KYBERNETIKA - VOLUME 31 (1995), PAGES 45 - 64

ON CONSISTENCY OF THE MLE

František Rublík[1]

Convergence of the maximum likelihood estimator is established without the assumption that the true value of the parameter belongs to the null hypothesis Ω_0. It is shown, that the MLE exists with probability tending to 1, and that the distance of the MLE from a set H of parameters from Ω_0 tends to zero almost everywhere, where H are the parameters of the probabilities best fitting the true distribution in the sense that they maximize the mean of logarithm of the likelihood function.

1. INTRODUCTION AND THE MAIN RESULTS

It is well known that the MLE is consistent if the true parameter belongs to the null hypothesis and certain regularity conditions are fulfilled, and some consistency results can be proved also in the case, when the theoretical model is misspecified. Let us mention some papers, whose results or methods are related to those in this paper. Compactness of the parametric set is an essential condition for consistency of the MLE in the paper [11]. Consistency for exponential families is investigated in [1], existence and uniqueness of the unrestricted MLE is a topic of [8]. Consistency of the MLE for Markov processes is under misspecification assumption established in [4]. A misspecified i.i.d. case is treated in [12], and consistency of MLE in a mispecified model is under general conditions on dependence of observations established in [13] and [3]. The aim of this paper is to establish convergence of the MLE in the setting admitting misspecification of the model, and with emphasize on the case of independent sampling from finitely many statistical populations.

We shall assume, that $(S, \mathcal{S})$ is a measurable space, Θ is a parameter set, $\{\mathcal{F}_u\}_{u=1}^{\infty}$ are sub-σ-algebras of $\mathcal{S}$, ν_u is a measure on $\mathcal{F}_u$ and $\{P_\theta^{(u)}, \theta \in \Theta\}$ are probabilities on $(S, \mathcal{F}_u)$ defined by means of the densities

$$f_u(s, \theta) = \frac{\mathrm{d}P_\theta^{(u)}}{\mathrm{d}\nu_u}(s)\,. \tag{1.1}$$

To ensure measurability of the likelihood function we shall impose the following assumption.

[1]This research was supported by a grant from the Slovak Academy of Sciences n. 999366.

(**A 1**) Θ *is a separable metric space endowed with a metric* ρ, *and* $f_u(s,\theta)$ *is a continuous function of* $\theta \in \Theta$ *for each* $s \in S$.

We shall admit a possible misspecification of the model, but the true probability P, defined on $(S, \mathcal{S})$, will be subjected to some regularity assumptions. In these for $\Omega \subset \Theta$ and $s \in S$ we put

$$L_u(s,\Omega) = \sup\{f_u(s,\theta);\ \theta \in \Omega\}\,. \tag{1.2}$$

(**A 2**) *There exist norming constants* $\{n_u\}_{u=1}^{\infty}$ *such that for each* $\theta^* \in \Theta$

$$\lim_{u\to\infty} \frac{1}{n_u} \log f_u(s,\theta^*) = I(\theta^*) \tag{1.3}$$

P *almost everywhere, and* I *is a continuous function of* θ^*.

(**A 3**) *For each* $\theta^* \in \Theta$ *there exists a positive number* $\Delta^* = \Delta^*(\theta^*)$ *such that for any arbitrary fixed number* $\Delta \in (0, \Delta^*\rangle$ *and for the set*

$$V(\theta^*,\Delta) = \{\tilde\theta \in \Theta;\ \rho(\tilde\theta,\theta^*) < \Delta\} \tag{1.4}$$

one can find an $\mathcal{F}_u$ *measurable function* $g_u(s,\theta^*,\Delta)$ *of the argument* s *such that on* S *for the constants* $\{n_u\}_{u=1}^{\infty}$ *from (A 2)*

$$\frac{1}{n_u} \log L_u(s, V(\theta^*,\Delta)) \le g_u(s,\theta^*,\Delta) \tag{1.5}$$

and

$$\lim_{u\to\infty} g_u(s,\theta^*,\Delta) = I(\theta^*,\Delta)\,, \tag{1.6}$$

P *a.e., where with the notation from (A 2)*

$$\lim_{\Delta\to 0^+} I(\theta^*,\Delta) = I(\theta^*)\,. \tag{1.7}$$

(**A 4**) *Let* $\{n_u\}_{u=1}^{\infty}$ *be the norming constants from (A 2). If* c *is a real number, then there exists a compact set* $\Gamma \subset \Theta$ *such that*

$$\limsup_{u\to\infty} \frac{1}{n_u} \log L_u(s, \Theta - \Gamma) < c \tag{1.8}$$

P a.e. .

In the following theorem we shall use for $\Omega \subset \Theta$ in accordance with (A 2) the notation

$$I(\Omega_0) = \sup\{I(\theta^*);\ \theta^* \in \Omega_0\}\,. \tag{1.9}$$

Theorem 1.1. *Let us assume that the regularity conditions (A 1) - (A 3) hold and the set Θ is either compact or (A 4) is fulfilled. Further we suppose that $\Omega_0 \subset \Theta$ is a closed set and*

$$I(\theta^*) > -\infty \tag{1.10}$$

for some $\theta^* \in \Omega_0$.

(I) The set

$$H = \{\theta^* \in \Omega_0;\, I(\theta^*) = I(\Omega_0)\} \tag{1.11}$$

is compact and non-empty.

(II) There exist mappings

$$\tilde{\theta}_u : S \longrightarrow \Omega_0 \tag{1.12}$$

measurable with respect to $\mathcal{F}_u$ and such that

$$P\Big[\, L_u(s,\Omega_0) = L_u(s,\tilde{\theta}_u(s)) \text{ for all } u \geq u(s) \,\Big] = 1\,, \tag{1.13}$$

with $L_u(s,\Omega_0) = L_u(s,\tilde{\theta}_u(s))$ for all $s \in S$ in the case when Ω_0 is compact. If (1.12) are any measurable mappings satisfying (1.13), then the random variables

$$\rho(\tilde{\theta}_u, H) = \inf\{\, \rho(\tilde{\theta}_u(s),\theta^*); \theta^* \in H\,\} \tag{1.14}$$

converge to zero P a.e. .

To pronounce the Corollary 1.1 we shall suppose that ν is a measure on $(X,\mathcal{F})$ and $\{f_\theta(x|y);\, \theta \in \Theta\}$ is a family of transition density functions (with respect to ν) of transition probabilities $P_\theta(A|y)$ of a Markov process.

(**A*** **1**) *Θ is a separable metric space endowed with a metric ρ, and $f_\theta(x|y)$ is a continuous function of $\theta \in \Theta$ for all $x,\ y \in X$.*

Let us further suppose that $X_1, X_2, \ldots$ is a process whose true distribution P is such that the law of large numbers holds for the sequencies $\{\xi(X_i, X_{i+1})\}_{i=1}^{\infty}$ with finite $\mathrm{E}(\xi(X_1,X_2))$ (this is for Markov process according to remark 2.2 in [4] guaranteed by the validity of the condition (A_1) in [4]). Obviously, with the notation $S = X^\infty$, $\mathcal{S} = \mathcal{F}^\infty$, $\mathcal{F}_n = \mathcal{F} * \ldots * \mathcal{F}$, $s = \{x_n\}_{n=1}^{\infty} \in S$ and

$$f_n(s,\theta) = f_n(x_1,\ldots,x_n,\theta) = f(x_1)\prod_{i=1}^{n-1} f_\theta(x_{i+1}|x_i) \tag{1.15}$$

the assumption (A 1) is fulfilled, provided that *(A* 1)* holds. As observed in [2] p.4, when instead of $f(x)$ some $f(x,\theta)$ is used in (1.15), the information about the initial density $f(x,\theta)$ does not increase with n, and [2] therefore uses as log-likelihood the function

$$L(\theta) = \sum_{i=1}^{n-1} \log f_\theta(x_{i+1}|x_i)\,, \tag{1.16}$$

which is also the approach used in [4]. The initial term $f(x)$ is in (1.15) included to fulfill the requirement (postulated at the begining of the paper) that $\{f_u\}$ are probability densities, and simultaneously to ensure that the MLE based on (1.15) is the same as the one based on (1.16).

Putting $n_u = n$ and denoting

$$\begin{aligned} L(x_i, x_{i+1}, V) &= \sup\{f_\theta(x_{i+1}|x_i);\ \theta \in V\} \\ g_n(s, \theta^*, \Delta) &= \frac{1}{n}\Big[\log f(x_1) + \sum_{i=1}^{n-1} \log L\Big(x_i, x_{i+1}, V(\theta^*, \Delta)\Big)\Big], \end{aligned}$$

we see that (owing to the postulated ergodicity) (A 2), (A 3) follow from *(A^* 1)* and the following assumption, where E denotes expectation with respect to the true distribution P.

(**A* 2**) *For each $\theta^* \in \Theta$ there exists a number $\Delta^* = \Delta^*(\theta^*)$ such that for the set (1.4)*

$$E\Big[\max\{0, \log L(X_1, X_2, V(\theta^*, \Delta))\}\Big] < +\infty \tag{1.17}$$

and $I(\theta^*) = E(\log f_{\theta^*}(X_2|X_1))$ is a continuous function of $\theta^* \in \Theta$.

From Theorem 1.1 we therefore immediately obtain the following assertion.

Corollary 1.1. *Let us assume that the assumptions (A^* 2) and (A^* 1) are fulfilled, the set Ω_0 is closed in Θ, and for some $\theta^* \in \Omega_0$ the inequality (1.10) holds. Then the assertions (I) and (II) of Theorem 1.1 remain true provided that either Θ is compact or the assumption (A 4) holds.*

We remark, that on the one hand the asssumptions used in the Corollary 1.1 are less restrictive, than the assumptions of [4] imposing on the functions $\log f_\theta(x|y)$ and on their expectations some differentiability conditions, not required in (A^*2). On the other hand, the condition (A^*2) requires validity of (1.17) for all $\theta^* \in \Theta$, while (A_6) of [4] requires uniform convergence of $\frac{1}{n}\sum_{i=1}^{n-1} \log f_\theta(X_{i+1}|X_i)$ on a neighbourhood V^* only for the unique $\theta^* \in \Theta$, postulated in [4] to maximize $\mathrm{E}(\log f_\theta(X_2|X_1))$. However, [4] does not deal with the general case $\Omega_0 = \Theta \cap C$, where C is a closed subset of R^k, while the Corollary 1.1 guarantees consistency of MLE for any closed subset of the parameter space.

A more general framework is used in [13] and [3], where the conditional densities are allowed to vary for all $t = 1, 2, \ldots$, and the limit in (1.3) may not exist in such a case. In these papers convergence to zero of $\hat\theta_n - \theta_n^*$ is established under different conditions, with $\hat\theta_n$ being the MLE and θ_n^* denoting the postulated maximizer θ_n^* of the mean of logarithm of likelihood function of n observations; the parameter space is in [13] assumed to be compact, and [3] deals with the case $\Omega_0 = B$, where B is an open subset of R^p. Thus while in [13] and [3] dependence of observations may be of a type not allowed by (A 1) - (A 4), these conditions allow to establish consistency of MLE for types of the null hypotheses, not included in [13] or [3].

The main goal of this paper is to prove consistency of the MLE in the case when inference is based on independent random samples from q populations.

Let $\{\overline{P}_\gamma;\ \gamma \in \Xi\}$ be a family of probability measures, defined on $(X, \mathcal{F})$ by means of the densities

$$f(x, \gamma) = \frac{\mathrm{d}\overline{P}_\gamma}{\mathrm{d}\nu}(x) \tag{1.18}$$

with respect to a measure ν. Let us denote the q-fold products

$$S = X^\infty \times \ldots \times X^\infty\ , \qquad \mathcal{S} = \mathcal{F}^\infty * \ldots * \mathcal{F}^\infty\ , \qquad \Theta = \Xi^q\,. \tag{1.19}$$

If $\theta = (\theta_1, \ldots, \theta_q) \in \Theta$, then θ_j is the parameter assigned to the j-th population. We shall asssume, that statistical conclusions about θ are for

$$s = (\{x_n^{(1)}\}_{n=1}^\infty, \ldots, \{x_n^{(q)}\}_{n=1}^\infty) \in S$$

based on

$$s_u = x^{(u)} = (y(1, n_u^{(1)}), \ldots, y(q, n_u^{(q)})), \tag{1.20}$$

where $y(1, n_u^{(1)}), \ldots, y(q, n_u^{(q)}))$ are independent random vectors, and

$$y(j, n_u^{(j)}) = (x_1^{(j)}, \ldots, x_{n_u^{(j)}}^{(j)})$$

is a random sample (with the sample size $n_u^{(j)}$) from the j-th population. In this q-sample case the σ-algebras $\mathcal{F}_u$ and the norming constants n_u are determined with

$$\mathcal{F}_u = \mathcal{F}^{n_u^{(1)}} * \ldots * \mathcal{F}^{n_u^{(q)}}\ , \qquad n_u = n_u^{(1)} + \ldots + n_u^{(q)}\,, \tag{1.21}$$

and the densities on which the MLE is based

$$f_u(s, \theta) = f_u(x^{(u)}, \theta) = \prod_{j=1}^{q} \prod_{i=1}^{n_u^{(j)}} f(x_i^{(j)}, \theta_j)\ . \tag{1.22}$$

(**RA 1**) *Ξ is a separable metric space endowed with a metric ρ, and the function $f(x, .)$ is continuous on Ξ for each $x \in X$.*

The true distributions $P_1, \ldots, P_q$ of the q underlying populations will be subjected to the following regularity assumptions.

(**RA 2**) *For each $\gamma \in \Xi$ there exists a positive real number $\Delta^* = \Delta^*(\gamma)$ such that for the set*

$$V(\gamma, \Delta^*) = \{\tilde{\gamma} \in \Xi;\ \rho(\tilde{\gamma}, \gamma) < \Delta^*\} \tag{1.23}$$

with the notation $L(x, V) = \sup\{f(x, \tilde{\gamma});\ \tilde{\gamma} \in V\}$ the inequality

$$\int \max\{0, \log L(x, V(\gamma, \Delta^*))\}\, dP_j(x) < +\infty \tag{1.24}$$

holds for $j = 1, \ldots, q$.

(**RA 3**) *The integral*

$$I_j(\gamma) = \int_X \log f(x, \gamma)\, dP_j(x), \tag{1.25}$$

which according to (RA 2) exists with $-\infty$ as a possible value, is a continuous function of $\gamma \in \Xi$ for all $j = 1, \ldots, q$.

(**RA 4**) *If c is a real number, then for $j = 1, \ldots, q$ there exists a compact set $\Gamma_j \subset \Xi$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log L(x_1, \ldots, x_n, \Xi - \Gamma_j) < c \tag{1.26}$$

a.e. P_j^∞, where in accordance with (1.2)

$$L(x_1, \ldots, x_n, U) = \sup \{ \prod_{i=1}^{n} f(x_i, \gamma);\ \gamma \in U \} \ . \tag{1.27}$$

In this setting and the notation

$$P = P_1^\infty \times \ldots \times P_q^\infty \tag{1.28}$$

Theorem 1.1 has the following form, where measurability of (1.12) with respect to $\mathcal{F}_u$ obviously means that $\tilde{\theta}_u = \tilde{\theta}_u(x^{(u)})$, with $x^{(u)}$ described by (1.20).

Corollary 1.2. *Let us assume that the regularity conditions (RA 1) - (RA 3) hold and with the notation (1.21)*

$$\lim_{u\to\infty} n_u = +\infty\,, \quad \lim_{u\to\infty} \frac{n_u^{(j)}}{n_u} = p_j \in (0,1\rangle \ \textit{for}\ j = 1, \ldots, q\ . \tag{1.29}$$

Further we assume, that the set Ξ is either compact or (RA 4) holds, and the set $\Omega_0 \subset \Theta = \Xi^q$ is closed. If for some $\theta^* = (\theta_1^*, \ldots, \theta_q^*) \in \Omega_0$ and the quantity

$$I(\theta^*, p) = \sum_{j=1}^{q} p_j \int \log f(x, \theta_j^*)\, dP_j(x) \tag{1.30}$$

the inequality

$$I(\theta^*, p) > -\infty \tag{1.31}$$

holds, then with the notation

$$I(\Omega_0, p) = \sup\{I(\theta^*, p);\ \theta^* \in \Omega_0\} \tag{1.32}$$

the set

$$H = H(p) = \{\theta^* \in \Omega_0;\ I(\theta^*, p) = I(\Omega_0, p)\} \tag{1.33}$$

is compact and non-empty, and the assertion (II) of Theorem 1.1 remains true.

A similar assertion for the one-sample case $q = 1$ and based on stronger conditions can be found in [12], where convergence to the postulated unique parameter minimizing the Kullback - Leibler information quantity $K(g : f, \theta)$ has been proved under the assumptions, that the parameter set is compact, the true distribution G has a density g with respect to the measure ν from (1.18), $\mathrm{E}(\log g)$ exists and the densities (1.18) are uniformly bounded with a G-integrable function. These assumptions remain in force also when the general setting in [13] is applied to the i.i.d. case. We remark, that in the q-sample case described in Corollary 1.2, the assumption (RA 4) is imposed to remove the condition of compactness of the parameter space. However, as we shall prove in the following theorem, in the case of exponential families of probabilities even this condition may be omitted.

(**RC 1**) *The measurable space* $(X, \mathcal{F}) = (R^m, \mathcal{B}^m)$*, the dominating measure* ν *is not suppported on a flat, the parameter set*

$$\Xi = \left\{ \gamma \in R^m;\ \int e^{\gamma' x}\, d\nu(x) < +\infty \right\} \tag{1.34}$$

is open, and the densities are determined by the formula

$$f(x, \gamma) = \frac{d\overline{P}_\gamma}{d\nu}(x) = e^{\gamma' x - C(\gamma)}\,, \tag{1.35}$$

where prime denotes the transpose of the vector, and

$$C(\gamma) = \log \int \mathrm{e}^{\gamma' x}\, \mathrm{d}\nu(x)\,. \tag{1.36}$$

(**RC 2**) *The true distributions* $P_1, \ldots, P_q$ *of the* q *populations are such that*

$$\int x\, dP_j(x) \in B(\nu)\,, \qquad j = 1, \ldots, q\,, \tag{1.37}$$

where

$$B(\nu) = \{\mathrm{E}\gamma(x);\ \gamma \in \Xi\}\,. \tag{1.38}$$

The condition that the dominating measure ν is not supported on a plane, i.e., that $\nu(R^m - N) > 0$ for every hyperplane $N = \{y \in R^m;\ c'\, x + b = 0\}$, is according to lema 2.1 in [1] equivalent to the fact that the probabilities $\{\overline{P}_\gamma;\ \gamma \in \Xi\}$ are mutually different. In (1.38) we use the notation

$$\mathrm{E}_\gamma(x) = \int x f(x, \gamma)\, \mathrm{d}\nu(x)\,, \tag{1.39}$$

where the integral is taken coordinate-wise. Since the set Ξ is open, all derivatives of the function $\int \mathrm{e}^{\gamma' x}\, \mathrm{d}\nu(x)$ of γ may be computed, according to Theorem 9 Chapter 2 in [7], by differentiating under the integration sign, which together with Lemma 2.2 in [1] means that the mapping

$$\xi(\gamma) = \mathrm{E}_\gamma(x) \tag{1.40}$$

is $1-1$ on Ξ. Hence under validity of (1.37) there exist unique parameters $\theta_1, \ldots, \theta_q$ from Ξ such that

$$\int x\, \mathrm{d}P_j(x) = \mathrm{E}_{\theta_j}(x)\,, \qquad j = 1, \ldots, q\,. \tag{1.41}$$

Similarly as in Theorem 1.1 and Corollary 1.2, also in the following statement we use the notations (1.19) - (1.22) and (1.28).

Theorem 1.2. *Let us assume that the regularity condition (RC 1) holds, and with the notation* $\Theta = \Xi^q$

$$\Omega_0 = \Theta \cap C\,, \tag{1.42}$$

where $C \subset R^{mq}$ *is a closed set.*

(I) Let us denote $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ *and put (cf. (1.38))*

$$A_n = \{(x_1,\ldots,x_n) \in R^{mn}; \overline{x} \in B(\nu)\}\,. \tag{1.43}$$

This set is open and there exist measurable mappings

$$\tilde{\theta}_u : D_u = A_{n_u^{(1)}} \times \ldots \times A_{n_u^{(q)}} \longrightarrow \Omega_0 \tag{1.44}$$

such that (c.f. (1.2))

$$L(x^{(u)},\Omega_0) = L(x^{(u)},\tilde{\theta}_u(x^{(u)})) \tag{1.45}$$

for all $x^{(u)} \in D_u$*.*

(II) Let also both (RC 2) and (1.29) hold. Then with the notation (1.28) and (1.20)

$$P\Big[s_u \in D_u \ \mathit{for\ all}\ u \geq u(s)\Big] = 1\,, \tag{1.46}$$

the set (1.33), where $p = (p_1,\ldots,p_q)$ *are the numbers from (1.29), is compact and non-empty, and if* $\tilde{\theta}_u : X^{n_u} \to \Omega_0$ *are any measurable mappings such that*

$$P\Big[\,s \in S;\ L(x^{(u)},\Omega_0) = L(x^{(u)},\tilde{\theta}_u(x^{(u)}))\ \mathit{for\ all}\ u \geq u(s)\,\Big] = 1\,, \tag{1.47}$$

then the random variables

$$\rho(\tilde{\theta}_u,H) = \inf\{\,\rho(\tilde{\theta}_u(x^{(u)}),\theta^*);\theta^* \in H\,\}\,, \tag{1.48}$$

where $\rho(\tilde{\theta},\theta^*)$ *is the usual Euclidean distance, converge to zero a.e. P.*

Let us denote for γ, γ^* from Ξ

$$K(\gamma,\gamma^*) = \int \log\left(\frac{f(x,\gamma)}{f(x,\gamma^*)}\right) f(x,\gamma)\,\mathrm{d}\nu(x) = (\gamma-\gamma^*)'\mathrm{E}_\gamma(x) - C(\gamma) + C(\gamma^*)\,, \tag{1.49}$$

and for $\theta = (\theta_1,\ldots,\theta_q)$, $\theta^* = (\theta_1^*,\ldots,\theta_q^*)$ belonging to $\Theta = \Xi^q$ and a vector $p = (p_1,\ldots,p_q)'$ with positive coordinates

$$K(\theta,\theta^*,p) = \sum_{j=1}^{q} p_j K(\theta_j,\theta_j^*)\,. \tag{1.50}$$

If the parameter θ determined with (1.41) belongs to Ω_0, then making use of (1.35) and (1.49) we get that

$$\int \log f(x,\theta_j)\,\mathrm{d}P_j(x) - \int \log f(x,\theta_j^*)\,\mathrm{d}P_j(x) = K(\theta_j,\theta_j^*)\,.$$

Thus with the notation

$$K(\theta,\Omega_0,p) = \inf\{K(\theta,\theta^*,p);\ \theta^* \in \Omega_0\} \tag{1.51}$$

the set H from Theorem 1.2(II) can be written as

$$H = \{\theta^* \in \Omega_0;\ K(\theta,\theta^*,p) = K(\theta,\Omega_0,p)\,\} = \{\theta\}\,,$$

because for $\theta \neq \theta^*$ according to the remark following *(RC 2)* the probabilities $\overline{P}_{\theta_j}$, $\overline{P}_{\theta_j^*}$ are different for some j, and therefore $K(\theta,\theta^*,p) > 0 = K(\theta,\Omega_0,p)$. Hence from Theorem 1.2(II) we obtain the following assertion.

Corollary 1.3. *Let us assume that the assumptions of Theorem 1.2 hold, and the parameter $\theta = (\theta_1, \ldots, \theta_q)$ determined with (1.41) belongs to Ω_0. If $\tilde{\theta}_u : X^{n_u} \to \Omega_0$ are measurable mappings satisfying (1.47), then*

$$\tilde{\theta}_u \to \theta$$

a.e. P for $u \to \infty$.

The previous theorem and its corollary are an extension of Theorem 3.1 and Lemma 3.2 in [1], where in the one-sample case $q = 1$ the true distribution P_1 is supposed to fulfill (1.41) for some $\theta_1 \in \Omega_0$, and existence of a measurable restricted MLE is guaranteed if the maximizers of $\log f(x_1, \ldots, x_n, \gamma)$ on Ω_0 are unique.

Let $k > 1$ be an integer and $a = k(k+1)/2$. Let us put $m = k + a$ and denote

$$\Xi = \{\, \gamma = (\mu', \sigma')' \in R^m;\ \mu \in R^k,\ \ \sigma \in R^a \ \textit{ and } V(\sigma) \textit{ is positive definite} \} \quad (1.52)$$

the set of parameters of the non-singular k-dimensional normal distributions, i.e., μ is the vector of means, $\sigma = (v_{11}, \ldots, v_{1k}, v_{22}, \ldots, v_{2k}, \ldots, v_{kk})'$ are elements of the covariance matrix and $V(\sigma)$ is the symmetric matrix with $V(\sigma)_{ij} = v_{ij}$ for $i \le j$. For $\gamma = (\mu', \sigma')' \in \Xi$ let $f(x, \gamma)$ be density of the normal distribution $N(\mu, V(\sigma))$. In this setting from Theorem 1.2 we obtain the following assertion.

Corollary 1.4. *(I) Let with the notation $\Theta = \Xi^q$ and (1.52)*

$$\Omega_0 = \Theta \cap C\,, \quad (1.53)$$

where $C \subset R^{mq}$ is a closed set. If we put

$$A_n = \{(x_1, \ldots, x_n) \in R^{kn};\ \det \hat{\Sigma} > 0\}\,, \quad (1.54)$$

where $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})(x_j - \overline{x})'$, $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, then there exist measurable mappings (1.44) such that (1.45) holds. Moreover, if the true distributions of the q populations are such that the covariance matrices

$$V_j = cov\,(x|P_j) \quad (1.55)$$

are positive definite for $j = 1, \ldots, q$, then the assertion (II) of Theorem 1.2 is true.

Since for any normal distribution with parameter from (1.52) probability of the set (1.54) according to [10], p. 73, equals 1 if $n > k$, from the previous Corollary we get as an immediate consequence the following existential assertion concerning the Behrens-Fisher problem.

Corollary 1.5. *Let*

$$\Omega_0 = \{\, (\theta_1^*, \ldots, \theta_q^*) \in \Xi^q;\ E_{\theta_1^*}(x) = \ldots = E_{\theta_q^*}(x) \,\} \quad (1.56)$$

denote the hypothesis of equality of means of q normal populations without restriction on their positive definite covariance matrices. If the true distributions

$$P_j = N(\mu_j, \Sigma_j)$$

where covariance matrices of these normal distributions are regular, then a measurable MLE $\tilde{\theta}_u$ of the parameter from (1.56) exists on the set where the sample covariance matrices are positive definite, which has probability 1 provided that $\min_j n_u^{(j)} > k$, *and under validity of (1.29) the random variables (1.48) converge to zero* P a.e., *where the non-empty compact set*

$$H = \{\theta^* \in \Omega_0;\ K(\theta, \theta^*, p) = K(\theta, \Omega_0, p)\}\,, \tag{1.57}$$

with the notation (1.50), (1.51) the symbol $K(\theta_j, \theta_j^*) = K(\overline{P}_{\theta_j}, \overline{P}_{\theta_j^*})$ *denotes the Kullback-Leibler information quantity,* $\theta = (\theta_1, \ldots, \theta_q)$ *and* θ_j *corresponds to* $N(\mu_j, \Sigma_j)$.

We remark, that if $X = \{0, 1, 2, \ldots\}$, $\mathcal{F} = 2^X$ is the system of all subsets of X, $f(x, \gamma) = \mathrm{e}^{-\gamma}\gamma^x/x!$ denotes density of the Poissson distribution and $\Xi = (0, +\infty)$, then the assertions (I) and (II) of Theorem 1.2 remain true, if we put $B(\nu) = (0, +\infty)$ and $\int x\,\mathrm{d}P_j(x) = \sum_{x=0}^{\infty} x P_j(\{x\})$. The assertion (I) can be easily improved in the sense that (1.44) can be written in the form $\tilde{\theta}_u : X^{n_u^{(1)}} \times \ldots \times X^{n_u^{(q)}} \longrightarrow \overline{\Omega}_0$, where $\overline{\Omega}_0$ denotes the closure of Ω_0, (1.45) remains unchanged and if computation of (1.45) involves zero value of the parameter, by density for $\lambda = 0$ we understand $f(x, 0) = \delta_{x,0}$, where $\delta_{x,0}$ is the Kronecker delta.

2. PROOFS

The following assertion is an extension of Lemma 3.3, p. 307 in [9] in the sense that the compact set Γ need not be a subset of R^k. We remark that the presented proof seems to be simpler also in the case when $\Gamma \subset R^k$.

Lemma 2.1. *Let us assume that* $\mathcal{S}$ *is a* σ*-algebra of subsets of* S, Γ *is a compact metric space and* $\mathcal{B}$ *are borel subsets of* Γ. *If* $g : S \times \Gamma \to R$ *is such that the function*

(a) $g(s, .)$ *is continuous for each* $s \in S$,

(b) $g(., \gamma)$ *is measurable for each* $\gamma \in \Gamma$,

then there exists measurable mapping $T : S \to \Gamma$ *such that with the notation*

$$g(s, A) = \sup\{\, g(s, \gamma);\ \gamma \in A\,\} \tag{2.1}$$

the equality

$$g(\,s, T(s)\,) = g(s, \Gamma) \tag{2.2}$$

holds for each $s \in S$.

Proof. Since Γ is a compact metric space, there exist $\{\gamma_j\}_{j=1}^{\infty}$ from Γ and a non-decreasing sequence $\{m_n\}_{n=1}^{\infty}$ of positive integers such that for each n

$$S_n = \{\gamma_1, \ldots, \gamma_{m_n}\}$$

is a 2^{-n} net in Γ, i.e.

$$\Gamma = \bigcup_{j=1}^{m_n} U(\gamma_j, 2^{-n})\,,$$

where $U(\gamma_j,\delta)=\{\,\gamma^*\in\Gamma;\ \rho(\gamma_j,\gamma^*)<\delta\,\}$. We shall utilize the fact, that the sets

$$W_i^{(1)}=\{\,s\in S;\ g(s,\Gamma)=g\Big(s,U(\gamma_i,\frac{1}{2})\Big)\,\}$$

are measurable, and

$$V_i^{(1)}=W_i^{(1)}-\bigcup_{j=1}^{i-1}W_j^{(1)}\,,\ \ i=1,\dots,m_1$$

is a measurable partition of S. Thus the mapping

$$T_1(s)=\sum_{i=1}^{m_1}\gamma_i\chi_{V_i^{(1)}}(s)$$

for which $T_1(s)=\gamma_i$ if $s\in V_i^{(1)}$, is a measurable mapping from S into Γ.

Let us assume, that for $n=1,\dots,k$ we have already constructed measurable mapings $T_n:S\to\Gamma$ such that for all n

$$T_n(s)=\sum_{i=1}^{m_n}\gamma_i\chi_{V_i^{(n)}}(s)\,,\quad \rho(\,T_n(s),T_{n+1}(s)\,)<\frac{1}{2^{n-1}} \tag{2.3}$$

and for all $s\in V_i^{(n)}$

$$g(s,\Gamma)=g\Big(s,U(\gamma_i,\frac{1}{2^n})\Big)\,. \tag{2.4}$$

Denoting

$$W_j^{(k+1,i)}\ =\ \{\,s\in V_i^{(k)};\ g(s,\Gamma)=g\Big(s,U(\gamma_i,\frac{1}{2^k})\cap U(\gamma_j,\frac{1}{2^{k+1}})\Big)\,\}$$

$$V_j^{(k+1,i)}=W_j^{(k+1,i)}-\bigcup_{r=1}^{j-1}W_r^{(k+1,i)}\,,\quad j=1,\dots,m_{k+1}\,,$$

taking into account the fact that $\{V_j^{(k+1,i)};\ i=1,\dots,m_k,\ j=1,\dots m_{k+1}\}$ is a measurable partition of S, and putting

$$T_{k+1}(s)=\sum_{j=1}^{m_{k+1}}\Big[\sum_{i=1}^{m_k}\gamma_j\chi_{V_j^{(k+1,i)}}(s)\Big]=\sum_{j=1}^{m_{k+1}}\gamma_j\chi_{V_j^{(k+1)}}(s)\,, \tag{2.5}$$

where

$$V_j^{(k+1)}=\bigcup_{i=1}^{m_k}V_j^{(k+1,i)}\,,$$

we see that

$$\rho(T_k(s),T_{k+1}(s))\le\frac{1}{2^k}+\frac{1}{2^{k+1}}<\frac{1}{2^{k-1}}\,,$$

and for $s\in V_j^{(k+1)}$

$$g(s,\Gamma)=g\Big(s,U(\gamma_j,\frac{1}{2^{k+1}})\Big)\,.$$

Hence existence of measurable mappings $\{T_n\}_{n=1}^{\infty}$ satisfying (2.3) and (2.4) is proved, and their limit

$$T(s)=\lim_{n\to\infty}T_n(s)$$

is obviously measurable. Since according to (2.4)

$$|g(s,\Gamma) - g(s,T_n(s))| \le \sup\{|g(s,\gamma) - g(s,\gamma^*)|;\ \rho(\gamma,\gamma^*) < \frac{1}{2^n}\},$$

also the equality (2.2) holds. □

Proof of Theorem 1.1. (a) Let us first suppose that Θ is a compact set. Since the function $I(\theta^*)$ in (1.3) is continuous, the assertion (I) is obviously true. Compactness of Ω_0 together with (A 1) according to Lemma 2.1 mean that there exist measurable mapings (1.12) satisfying $L_u(s,\Omega_0) = L_u(s,\tilde\theta_u(s))$ for all $s \in S$.

Let (1.12) be any measurable mapings satisfying (1.13) and a fixed number $\varepsilon > 0$ be such that the set

$$U_\varepsilon = \{\theta^* \in \Omega_0;\ \rho(\theta^*, H) \ge \varepsilon\} \tag{2.6}$$

is non-empty. It is easy to see that continuity of $I(\theta^*)$ together with (1.11) imply existence of a real number $M < I(\Omega_0)$ such that

$$U_\varepsilon \subset \Omega_0(M) = \{\theta^* \in \Omega_0;\ I(\theta^*) \le M\}. \tag{2.7}$$

Now we shall proceed similarly as in the proof of Theorem 1 in [11]. Since the set $\Omega_0(M)$ is compact, validity of (A 3) implies existence of finitely many open sets $V_i = V(\theta_i^*, \Delta_i)$, $i = 1,\dots,r$ such that

$$\Omega_0(M) \subset \bigcup_{i=1}^{r} V_i \subset \Theta - H\ ,$$

$$\max\{I(\theta_i^*,\Delta_i);\ i = 1,\dots,r\} < I(\Omega_0)\ .$$

Hence if $\eta \in H$, then putting

$$\log\frac{0}{0} = 0 \tag{2.8}$$

and making use of (A 3) we see that

$$\limsup_{u\to\infty} \frac{1}{n_u}\log\frac{L_u(s,\Omega_0(M))}{L_u(s,\Omega_0)} \le \max_{i=1,\dots,r}\ \limsup_{u\to\infty}\frac{1}{n_u}\log\frac{L_u(s,V_i)}{L_u(s,\eta)} < 0$$

P a.e., and (II) follows from (2.6) and (2.7).

(b) Let us drop the assumption of compactness of Θ. Since (1.10) holds, there exist a real number c such that

$$c < I(\Omega_0). \tag{2.9}$$

Let $\Gamma \subset \Theta$ be the compact set satisfying (1.8). If $\tilde\theta \in \Theta - \Gamma$, then $I(\tilde\theta) < c$, because in the opposite case according to (A 2)

$$\liminf_{u\to\infty}\frac{1}{n_u}\log L_u(s,\Theta-\Gamma) \ge \liminf_{u\to\infty}\frac{1}{n_u}\log f_u(s,\tilde\theta) \ge c$$

P a.e., which is a contradiction with (1.8). Thus $I(\Omega_0) = I(\Omega_0\cap\Gamma)$ and

$$\{\theta^* \in \Omega_0;\ I(\theta^*) = I(\Omega_0)\} = \{\theta^* \in \Omega_0\cap\Gamma;\ I(\theta^*) = I(\Omega_0\cap\Gamma)\}. \tag{2.10}$$

Further, if η is a fixed point from H, then with the notation (2.8)

$$\limsup_{u\to\infty} \frac{1}{n_u} \log \frac{L_u(s,\Omega_0-\Gamma)}{L_u(s,\Omega_0)} \le \limsup_{u\to\infty} \frac{1}{n_u} \log \frac{L_u(s,\Theta-\Gamma)}{L_u(s,\eta)} < 0$$

P a.e., because $c < I(\eta)$. Hence P a.e.

$$L_u(s,\Omega_0-\Gamma) < L_u(s,\Omega_0)$$

for all $u \ge u(s)$. Since the set Γ is compact, putting $\Theta = \Omega_0 \cap \Gamma$ and taking into account (2.10) and the part (a) of this proof we easily obtain that Theorem 1.1 is true. □

Proof of Corollary 1.2. From (1.24) we obtain that the integral

$$I_j(\gamma,\Delta^*) = \int \log L(x, V(\gamma,\Delta^*))\, \mathrm{d}P_j(x) \tag{2.11}$$

exists with $-\infty$ as a possible value. Employing the monotone convergence theorem we get that with the notation (1.25)

$$\lim_{\Delta\to 0^+} I_j(\gamma,\Delta) = I_j(\gamma)\,. \tag{2.12}$$

Hence putting

$$g_u(s,\theta^*,\Delta) = \frac{1}{n_u} \log \left[\prod_{j=1}^{q} \prod_{i=1}^{n_u^{(j)}} L(x_i^{(j)}, V(\theta_j^*,\Delta)) \right]\,, \tag{2.13}$$

$$I(\theta^*,\Delta) = \sum_{j=1}^{q} p_j I(\theta_j^*,\Delta) \tag{2.14}$$

and utilizing the law of large numbers we obtain that with the notation (1.30) and $I(\theta^*) = I(\theta^*, p)$ the conditions (A 1) - (A 3) hold. Thus according to Theorem 1.1 it is sufficient to prove (A 4).

First we show that for $j = 1,\dots,q$

$$\limsup_{n\to\infty} \frac{1}{n} \log L(x_1,\dots,x_n,\Xi) < d \tag{2.15}$$

P_j^∞ a.e. for some real d. If Γ is the compact set from (RA 4) satisfying (1.26) with $c = 0$, then according to the previous part of the proof there exist finitely many sets $V(\gamma_k,\Delta_k)$, $k = 1,\dots,r$ such that with the notation (2.11)

$$\Gamma \subset \bigcup_{k=1}^{r} V(\gamma_k,\Delta_k)\,, \qquad \max_j \max_k I_j(\gamma_k,\Delta_k) < +\infty\,.$$

Thus

$$\limsup_{n\to\infty} \frac{1}{n} \log L(x_1,\dots,x_n,\Xi) \le$$

$$\leq \limsup_{n\to\infty} \max\Big\{ \frac{1}{n} \log L(x_1,\ldots,x_n,\Xi-\Gamma)\,,\ \frac{1}{n}\log\Big(\prod_{i=1}^{n} L(x_i, V(\gamma_k,\Delta_k))\Big);\ k=1,\ldots,r\Big\}.$$

Choosing real numbers $d_k > \max_j I_j(\gamma_k,\Delta_k)$, putting $d = \max\{0, d_1,\ldots,d_k\}$, utilizing (1.26) and the law of large numbers we get (2.15).

Now we utilize (RA 4) according to which there exists a compact set $\Gamma_j \subset \Xi$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\log L(x_1,\ldots,x_n,\Xi-\Gamma_j) < c_j \tag{2.16}$$

P_j^∞ a.e. . Obviously, $\Gamma = \Gamma_1 \times \ldots \times \Gamma_q$ is a compact subset of Θ and

$$\frac{1}{n_u}\log L_u(s,\Theta-\Gamma) \leq \max_{j=1,\ldots,q} \frac{1}{n_u}\log L_u(s,D_j)\,, \tag{2.17}$$

where $D_j = \Xi\times\ldots\times\Xi\times(\Xi-\Gamma_j)\times\Xi\times\ldots\times\Xi$. But if we denote $\sum^*$ the sum over the indices $1\leq j\leq q$, $j\neq i$, then in accordance with (2.15), (1.20) and (1.22)

$$\limsup_{u\to\infty}\frac{1}{n_u}\log L_u(s,D_i) \leq$$

$$\leq \ \sum{}^{*}\ \limsup_{u\to\infty}\frac{1}{n_u}\log L(y(j,n_u^{(j)}),\Xi)\ +\ \limsup_{u\to\infty}\ \frac{1}{n_u}\log L(y(i,n_u^{(i)}),\Xi-\Gamma_i) <$$

$$< \ \sum{}^{*} p_j d + p_i c_i \tag{2.18}$$

P almost everywhere. Hence if c is a fixed real number and the c_j's in (2.16) are such that

$$\max_i \Big((1-p_i)d + p_i c_i\Big) < c\,,$$

then combining (2.17) and (2.18) we get (1.8). □

The proof of Theorem 1.2 is based on the following lemmas. In these we use for γ, γ^* from (1.34) with the notation (1.35) the quantity (1.49).

Let γ be an interior point of (1.34). As pointed out in [1], p.195, since according to Theorem 9, Chapter 2 in [7] differentiating in (1.36) may be performed under the integration sign,

$$\frac{\partial}{\partial\gamma}C(\gamma) = \mathrm{E}_\gamma(x)\,,\qquad \frac{\partial}{\partial\gamma}\mathrm{E}_\gamma(x) = \mathrm{Var}(x\,|\,\overline{P}_\gamma)\,, \tag{2.19}$$

where Var denotes the covariance matrix.

Lemma 2.2. *Let the condition (RC 1) hold. If Γ is a non-empty compact subset of Ξ and c is a positive real number, then*

$$\mathcal{K} = \{\,\gamma^*\in\Xi;\ \textit{there exists}\ \gamma\in\Gamma\ \textit{such that}\ K(\gamma,\gamma^*)\leq c\,\} \tag{2.20}$$

is a compact subset of R^m.

Proof. (a) First we show that this set is closed. Let us assume that $\{\gamma_n^*\}_{n=1}^\infty$ belong to $\mathcal{K}$ and

$$\lim_{n\to\infty} \gamma_n^* = \gamma^* \in R^m . \tag{2.21}$$

According to (2.20) there exist parameters $\gamma_n \in \Gamma$ such that

$$K(\gamma_n, \gamma_n^*) \le c . \tag{2.22}$$

Since the set Γ is compact,

$$\lim_{k\to\infty} \gamma_{n_k} = \gamma \in \Gamma \tag{2.23}$$

for some subsequence $\{n_k\}_{k=1}^\infty$ of $1, 2, \dots$. But $C(\gamma)$, $\mathrm{E}_\gamma(x)$ are according to (2.19) continuous, and from (2.22) and (1.49) we get that

$$\liminf_{k\to\infty} C(\gamma_{n_k}^*) < +\infty .$$

This together with (1.36), (2.21) and the Fatou lemma means that $\gamma^* \in \Xi$. From (2.23), (2.21), (2.22) and continuity of (1.49) we obtain that γ^* belongs to (2.20).

(b) It remains to prove boundedness of (2.20). Since $\Gamma \subset \Xi \subset R^m$, Γ is compact and Ξ is open, according to the Lebesgue covering lemma (cf. [6], p.154) there exists a positive number δ such that

$$\gamma \in \Gamma , \qquad \|\gamma - \tilde\gamma\| \le \delta \Longrightarrow \tilde\gamma \in \Xi . \tag{2.24}$$

Suppose that the set (2.20) is not bounded. Then there exist $\{\gamma_n^*\}_{n=1}^\infty$ from $\mathcal{K}$ and $\{\gamma_n\}_{n=1}^\infty$ from Γ such that

$$K(\gamma_n, \gamma_n^*) \le c \textit{ for all } n , \ \lim_{n\to\infty} \|\gamma_n^*\| = +\infty , \ \lim_{n\to\infty} \gamma_n = \gamma \in \Gamma . \tag{2.25}$$

Thus the sequence $\{\gamma_n\}_{n=1}^\infty$ is bounded, and we may assume that $\gamma_n \ne \gamma_n^*$ for all n and for the vectors

$$h_n = \delta \frac{\gamma_n^* - \gamma_n}{\|\gamma_n^* - \gamma_n\|} \tag{2.26}$$

there exists the limit

$$\lim_{n\to\infty} h_n = h . \tag{2.27}$$

From (2.25) and (2.26) we get

$$\gamma_n^* = \gamma_n + \alpha_n h_n , \qquad \alpha_n = \frac{\|\gamma_n^* - \gamma_n\|}{\delta} \to +\infty . \tag{2.28}$$

But if α is a real number and $\tilde\gamma$, $\tilde\gamma + \alpha\tilde h$ belong to Ξ, then making use of (1.49) and (2.19) we get

$$\frac{\partial K(\tilde\gamma, \tilde\gamma + \alpha\tilde h)}{\partial\alpha} = -\tilde h' \mathrm{E}_{\tilde\gamma}(x) + \mathrm{E}_{\tilde\gamma+\alpha\tilde h}(x)'\tilde h , \quad \frac{\partial^2 K(\tilde\gamma, \tilde\gamma + \alpha\tilde h)}{\partial\alpha^2} = \tilde h' \mathrm{Var}(x \mid \overline{P}_{\tilde\gamma+\alpha\tilde h})\tilde h , \tag{2.29}$$

where the second derivative is positive for $\tilde h \ne 0$, because ν is suposed not to be concentrated on a flat. Since according to Lemma 7, Chapter II in [7] the set Ξ is convex, (2.28) and (2.29) yield

$$
\begin{aligned}
K(\gamma_n, \gamma_n^*) &= K(\gamma_n, \gamma_n + h_n) + \int_1^{\alpha_n} \frac{\partial K(\gamma_n, \gamma_n + \alpha h_n)}{\partial \alpha} \, \mathrm{d}\alpha \geq \\
&\geq \left. \frac{\partial K(\gamma_n, \gamma_n + \alpha h_n)}{\partial \alpha} \right|_{\alpha=1} (\alpha_n - 1) \,, \qquad (2.30)
\end{aligned}
$$

because the Kullback-Leibler information quantity is non-negative. Since (2.24) - (2.27) hold, $K(\gamma, \gamma + 0h) = 0 < K(\gamma, \gamma + h)$, which together with (2.29) means that

$$\left. \frac{\partial K(\gamma, \gamma + \alpha h)}{\partial \alpha} \right|_{\alpha=1} > 0 \,.$$

From continuity of this partial derivative, (2.30) and (2.28) we therefore obtain, that

$$\lim_{n \to \infty} K(\gamma_n, \gamma_n^*) = +\infty \,,$$

which is a contradiction with (2.25). $\square$

In the following assertion we use the notation

$$I = \{ (p_1, \ldots, p_q); \sum_{j=1}^{q} p_j = 1 \text{ and } \min_j p_j > 0 \} \,. \qquad (2.31)$$

Lemma 2.3. *Let the condition (RC 1) hold, and the null hypothesis*

$$\Omega_0 = \Theta \cap C \,, \qquad (2.32)$$

where $\Theta = \Xi^q$ and C is a closed subset of R^{mq}.

(I) If $\theta \in \Theta$ and $p \in I$, then there exists an $\eta \in \Omega_0$ such that (cf. (1.49)) - (1.51))

$$K(\theta, \Omega_0, p) = K(\theta, \eta, p) \,.$$

(II) If $W \subset \Theta$ and $T \subset I$ are non-empty compact sets, then

$$D = \{\theta^* \in \Omega_0; \text{ there exist } \theta \in W, \ p \in T \text{ such that } K(\theta, \theta^*, p) = K(\theta, \Omega_0, p)\} \qquad (2.33)$$

is a compact subset of R^{mq}.

(III) The function $K(., \Omega, .)$ is continuous on $\Theta \times I$ for every non-empty $\Omega \subset \Theta$.

Proof. (I) If $\eta^* \in \Omega_0$ and $K(\theta, \tilde{\theta}, p) \leq K(\theta, \eta^*, p)$, then for $j = 1, \ldots, q$ with the notation $d(p) = \min\{p_1, \ldots, p_q\}$

$$d(p) K(\theta_j, \tilde{\theta}_j) \ \leq \ K(\theta, \eta^*, p) \ \leq \ \sum_{j=1}^{q} K(\theta_j, \eta_j^*) \,, \qquad (2.34)$$

where θ_j denotes the j-th component of $\theta = (\theta_1, \ldots, \theta_q)$. Hence for $c > 0$ sufficiently large and

$$\Gamma = \bigcup_{j=1}^{q} \{\gamma^* \in \Xi; K(\theta_j, \gamma^*) \leq c\} \,, \qquad \Omega_1 = \Omega_0 \cap (\Gamma \times \ldots \times \Gamma) \qquad (2.35)$$

the equality

$$K(\theta, \Omega_0, p) = K(\theta, \Omega_1, p) \qquad (2.36)$$

holds. Taking into account (2.32) and Lemma 2.2 we get that the set Ω_1 is compact, and therefore $K(\theta, ., p)$ attains on Ω_1 its minimum.

(II) Owing to compactnes of T the number $\Delta = \inf\{p_j;\, j = 1, \dots, q\,,\; p \in T\}$ is positive, and similarly as in (2.34) - (2.36) one can show by means of Lemma 2.2 that

$$D \subset (\Gamma \times \ldots \times \Gamma) \subset \Theta\,, \tag{2.37}$$

where Γ is a compact set. Thus it is sufficient to prove that the set D is closed.

If $\{\theta_n^*\}_{n=1}^{\infty}$ belong to D and $\lim_{n\to\infty} \theta_n^* = \theta^*$, then (2.37), (2.32) imply that $\theta^* \in \Omega_0$. According to (2.33)

$$K(\theta_n, \theta_n^*, p_n) = K(\theta_n, \Omega_0, p_n) \tag{2.38}$$

for some $\theta_n \in W$ and $p_n \in T$. Since the compactness assumptions allow us to assume that $\theta_n \to \theta \in W$, $p_n \to p \in T$, making use of the assertion (I) of this lemma and (2.38) we get

$$K(\theta, \Omega_0, p) = K(\theta, \eta, p) = \lim_{n\to\infty} K(\theta_n, \eta, p_n) \geq \limsup_{n\to\infty} K(\theta_n, \Omega_0, p_n) = K(\theta, \theta^*, p)$$

and $\theta^* \in D$.

(III) Since the function $K(\theta, ., p)$ is continuous on Θ, the set Ω may be replaced with $\overline{\Omega} \cap \Theta$, and we shall therefore assume that $\Omega = \Omega_0$ is the set (2.32).

Let $\theta_n \to \theta$ belong to Θ and $p_n \to p$ belong to I. If

$$W = \{\theta_n;\, n = 1, 2, \dots\} \cup \{\theta\}\,, \qquad T = \{p_n;\, n = 1, 2, \dots\} \cup \{p\}\,,$$

then according to (I) of this lemma there exist η, η_n from the set D defined by means of (2.33), satisfying the equalities

$$K(\theta, \Omega_0, p) = K(\theta, \eta, p)\,, \quad K(\theta_n, \Omega_0, p_n) = K(\theta_n, \eta_n, p_n)\,.$$

Since the set D is according to (II) compact, there exists a sequence $\{n_k\}_{k=1}^{\infty}$ such that $\eta_{n_k} \to \eta^* \in D$ and $\liminf_{n\to\infty} K(\theta_n, \Omega_0, p_n) = \lim_{k\to\infty} K(\theta_{n_k}, \Omega_0, p_{n_k})$. Thus

$$K(\theta, \Omega_0, p) = \lim_{n\to\infty} K(\theta_n, \eta, p_n) \geq \limsup_{n\to\infty} K(\theta_n, \Omega_0, p_n) \geq$$

$$\geq \liminf_{n\to\infty} K(\theta_n, \Omega_0, p_n) = \lim_{k\to\infty} K(\theta_{n_k}, \eta_{n_k}, p_{n_k}) = K(\theta, \eta^*, p) \geq K(\theta, \Omega_0, p)\,,$$

and the continuity is proved. $\square$

Proof of Theorem 1.2. (I) According to Lemma 2.2 in [1] the mapping (1.40) is 1 - 1 on Ξ. Since differentiating of (1.36) can be performed under integration sign, (2.19) holds and ν is not concentrated on a flat, Jacobian of $\xi(\gamma)$ is positive on Ξ and ξ has continuous derivatives. This according to Theorem 212 in [5] means, that the set $B(\nu)$ is open and ξ^{-1} has continuous derivatives on $B(\nu)$. Thus A_n is open and ξ^{-1} is measurable. Let for $(x_1, \dots, x_n) \in A_n$

$$\hat{\theta}_n(x_1, \dots, x_n) = \xi^{-1}(\overline{x})\,, \tag{2.39}$$

and for $x^{(u)} \in D_u$ with the notation (1.20)

$$\hat{\theta}_{(u)}(x^{(u)}) = \left(\hat{\theta}_{n_u^{(1)}}(y(1, n_u^{(1)})), \ldots, \hat{\theta}_{n_u^{(q)}}(y(q, n_u^{(q)}))\right) . \tag{2.40}$$

Since for $(x_1, \ldots, x_n) \in A_n$ with the notation (1.49)

$$\log L(x_1, \ldots, x_n, \gamma) = g_n(x_1, \ldots, x_n) - nK(\hat{\theta}_n, \gamma) \tag{2.41}$$

where

$$g_n(x_1, \ldots, x_n) = n\overline{x}'\hat{\theta}_n - nC(\hat{\theta}_n) , \tag{2.42}$$

we see that on D_u for $\theta^* \in \Theta$

$$\log L(x^{(u)}, \theta^*) = G^{(u)}(x^{(u)}) - n_u K(\hat{\theta}_{(u)}, \theta^*, p_u) , \tag{2.43}$$

where $p_u = \left(\frac{n_u^{(1)}}{n_u}, \ldots, \frac{n_u^{(q)}}{n_u}\right)$ and

$$G^{(u)}(x^{(u)}) = \sum_{j=1}^{q} g_{n_u^{(j)}}(y(j, n_u^{(j)})) . \tag{2.44}$$

Let us denote by $\mathcal{H}(D_u)$ the boundary of D_u, and put

$$D_u^{(M)} = \{\, x^{(u)} \in D_u;\ \|x^{(u)}\| \leq M\ ,\ \ \rho(x^{(u)}, \mathcal{H}(D_u)) \geq \frac{1}{M}\ \} ,$$

where $\rho(x, A) = \inf\{\|x - y\|;\ y \in A\}$ and $\rho(x, \emptyset) = +\infty$. Then $\{D_u^{(M)}\}_{M=1}^{\infty}$ is an increasing sequence of compact sets and

$$D_u = \bigcup_{M=1}^{\infty} D_u^{(M)}$$

because D_u is open. Since continuous image of a compact set is again compact,

$$W_u^{(M)} = \{\, \hat{\theta}_{(u)}(x^{(u)});\ x^{(u)} \in D_u^{(M)}\, \}$$

is a compact set. Hence

$$\mathcal{D}_u^{(M)} = \{\, \tilde{\theta} \in \Omega_0;\ \textit{there exists } \hat{\theta} \in W_u^{(M)} \textit{ such that } K(\hat{\theta}, \tilde{\theta}, p_u) = K(\hat{\theta}, \Omega_0, p_u)\, \}$$

is according to Lemma 2.3(II) compact, and from (2.43) and Lemma 2.3(I) we obtain that on $D_u^{(M)}$

$$\log L(x^{(u)}, \Omega_0) = \log L(x^{(u)}, \mathcal{D}_u^{(M)}) .$$

From Lemma 2.1 we therefore get existence of measurable mappings $\tilde{\theta}_u^{(M)} : D_u^{(M)} \to \mathcal{D}_u^{(M)}$ such that

$$\log L(x^{(u)}, \Omega_0) = \log L(x^{(u)}, \tilde{\theta}_u^{(M)}(x^{(u)}))$$

for all $x^{(u)} \in D_u^{(M)}$. Hence if we put

$$\tilde{\theta}_u(x^{(u)}) = \tilde{\theta}_u^{(M)}(x^{(u)})$$

for $x^{(u)} \in D_u^{(M)} - D_u^{(M-1)}$, the assertion (I) is proved.

(II) According to the law of large numbers and (1.41)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} x_j = \mathrm{E}_{\theta_j}(x) \in B(\nu) \tag{2.45}$$

a.e. P_j^∞. Since the set $B(\nu)$ is according to (I) of this proof open, (2.45) implies (1.46).

Now it remains to prove the validity of the assumptions of Corollary 1.2. Since (RA 1) - (RA 3) obviously hold, it is sufficient to prove (RA 4). If the set $\Xi - \Gamma$ is non-empty, then on A_n according to (2.41)

$$\frac{1}{n}\log L(x_1,\ldots,x_n,\Xi-\Gamma) = \frac{1}{n}\, g_n(x_1,\ldots,x_n) - K(\hat\theta_n,\Xi-\Gamma)\,.$$

This together with (2.45), (2.42), (2.39), continuity of ξ^{-1} and Lemma 2.3(III) means, that

$$\lim_{n\to\infty}\frac{1}{n}\log L(x_1,\ldots,x_n,\Xi-\Gamma) = \mathrm{E}_{\theta_j}(x)'\theta_j - C(\theta_j) - K(\theta_j,\Xi-\Gamma)$$

P_j^∞ a.e., and existence of the compact set Γ fulfilling (1.26) can be easily proved by means of Lemma 2.2. $\square$

Proof of Corollary 1.4. If we put for $x \in R^k$

$$T(x) = \Big(x_1,\ldots,x_k,-\frac{x_1^2}{2},-x_1x_2,\ldots,-x_1x_k,-\frac{x_2^2}{2},-x_2x_3,\ldots,-x_2x_k,\ldots,-\frac{x_k^2}{2}\Big)'$$

and analogously for $\gamma = (\mu',\sigma')' \in \Xi$

$$e(\gamma) = \Big((V^{-1}(\sigma)\mu)', V^{-1}(\sigma)_{11}, V^{-1}(\sigma)_{12},\ldots,\ldots,V^{-1}(\sigma)_{kk}\Big)'\,,$$

then

$$e(\gamma)'T(x) = -\frac{1}{2}(x-\mu)'V(\sigma)^{-1}(x-\mu) + \frac{1}{2}\mu' V(\sigma)^{-1}\mu \tag{2.46}$$

and e, e^{-1} are continuous mappings of Ξ onto Ξ. Since the set Ξ is open, the Corollary 1.4 will be proved if we prove the following lemma, by means of which one can easily show that the assumptions of Theorem 1.2 are fulfilled.

Lemma 2.4. *(I) If we denote for $A \in \mathcal{B}^m$*

$$\nu(A) = \mu_L(T^{-1}A)$$

where μ_L is the Lebesgue measure on $(R^k,\mathcal{B}^k)$, then the measure ν is not supported on a flat.

(II) The natural set of parameters (1.34) coincides with (1.52).

Proof. Since with the notation (1.35) according to Lemma 2.2 in [1] the measure ν is not supported on a flat if and only if the mapping $\gamma \to \overline{P}_\gamma$ is 1 - 1, making use of (2.46) we see that it is sufficient to prove (II). However, if $\gamma = (\mu',\sigma')' \in R^m$ and the matrix $V(\sigma)$ is not positive definite, then

$$\gamma' T(x) = x'\mu - \frac{1}{2} x' V(\sigma) x = z' P\mu - \frac{1}{2} \sum_{i=1}^{k} \lambda_i z_i^2 \,,$$

where P is an orthogonal matrix, $z = Px$ and $\lambda_k \leq 0$. Thus after some calculation

$$\int_{R^m} e^{\gamma' y} \, \mathrm{d}\nu(y) = \int_{R^k} e^{\gamma' T(x)} \, \mathrm{d}\mu_L(x) = +\infty$$

and the set of natural parameters (1.34) is a subset of (1.52). Since the reverse is also true, the lemma is proved. □

References

[1] R. H. Berk: Consistency and asymptotic normality of mle's for exponential models. *Ann. of Math. Statist.* 43(1972), 193–204.

[2] P. Billingsley: *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago l961.

[3] L. Fahrmeir: Maximum likelihood estimation in misspecified generalized linear models. *Statistics* 21(1990), 487–502.

[4] R. V. Foutz and R. C. Srivastava: Statistical inference for Markov processes when the model is incorrect. *Adv. in Applied Prob.* 11(1979), 737–749.

[5] V. Jarník: *Diferenciální počet II.* ČSAV, Prague 1956 (in Czech).

[6] J. L. Kelley: *General Topology.* van Nostrand, Princeton 1968.

[7] E. L. Lehmann: *Testing Statistical Hypotheses.* Wiley & Sons, New York 1959.

[8] T. Mäkelainen, K. Schmidt, and G. P. Styan: On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Ann. Statist.* 9(1981), 758–767.

[9] L. Schmetterer: *Introduction to Mathematical Statistics.* Springer Verlag, Berlin 1974.

[10] M. S. Srivastava and C. G. Khatri: *Introduction to Multivariate Statistics.* North Holland, New York 1979.

[11] A. Wald: Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* 20(1949), 595–601.

[12] H. White: Maximum-likelihood estimation of misspecified models. *Econometrica* 50(1982), 1–25.

[13] H. White: Maximum likelihood estimation of misspecified dynamic models. in: *Missspecification analysis* (ed.: T. Dijkstra), Lecture Notes in Econom. and Math. Systems no. 237, Springer Verlag, Berlin - New York 1984, pp. 1–19.

RNDr. František Rublík, CSc., Institute of Measurement of SAV, Dúbravská cesta 9, 84219 Bratislava, Slovak Republic.