**A TEST OF THE HYPOTHESIS ABOUT THE HOMOGENEITY
OF COMPONENTS OF A MIXTURE
WITH VARYING CONCENTRATIONS
BY USING CENSORED DATA**

UDC 519.21

A. YU. RYZHOV

ABSTRACT. We consider a statistical test for the homogeneity of components in a model of mixtures with varying concentrations by using censored data. The test is constructed with the help of adaptive estimators of components of a mixture.

1. INTRODUCTION

The problem of testing the homogeneity of groups of observations appears often in different areas of statistical analysis. Another example of applications of this problem is presented by survival analysis. In survival analysis a statistician separates the observations and forms two or more samples according to some separating parameter or to a family of separating parameters. Below are some examples of such parameters: an age of a patient, a morphological type, a stage of a disease, a type of treatment, etc. Several methods to test the homogeneity of groups are known in the literature. Among those methods are the log-rank test, the generalized Wilcoxon–Gehan test, the Prentice test, etc. [1]. However, the known tests cannot be used in the case where one does not know *a priori* the population containing a certain individual and where the samples are mixtures of individuals from different populations with known varying concentrations. This is precisely the case considered in this paper. Below we discuss the setting of the problem in more detail.

Let $\Omega_1, \Omega_2, \dots, \Omega_m$ be m given populations. Denote by

$$H_1(t), H_2(t), \dots, H_m(t), \quad t \geq 0,$$

the distribution functions of the lifetime of an individual in the populations $\Omega_1, \Omega_2, \dots, \Omega_m$, respectively. A statistician studies $N \geq m$ samples; each of these samples is a mixture of individuals of the populations $\Omega_1, \Omega_2, \dots, \Omega_m$. Observations are randomly censored on the right; the probability of censoring for every member of a sample j depends on a sample and does not depend on the number m of the population containing this member. For all samples, the numbers $w_1^{(j)}, w_2^{(j)}, \dots, w_m^{(j)}$ are known where $w_l^{(j)}$ stands for the probability that a member of a sample j belongs to a population l . In other words, $w_1^{(j)}, w_2^{(j)}, \dots, w_m^{(j)}$ are concentrations of components in a mixture. The distribution of

2000 *Mathematics Subject Classification.* Primary 62N03; Secondary 62G10.

Key words and phrases. Survival analysis, mixtures with varying concentrations, censoring, Kaplan–Meier estimator, dispersion coefficient.

the lifetime of an arbitrary member of a sample j is given by

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

where

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

and $j = 1, \dots, N$. The models without censoring are studied in [2]. The problem is to test the hypothesis $H_p(t) = H_q(t)$ for some fixed p and q . The test is constructed in Section 2; it uses adaptive estimators of the distribution functions $H_l(t)$, $l = 1, \dots, m$. The proofs of the results are given in Section 3.

2. THE CONSTRUCTION OF THE TEST

Let N samples of observations be given. The distribution function of the lifetime of an arbitrary object of the sample j is given by relation (1) where the concentrations $w_l^{(j)}$ are known. Let ordered pairs of random variables (T_{jk}, U_{jk}) , $1 \leq k \leq n_j$, $1 \leq j \leq N$, denote the finite refusal time and the censoring time, respectively, for a member k of a sample j . Put $n \equiv \sum_{j=1}^N n_j$, where n_j is the size of the sample j . In fact, a statistician observes the random variables

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

and

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

where I_A denotes the indicator of the random event A . Consider the following functions:

$$S_j(t) = \mathbf{P}\{T_{jk} > t\} = 1 - F_j(t),$$

$$C_{jk}(t) = \mathbf{P}\{U_{jk} > t\},$$

$$\pi_{jk}(t) = \mathbf{P}\{X_{jk} > t\}.$$

Definition 2.1. The function $S_j(t)$ is called the *survival function* of the sample j ; the function

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

is called the *cumulative intensity function*.

In what follows we assume that

- a) the random variables U_{jk} have an identical distribution within a fixed sample; that is, the functions $C_{jk}(t)$ do not depend on the number k of a member of a sample if j is fixed;
- b) the random variables T_{jk} and U_{jk} are independent for all $k = 1, \dots, n_j$ and $j = 1, \dots, N$; the corresponding functions $S_j(t) = \mathbf{P}\{T_{jk} > t\}$ and

$$C_j(t) = \mathbf{P}\{U_{jk} > t\}$$

are continuous;

- c) for every j , there are constants $0 < h_j < 1$ such that

$$\frac{n_j}{n} \rightarrow h_j, \quad n \rightarrow \infty;$$

d) the columns of the matrix

$$X = \begin{pmatrix} w_1^{(1)} & w_2^{(1)} & \dots & w_m^{(1)} \\ w_1^{(2)} & w_2^{(2)} & \dots & w_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ w_1^{(N)} & w_2^{(N)} & \dots & w_m^{(N)} \end{pmatrix}$$

are linearly independent.

Condition b) implies that $\pi_{jk}(t) = \pi_j(t) = S_j(t) \cdot C_j(t)$ for all j . Let

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

For every sample, one can construct the Kaplan–Meier estimators $\hat{S}_1(t), \dots, \hat{S}_N(t)$ of the survival functions $S_1(t), \dots, S_N(t)$, respectively (see [3]). If $t \in \Pi_N$, $t < \infty$, and conditions a)–d) hold for $H_l(\cdot)$, then one can also use the following “linear” estimators

$$(2) \quad \hat{H}(s, \mathbf{a}^{(l)}(s)) = \sum_{j=1}^N a_j^{(l)}(s) \hat{F}_j(s), \quad s \in [0, t]$$

(see [4, Theorem 2]), where $\hat{F}_j(s) = 1 - \hat{S}_j(s)$ and $\mathbf{a}^{(l)}(s) = (a_1^{(l)}(s), \dots, a_N^{(l)}(s))^T$ is a nonrandom vector of weight coefficients that are continuous functions on $[0, t]$ and such that

$$(3) \quad \sum_{j=1}^N a_j^{(l)}(s) w_i^{(j)} = \delta_{li}, \quad i = 1, \dots, m.$$

Moreover,

$$\sqrt{n}(\hat{H}(s, \mathbf{a}^{(l)}(s)) - H_l(s)) \Longrightarrow W(\sigma^2(s, \mathbf{a}^{(l)}(s))), \quad n \rightarrow \infty,$$

for $0 \leq s \leq t$, where W is a Wiener process and

$$\sigma^2(s, \mathbf{a}^{(l)}(s)) = \sum_{j=1}^N \frac{1}{h_j} (a_j^{(l)}(s) S_j(s))^2 v_j(s),$$

$$v_j(s) = \int_0^s \frac{d\Lambda_j(u)}{\pi_j(u)}.$$

The optimal linear estimator $\hat{H}_l(s)$, $l = 1, \dots, m$, whose asymptotic dispersion coefficient $\sigma^2(s, \mathbf{a}^{(l)}(s))$ is minimal among estimators of the form (2) corresponds to the case of

$$(4) \quad \mathbf{a}^{(l)}(s) = (V(s)\Gamma^{-1}(s))\mathbf{e}_l,$$

where

$$(5) \quad \Gamma(s) = (\gamma_{pi}(s))_{m \times m},$$

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

$$V(s) = \left(\frac{h_j}{S_j^2(s)v_j(s)} w_i^{(j)} \right)_{j=1}^N_{i=1}^m,$$

$\mathbf{e}_l = (\delta_{l1}, \dots, \delta_{lm})^T$ is the l -th column of the unit matrix, and δ_{ij} is the Kronecker symbol.

A disadvantage of the optimal estimators $\hat{H}_1(s), \dots, \hat{H}_m(s)$ is that their weight coefficients depend on unknown functions $H_1(s), \dots, H_m(s)$. Therefore these estimators are useless in practical problems. On the other hand, one can use *adaptive* estimators

$$\tilde{H}_l(s) = \sum_{j=1}^N \hat{a}_{j,n}^{(l)}(s) \hat{F}_j(s), \quad l = 1, \dots, m,$$

instead of the optimal estimators, where the vector of weight coefficients

$$\hat{\mathbf{a}}_n^{(l)}(s) = (\hat{a}_{1,n}^{(l)}(s), \dots, \hat{a}_{N,n}^{(l)}(s))^T$$

can be found for arbitrary l from the equations

$$(6) \quad \hat{\mathbf{a}}_n^{(l)}(s) = (\hat{V}_n(s) \hat{\Gamma}_n^{-1}(s)) \mathbf{e}_l$$

and

$$\hat{\Gamma}_n(s) = (\hat{\gamma}_{pi}^n(s))_{m \times m},$$

where

$$(7) \quad \begin{aligned} \hat{\gamma}_{pi}^n(s) &= \sum_{j=1}^N \frac{h_j}{\hat{S}_j^2(s) \hat{v}_j(s)} w_i^{(j)} w_p^{(j)}, \\ \hat{V}_n(s) &= \left(\frac{h_j}{\hat{S}_j^2(s) \hat{v}_j(s)} w_i^{(j)} \right)_{j=1}^N \quad m. \end{aligned}$$

Here $\hat{S}_j(s)$ is the Kaplan–Meier estimator of the function $S_j(s)$ constructed from the sample j of size n_j , and the estimators $\hat{v}_j(s)$ are defined by

$$\hat{v}_j(s) = n_j \int_0^s [\{\bar{Y}_j(u) - \Delta \bar{N}_j(u)\} \bar{Y}_j(u)]^{-1} d\bar{N}_j(u),$$

where

$$\begin{aligned} \bar{N}_j(u) &= \sum_{k=1}^{n_j} I_{\{X_{jk} \leq u, \eta_{jk}=1\}}, \\ \bar{Y}_j(u) &= \sum_{k=1}^{n_j} I_{\{X_{jk} \geq u\}}, \\ \Delta \bar{N}_j(u) &= \bar{N}_j(u) - \bar{N}_j(u-). \end{aligned}$$

It is shown in [4, Theorem 3] that the adaptive estimators $\tilde{H}_l(s)$ have asymptotically the same dispersion coefficient as the estimates $\hat{H}_l(s)$. More precisely, if conditions a)–d) hold, then

$$\sqrt{n}(\tilde{H}_l(s) - H_l(s)) \implies W(\sigma^2(s, \mathbf{a}^{(l)}(s))), \quad n \rightarrow \infty,$$

for finite $t \in \Pi_N$ and $s \in [0, t]$, where $\sigma^2(s, \mathbf{a}^{(l)}(s))$ is the asymptotic dispersion coefficient of the optimal estimator $\hat{H}_l(s)$. In what follows we use the adaptive estimators of distribution functions to construct the test.

Consider the hypothesis $H_0: \{H_p(t) = H_q(t) \text{ for all } t \geq 0\}$ for $1 \leq p, q \leq m$. Without loss of generality, one can restrict the consideration to the case of $p = 1$ and $q = 2$. Let $0 = t_0 < t_1 < t_2 < \dots < t_L < \infty$ be $L + 1$ points of the set Π_N . Also let $\tilde{H}_1(t)$ and

$\tilde{H}_2(t)$ be adaptive estimators of the distribution functions $H_1(t)$ and $H_2(t)$, respectively. Consider the random vector

$$\tilde{\xi}_n = \begin{pmatrix} \tilde{\xi}_{1,n} \\ \tilde{\xi}_{2,n} \\ \vdots \\ \tilde{\xi}_{L,n} \end{pmatrix},$$

where

$$(8) \quad \tilde{\xi}_{i,n} = \sqrt{n}(\tilde{H}_1(t_i) - \tilde{H}_2(t_i))$$

for all $1 \leq i \leq L$. We show in Lemma 3.1 that

$$\tilde{\xi}_n \implies \xi, \quad n \rightarrow \infty,$$

where the vector ξ has the multivariate Gaussian distribution with zero mean vector and some covariance matrix $\Sigma = \text{cov } \xi$. Using this result one can prove for nondegenerate matrices Σ that the sequence of random variables $\{\theta_n = \tilde{\xi}_n^T \Sigma^{-1} \tilde{\xi}_n, n \geq 1\}$ has the asymptotic χ^2 distribution with L degrees of freedom. It is natural to use this property when constructing the test; however, the entries of the matrix Σ also depend on unknown functions $H_1(t), \dots, H_m(t)$. One can construct an estimator $\tilde{\Sigma}_n$ of the matrix Σ (see (16)) such that the sequence

$$\{\tilde{\theta}_n = \tilde{\xi}_n^T \tilde{\Sigma}_n^{-1} \tilde{\xi}_n, n \geq 1\}$$

also has the asymptotic χ^2 distribution with L degrees of freedom.

Theorem 2.1. *Assume that conditions a)–d) hold and the functions $v_j(t)$ are increasing for $t \geq 0$ and all $j = 1, \dots, N$. If the hypothesis H_0 is true, then the random variable $\tilde{\theta}_n$, $n \geq 1$, has the asymptotic χ^2 distribution with L degrees of freedom. In other words,*

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\tilde{\theta}_n > h\} = \mathbf{P}\{\chi_L^2 > h\}$$

for all $h \geq 0$.

The test for the hypothesis H_0 is as follows: fix a significance level α and evaluate the $(1 - \alpha)$ -quintile $h_{1-\alpha}^{(L)}$ of the χ^2 distribution with L degrees of freedom. If $\tilde{\theta}_n > h_{1-\alpha}^{(L)}$, then the hypothesis H_0 is rejected; if $\tilde{\theta}_n \leq h_{1-\alpha}^{(L)}$, then H_0 is accepted. According to Theorem 2.1 the significance level of the test approaches α as $n \rightarrow \infty$.

3. PROOF OF RESULTS

Let

$$0 = t_0 < t_1 < t_2 < \dots < t_L < \infty$$

be $L + 1$ points of the set Π_N . Also let $\tilde{H}_1(t)$ and $\tilde{H}_2(t)$ be adaptive estimators of the distribution functions $H_1(t)$ and $H_2(t)$, respectively. Consider the sequence of random vectors

$$\{\tilde{\xi}_n, n \geq 1\}$$

whose components are given by (8) for all n .

Lemma 3.1. *Let conditions a)–d) hold. If $H_1(t) = H_2(t) = H(t)$ for all $t \geq 0$, then*

$$\tilde{\xi}_n \implies \xi, \quad n \rightarrow \infty,$$

where ξ has the multivariate Gaussian distribution with the zero mean vector and some covariance matrix $\Sigma = \text{cov } \xi$.

Proof. The weight coefficients of the estimators $\tilde{H}_1(t)$ and $\tilde{H}_2(t)$ satisfy condition (3), whence

$$H_l(t) = \sum_{j=1}^N \hat{a}_j^{(l)}(t) F_j(t), \quad l = 1, 2.$$

Then

$$\begin{aligned} \tilde{\xi}_{i,n} &= \sqrt{n}(\tilde{H}_1(t_i) - \tilde{H}_2(t_i)) = \sqrt{n}((\tilde{H}_1(t_i) - H(t_i)) - (\tilde{H}_2(t_i) - H(t_i))) \\ &= \sqrt{n}(\tilde{H}_1(t_i) - H_1(t_i)) - \sqrt{n}(\tilde{H}_2(t_i) - H_2(t_i)) \\ &\Rightarrow W_1(\sigma^2(t_i, \mathbf{a}^{(1)}(t_i))) - W_2(\sigma^2(t_i, \mathbf{a}^{(2)}(t_i))) \\ &= \sum_{j=1}^N \frac{1}{\sqrt{h_j}} \left((a_j^{(1)}(t_i) - a_j^{(2)}(t_i)) S_j(t_i) \right) W_j(v_j(t_i)), \quad n \rightarrow \infty, \end{aligned}$$

for all i . The convergence above follows from Theorem 3 in [4]. Thus any random variable $\tilde{\xi}_{i,n}$ has the asymptotic Gaussian distribution with zero mean and variance

$$\sum_{j=1}^N \frac{1}{h_j} \left((a_j^{(1)}(t_i) - a_j^{(2)}(t_i)) S_j(t_i) \right)^2 v_j(t_i).$$

Applying the Cramér–Wald method (see [5]) one can show that the sequence of vectors $\tilde{\xi}_n$ converges in distribution to the Gaussian vector ξ defined by

$$\xi = \begin{pmatrix} \sum_{j=1}^N \frac{1}{\sqrt{h_j}} \left((a_j^{(1)}(t_1) - a_j^{(2)}(t_1)) S_j(t_1) \right) W_j(v_j(t_1)) \\ \vdots \\ \sum_{j=1}^N \frac{1}{\sqrt{h_j}} \left((a_j^{(1)}(t_L) - a_j^{(2)}(t_L)) S_j(t_L) \right) W_j(v_j(t_L)) \end{pmatrix}. \quad \square$$

Now we find conditions that the matrix Σ is nonsingular.

Lemma 3.2. *Let all the assumptions of Lemma 3.1 hold. If the functions $v_j(t)$, $t \geq 0$, are increasing for all $j = 1, \dots, N$, then the matrix $\Sigma = \text{cov } \xi$ is nonsingular.*

Proof. Consider the vector

$$\psi = \left(\underbrace{W_1(v_1(t_1)), \dots, W_N(v_N(t_1))}_N, \dots, \underbrace{W_1(v_1(t_L)), \dots, W_N(v_N(t_L))}_N \right)^T$$

containing $L \cdot N$ components, and the matrix

$$B_{L \times L \cdot N} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1N} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & b_{21} & \cdots & b_{2N} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & b_{L1} & \cdots & b_{LN} \end{pmatrix},$$

where

$$b_{ij} = \frac{1}{\sqrt{h_j}} (a_j^{(1)}(t_i) - a_j^{(2)}(t_i)) S_j(t_i), \quad i = 1, \dots, L, \quad j = 1, \dots, N.$$

Then $\xi = B\psi$, whence

$$(9) \quad \Sigma = \text{cov } \xi = B \text{cov } \psi B^T.$$

Furthermore,

$$\text{cov}(W_j(v_j(s)), W_j(v_j(t))) = v_j(s)$$

for all j and $0 \leq s < t$. Moreover, $\text{cov}(W_k(v_k(t)), W_j(v_j(t))) = 0$ for all $t \geq 0$ and $k \neq j$, since the estimators $\hat{F}_j$ and $\hat{F}_k$ are independent. Thus

$$\text{cov } \psi = \begin{pmatrix} v_1(t_1) & \cdots & 0 & v_1(t_1) & \cdots & 0 & \cdots & v_1(t_1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & v_N(t_1) & 0 & \cdots & v_N(t_1) & \cdots & 0 & \cdots & v_N(t_1) \\ v_1(t_1) & \cdots & 0 & v_1(t_2) & \cdots & 0 & \cdots & v_1(t_2) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & v_N(t_1) & 0 & \cdots & v_N(t_2) & \cdots & 0 & \cdots & v_N(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ v_1(t_1) & \cdots & 0 & v_1(t_2) & \cdots & 0 & \cdots & v_1(t_L) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & v_N(t_1) & 0 & \cdots & v_N(t_2) & \cdots & 0 & \cdots & v_N(t_L) \end{pmatrix}.$$

This matrix is symmetric, and therefore there exists a nonsingular square matrix C such that

$$C \text{cov } \psi C^T = D,$$

where $D = \text{diag}(d_{11}, \dots, d_{NL, NL})$ is a diagonal matrix. The matrices C and D are easy to evaluate in this case. Namely, the entries of the matrix $C = (c_{ij})$, $i, j = 1, \dots, NL$, are such that

$$c_{ij} = \begin{cases} 1, & i = j, \\ -1, & i - j = N, \\ 0, & \text{otherwise,} \end{cases}$$

that is,

$$(10) \quad C = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & -1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

whence

$$D = \begin{pmatrix} v_1(t_1) & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & v_N(t_1) & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & v_1(t_2) - v_1(t_1) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & v_2(t_2) - v_2(t_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & v_N(t_L) - v_N(t_{L-1}) \end{pmatrix}.$$

It is easy to check that the entries of the inverse matrix C^{-1} are given by

$$c_{ij}^{-1} = \begin{cases} 1, & (i-j) = kN, \ k = 0, 1, \dots, L-1, \\ 0, & \text{otherwise} \end{cases}$$

and thus $(C^T)^{-1} = (C^{-1})^T$. Hence

$$\text{cov } \boldsymbol{\psi} = C^{-1}D(C^{-1})^T.$$

Substituting this expression into (9) we get

$$(11) \quad \Sigma = (BC^{-1})D(BC^{-1})^T.$$

This means that Σ is the Gram matrix of rows of the matrix $BC^{-1}D^{1/2}$. If the rows are linearly independent, then the matrix Σ is nonsingular. By this condition, all the functions $v_j(t)$ are increasing; thus the diagonal entries of the matrix D are nonzero and thus the rows are linearly independent. Since the matrix C^{-1} is nonsingular,

$$\text{rk}(BC^{-1}D^{1/2}) = \text{rk } B \leq L.$$

The inequality is strict if and only if the matrix B contains nonzero rows. We show below that the matrix B does not contain zero rows, which implies that Σ is nonsingular. Assume the converse; that is, assume that there is $1 \leq i_0 \leq L$ such that

$$(12) \quad b_{i_0 j} = 0$$

for all $j = 1, \dots, N$. By definition,

$$b_{i_0 j} = \frac{1}{\sqrt{h_j}}(a_j^{(1)}(t_{i_0}) - a_j^{(2)}(t_{i_0}))S_j(t_{i_0}).$$

The points $t_1, \dots, t_L$ belong to the set Π_N ; thus $S_j(t_{i_0}) > 0$. Condition c) implies that $h_j > 0$ for all $j = 1, \dots, N$ and thus equality (12) coincides with

$$(13) \quad a_j^{(1)}(t_{i_0}) = a_j^{(2)}(t_{i_0}).$$

Relation (4) implies that the vectors $\mathbf{a}^{(1)}(t)$ and $\mathbf{a}^{(2)}(t)$ of the optimal weight coefficients are the first and second columns of the matrix $V(t)\Gamma^{-1}(t)$, respectively. Thus equality (13) means that these vectors coincide at the point t_{i_0} . Denote by $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ the first and second columns of the matrix $\Gamma^{-1}(t_{i_0})$, respectively. We prove that equality (13) is equivalent to

$$V(t_{i_0})\boldsymbol{\alpha}_1 = V(t_{i_0})\boldsymbol{\alpha}_2$$

or to

$$(14) \quad V(t_{i_0})(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0.$$

Consider the matrix

$$(15) \quad Q(t) = \begin{pmatrix} \frac{h_1}{S_1^2(t)v_1(t)} & 0 & \cdots & 0 \\ 0 & \frac{h_2}{S_2^2(t)v_2(t)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{h_N}{S_N^2(t)v_N(t)} \end{pmatrix}.$$

Then $V(t) = Q(t)X$. Using this equality we rewrite (14) as follows:

$$Q(t_{i_0})X(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0.$$

Since all the entries of the matrix $Q(t_{i_0})$ are positive, the latter equality is equivalent to

$$X(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0.$$

Note that the columns of X are linearly independent by condition d); thus the zero vector is the only solution of the latter system of equations, whence

$$\alpha_1 = \alpha_2.$$

Since the matrix $\Gamma^{-1}(t_{i_0})$ is nonsingular, $\alpha_1 \neq \alpha_2$. This contradiction proves that the matrix B does not contain zero rows; thus its rank is equal to L . Therefore the matrix Σ is nonsingular. $\square$

We explained above that the matrix Σ is useless for practical purposes. On the other hand, one can approach Σ by a sequence of estimators $\{\tilde{\Sigma}_n, n \geq 1\}$. For $t \in \Pi_N$, $t < \infty$, and $0 \leq s \leq t$ let

- $\hat{\mathbf{a}}_n^{(1)}(s)$ and $\hat{\mathbf{a}}_n^{(2)}(s)$ be the vectors of weight coefficients for adaptive estimators $\tilde{H}_1(s)$ and $\tilde{H}_2(s)$, respectively;
- $\hat{S}_1(s), \dots, \hat{S}_N(s)$ be the Kaplan–Meier estimators of the functions

$$S_1(s), \dots, S_N(s);$$

- for all $j = 1, \dots, N$, the function

$$\hat{v}_j(s) = n_j \int_0^s [\{\bar{Y}_j(u) - \Delta \bar{N}_j(u)\} \bar{Y}_j(u)]^{-1} d\bar{N}_j(u)$$

is the estimator of the function $v_j(s)$.

It is shown in [6] that the estimators $\hat{S}_j(\cdot)$ and $\hat{v}_j(\cdot)$ are consistent in the interval $[0, t]$. Let t be such that $0 = t_0 < t_1 < t_2 < \dots < t_L \leq t$ and put

$$\hat{B}_n = \begin{pmatrix} \hat{b}_{11} & \hat{b}_{12} & \dots & \hat{b}_{1N} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \hat{b}_{21} & \dots & \hat{b}_{2N} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & \hat{b}_{L1} & \dots & \hat{b}_{LN} \end{pmatrix},$$

where

$$\hat{b}_{ij} = \frac{1}{\sqrt{\hat{h}_j}} (\hat{a}_{j,n}^{(1)}(t_i) - \hat{a}_{j,n}^{(2)}(t_i)) \hat{S}_j(t_i), \quad i = 1, \dots, L, \quad j = 1, \dots, N,$$

and

$$\hat{D}_n = \begin{pmatrix} \hat{v}_1(t_1) & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \hat{v}_N(t_1) & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \hat{v}_1(t_2) - \hat{v}_1(t_1) & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \hat{v}_2(t_2) - \hat{v}_2(t_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & \hat{v}_N(t_L) - \hat{v}_N(t_{L-1}) \end{pmatrix}.$$

Then the members of the desired sequence of matrices are given by

$$(16) \quad \tilde{\Sigma}_n = (\hat{B}_n C^{-1}) \hat{D}_n (\hat{B}_n C^{-1})^T, \quad n \geq 1,$$

where the matrix C is defined by (10).

Lemma 3.3. *If all assumptions of Lemma 3.2 hold, then*

$$\tilde{\Sigma}_n^{-1} \xrightarrow{P} \tilde{\Sigma}^{-1}, \quad n \rightarrow \infty.$$

Remark. The convergence in probability of a sequence of matrices means that every entry of matrices converges in probability.

Proof. First we prove that

$$\tilde{\Sigma}_n \xrightarrow{P} \Sigma, \quad n \rightarrow \infty.$$

Since the estimators $\hat{v}_j(\cdot)$ and $\hat{S}_j(\cdot)$, $j = 1, \dots, N$, are consistent in the interval $[t_1, t_L]$, they converge in probability at any point s of this interval. Thus

$$\hat{D}_n \xrightarrow{P} D, \quad n \rightarrow \infty,$$

and

$$(17) \quad \hat{V}_n(s) \xrightarrow{P} V(s), \quad \hat{\Gamma}_n(s) \xrightarrow{P} \Gamma(s), \quad n \rightarrow \infty,$$

where the matrices $V(s)$ and $\Gamma(s)$ are defined in (5) and the matrices $\hat{V}_n(s)$ and $\hat{\Gamma}_n(s)$ are defined in (7). Since the entries of the matrices $\{\hat{\Gamma}_n(s), n \geq 1\}$ converge in probability to the corresponding entries of the matrix $\Gamma(s)$ and $\det \Gamma(s) \neq 0$,

$$(18) \quad \hat{\Gamma}_n^{-1}(s) \xrightarrow{P} \Gamma^{-1}(s), \quad n \rightarrow \infty.$$

It follows from (17) and (18) that

$$(\hat{V}_n(s) \hat{\Gamma}_n^{-1}(s)) \xrightarrow{P} (V(s) \Gamma^{-1}(s)), \quad n \rightarrow \infty.$$

Now we use equality (6) and obtain for $l = 1, 2$ that

$$\hat{\mathbf{a}}_n^{(l)}(s) \xrightarrow{P} \mathbf{a}^{(l)}(s), \quad n \rightarrow \infty.$$

This means that

$$\hat{a}_{j,n}^{(1)}(s) \xrightarrow{P} a_j^{(1)}(s), \quad \hat{a}_{j,n}^{(2)}(s) \xrightarrow{P} a_j^{(2)}(s), \quad n \rightarrow \infty,$$

for all $j = 1, \dots, N$. Thus

$$\hat{b}_{ij} \xrightarrow{P} b_{ij}, \quad n \rightarrow \infty,$$

for all $i = 1, \dots, L$ and $j = 1, \dots, N$ where

$$b_{ij} = \frac{1}{\sqrt{h_j}} (a_j^{(1)}(t_i) - a_j^{(2)}(t_i)) S_j(t_i).$$

Therefore,

$$\hat{B}_n \xrightarrow{P} B, \quad n \rightarrow \infty,$$

and the sequence of matrices

$$\tilde{\Sigma}_n = (\hat{B}_n C^{-1}) \hat{D}_n (\hat{B}_n C^{-1})^T$$

converges in probability to the matrix Σ . Since the latter matrix is nonsingular, the sequence of inverse matrices $\tilde{\Sigma}_n^{-1}$ converges in probability to Σ^{-1} . $\square$

Finally we prove Theorem 2.1.

Proof. Consider the sequence

$$\{\tilde{\theta}_n = \tilde{\xi}_n^T \tilde{\Sigma}_n^{-1} \tilde{\xi}_n, n \geq 1\},$$

where the matrix $\tilde{\Sigma}_n$ is defined in (16). It is known (see Problem 87 in [7]) that if

$$\left\{ G_n = (g_{jk}^{(n)})_{j,k=1}^p, n \geq 1 \right\}$$

is a sequence of $p \times p$ random matrices (that is, their entries are random variables) such that

- for all j and k , the random variables $g_{jk}^{(n)}$ converge in probability to constants g_{jk} as $n \rightarrow \infty$;
- $G = (g_{jk})_{j,k=1}^p$;

- the sequence $\{\alpha_n, n \geq 1\}$ of random variables of $\mathbb{R}^p$ converges to a random variable α in distribution,

then the sequence

$$\{G_n \alpha_n, n \geq 1\}$$

converges in distribution to $G\alpha$. Put $p = L$, $G_n = \tilde{\Sigma}_n^{-1}$, and $\alpha_n = \tilde{\xi}_n$. Applying Lemmas 3.1–3.3 we prove that

$$\tilde{\Sigma}_n^{-1} \tilde{\xi}_n \implies \Sigma^{-1} \xi, \quad n \rightarrow \infty,$$

whence

$$\tilde{\theta}_n = \tilde{\xi}_n^T \tilde{\Sigma}_n^{-1} \tilde{\xi}_n \implies \xi^T \Sigma^{-1} \xi, \quad n \rightarrow \infty.$$

Now we show that the random variable $\xi^T \Sigma^{-1} \xi$ has the χ^2 distribution with L degrees of freedom. Since Σ is nonsingular and symmetric, the square root of the inverse matrix $\Sigma^{-1/2}$ exists and

$$(\Sigma^{-1/2})^T = \Sigma^{-1/2}.$$

The vector ξ contains L components and

$$\xi^T \Sigma^{-1} \xi = (\Sigma^{-1/2} \xi)^T (\Sigma^{-1/2} \xi).$$

On the other hand,

$$\begin{aligned} \mathbb{E}(\Sigma^{-1/2} \xi)(\Sigma^{-1/2} \xi)^T &= \mathbb{E}\left(\Sigma^{-1/2} (\xi \cdot \xi^T) \Sigma^{-1/2}\right) = \Sigma^{-1/2} \mathbb{E}(\xi \cdot \xi^T) \Sigma^{-1/2} \\ &= \Sigma^{-1/2} \Sigma \Sigma^{-1/2} = E. \end{aligned}$$

Thus $\Sigma^{-1/2} \xi$ has the multivariate distribution with zero mean vector and unit covariance matrix. Hence

$$\xi^T \Sigma^{-1} \xi \sim \chi_L^2,$$

and the proof of Theorem 2.1 is complete. $\square$

BIBLIOGRAPHY

1. D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman & Hall, New York, 1984. MR0751780 (86f:62170)
2. R. E. Maiboroda, *Statistical Analysis of Mixtures*, Kyiv University, Kyiv, 2003. (Ukrainian)
3. E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457–481. MR0093867 (20:387)
4. A. Yu. Ryzhov, *Estimates of distributions of components in a mixture from censoring data*, Teor. Imovir. ta Matem. Statist. **69** (2003), 140–147; English transl. in Theory Probab. Math. Statist. **69** (2004), 167–174. MR2110914 (2005i:62138)
5. P. Billingsley, *Convergence of Probability Measures*, J. Wiley & Sons, Inc., New York, 1999. MR1700749 (2000e:60008)
6. D. P. Harrington and T. R. Fleming, *Counting Processes and Survival Analysis*, J. Wiley & Sons, Inc., New York, 1991. MR1100924 (92i:60091)
7. A. Ya. Dorogovtsev, D. S. Silvestrov, A. V. Skorokhod, and M. I. Yadrenko, *Probability Theory: Collection of Problems*, “Vyshcha Shkola”, Kyiv, 1976; English transl., American Mathematical Society, Providence, RI, 1997. MR1444579 (98k:60001)

DEPARTMENT OF GENERAL MATHEMATICS, FACULTY FOR MATHEMATICS AND MECHANICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE 6, KYIV 03127, UKRAINE

E-mail address: ryzhov@univ.kiev.ua

Received 5/JUL/2004

Translated by S. KVASKO