic process in $[0, T]$. Under the assumptions of Theorem 3.1, the functions $N_m(t)$, $m = 1, \dots, M$, are continuously differentiable and $Y_k(t)$ is separated from zero in $[0, T]$. In view of definition (9),

$$\mathsf{E}\big(Z_k(t_2) - Z_k(t_1)\big)^2 \le K(t_2 - t_1)^2$$

for some $K < \infty$ and all $t_1, t_2 \in [0, T]$. Applying Kolmogorov's theorem [4, Theorem 7, Section III] we conclude that there exists an almost surely continuous version of the Gaussian process $Z_k(t)$ (denoted by the same symbol).

The asymptotic (as $n \to \infty$) normality of finite dimensional distributions of $Z_{k;n}(t)$ can be derived from the Lindeberg central limit theorem in the same way as this result is proved in [10] for weighted empirical functions. Equality (17) implies that

$$\mathsf{E}\big(Z_{n;k}(t_3) - Z_{n;k}(t_2)\big)\big(Z_{n;k}(t_2) - Z_{n;k}(t_1)\big) \le K(t_3 - t_2)(t_2 - t_1) \le K(t_3 - t_1)^2$$

for some $K < \infty$ and all $0 \le t_1 < t_2 < t_3 \le T$. Recalling the convergence of finite dimensional distributions we obtain the convergence of $Z_{k;n}$ to Z_k in the space $D[0, T]$ equipped with the Skorokhod metric [2, Theorem 13.5]. Since the limit process Z_k is almost surely continuous, the weak convergence in the Skorokhod metric implies the weak convergence in the uniform metric. □

Proof of Theorem 3.1 (*outlined*). Put

$$\begin{aligned}
\widehat{\Lambda}_{k;n}(t) &= \int_0^t \frac{\widehat{N}_{k;n}(du)}{\widehat{Y}_{k;n}(u)}, \\
Z^Y_{k;n}(t) &= \sqrt{n}\left(\widehat{Y}_{k;n}(t) - Y_k(t)\right), \\
Z^N_{k;n}(t) &= \sqrt{n}\left(\widehat{N}_{k;n}(t) - N_k(t)\right).
\end{aligned}$$

Then

$$Z_{k;n}(t) = \int_0^t \left(Z^N_{k;n}(du) - Z^Y_{k;n}(u)\Lambda_k(du)\right).$$

Similarly to the proof of Lemma 4.2, one can show that the process $\big(Z^Y_{k;n}(\cdot), Z^N_{k;n}(\cdot)\big)$ weakly converges to some process $\big(Z^Y_k(\cdot), Z^N_k(\cdot)\big)$ with respect to the uniform metric in the space of functions mapping the interval $[0, T]$ to $\mathbb{R}^2$; this process has no discontinuities of the second kind and its trajectories are almost surely continuous.

Using Lemma 4.1 and the method of a common probability space we conclude that

$$\sqrt{n}\left(\widehat{\Lambda}_{k;n}(t) - \Lambda_k(t)\right) = \sqrt{n}\left(\int_0^t \frac{1}{\widehat{Y}_{k;n}(u)}\,\widehat{N}_{k;n}(du) - \int_0^t \frac{1}{Y_k(u)}\,N_k(du)\right)$$

weakly converges in $[0,T]$ to

$$\int_0^t \left(\frac{1}{Y_k(u)} Z_k^N(du) - \frac{Z_k^Y(u)\, N_k(du)}{(Y_k(u))^2}\right) = \int_0^t \frac{\widetilde{Z}_k(du)}{Y_k(u)},$$

where

$$\widetilde{Z}_k(t) = Z_k^N(t) + \int_0^t Z_k^Y(u)\, \Lambda_k(du).$$

It is clear that $\widetilde{Z}_k(t)$ is the limit of $Z_{k;n}(t)$ and hence Lemma 4.2 yields $\widetilde{Z}_k(t) = Z_k(t)$.

Using equality (66) of [5] we obtain

$$\sqrt{n}\left(\widehat{F}_{k;n}(t) - F_k(t)\right) = \left(1 - F_k(t)\right) \int_0^t \frac{1 - \widehat{F}_{k;n}(u-)}{1 - \widehat{F}_{k;n}(u)} \sqrt{n}\left(\widehat{\Lambda}_{k;n} - \Lambda_k\right)(du).$$

Since $\left\|\widehat{F}_{k;n} - F_k\right\|_\infty \le K\sqrt{\log(n)/n}$ (see [11]), we get

$$\frac{1 - \widehat{F}_{k;n}(u-)}{1 - \widehat{F}_{k;n}(u)} \to 1$$

and

$$\sqrt{n}\left(\widehat{F}_{k;n}(t) - F_k(t)\right) \xrightarrow{\text{w}} \left(1 - F(t)\right) \int_0^t \frac{Z_k(du)}{Y_k(u)}$$

as $n \to \infty$ in the uniform norm. The rigorous proof of this convergence uses the same technique as that in the proof of Lemma 4.1. □

Proof of Corollary 3.1. The integral in (11) is considered as the mean square limit of integral sums

$$J(\mathcal{T}) = \sum_{i=1}^{I} \frac{Z_k(]t_{i-1}, t_i])}{Y_k(t_i)}$$

as $\operatorname{diam}(\mathcal{T}) \to 0$, where $\mathcal{T} = \{0 = t_0 < t_1 < \cdots < t_I = T\}$ is a partition of the interval $[0,1]$ and $\operatorname{diam}(\mathcal{T}) = \sup_{i=1,\dots,I}(t_i - t_{i-1})$. Such QM-integrals are usually defined for processes with independent increments; however, this property fails for Z_k in the case under consideration. The theory of QM-integration with respect to the processes with dependent increments is described in [7, Section 37.3]. In particular, it follows from results of [7, Section 37.3] that

$$\operatorname{Var}\left(\int_0^T \frac{Z_k(dt)}{Y_k(t)}\right) = \lim_{\operatorname{diam}(\mathcal{T})\to 0} \sum_{i\ne j} \frac{\mathsf{E}\, Z_k(]t_{i-1}, t_i])Z_k(]t_{j-1}, t_j])}{Y_k(t_i)Y_k(t_j)} + \lim_{\operatorname{diam}(\mathcal{T})\to 0} \sum_{i=1}^{I} \frac{\mathsf{E}\big(Z_k(]t_{i-1}, t_i])\big)^2}{(Y_k(t_i))^2}. \tag{19}$$

Under the assumptions of Theorem 3.1,

$$\mathsf{E}\, Z_k(]t_{i-1}, t_i])Z_k(]t_{j-1}, t_j]) = \frac{\partial^2 \rho(t_i, t_j)}{\partial t_i \partial t_j} + o\left(\operatorname{diam}(\mathcal{T})^2\right)$$

and

$$\mathsf{E}\big(Z_k(]t_{i-1}, t_i])\big)^2 = \sum_{m=1}^{M} \left\langle (\mathbf{a}^k)^2 \mathbf{p}^m \right\rangle N_m'(t_i) + o\big(\operatorname{diam}(\mathcal{T})\big).$$

This implies equality (12). □

Acknowledgment

The author is indebted to a referee for the attention paid to this manuscript and for valuable remarks.

Bibliography

[1] F. Autin and C. Pouet, *Minimax rates over Besov spaces in ill-conditioned mixture-models with varying mixing-weights*, J. Statist. Plann. Inference **146** (2014), 20–30. MR3132476
[2] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1999. MR1700749
[3] O. V. Doronin, *Adaptive estimation for a semiparametric model of mixture*, Theory Probab. Math. Statist. **91** (2015), 29–41. MR3364121
[4] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes*, vol. 1, "Nauka", Moscow, 1971; English transl., Springer-Verlag, Berlin–Heidelberg–New York, 1974. MR0346882
[5] R. D. Gill and S. Johansen, *A survey of product-integration with a view toward application in survival analysis*, Ann. Statist. **18** (1990), no. 4, 1501–1555. MR1074422
[6] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, Second edition, Wiley, New York, 2003. MR1940115
[7] M. Loéve, *Probability Theory II*, Fourth edition, Springer, New York, 1978. MR0651018
[8] R. E. Maiboroda, *On the estimation of parameters of variable mixtures*, Theory Probab. Math. Statist. **44** (1991), 87–92. MR1168432
[9] R. E. Maiboroda, *Estimates for distributions of components of mixtures with varying concentrations*, Ukrainian Math. J. **48** (1996), no. 4, 618–622. MR1417019
[10] R. Maiboroda and O. Sugakova, *Statistics of mixtures with varying concentrations with application to DNA microarray data analysis*, J. Nonparametr. Stat. **24** (2012), no. 1, 201–215. MR2885834
[11] R. E. Maiboroda and V. G. Khizanov, *A modified Kaplan–Meier estimator for a model of mixtures with varying concentrations*, Theory Probab. Math. Statist. **92** (2016), 109–116. MR3619048
[12] A. Yu. Ryzhov, *Estimates of distributions of components in a mixture from censoring data*, Teor. Imovir. Mat. Statist. **69** (2003), 154–161; English transl. in Theory Probab. Math. Statist. **69** (2004), 167–174. MR2110914
[13] J. Shao, *Mathematical Statistics*, Springer-Verlag, New York, 2003. MR2002723

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrs'ka Street, 64/13, 01601, Kyiv, Ukraine

Email address: mre@univ.kiev.ua

Received 21/DEC/2016

Translated by N. N. SEMENOV