Kybernetika

František Rublík

On d-optimality of the LR tests

Kybernetika, Vol. 31 (1995), No. 2, 189--206

Persistent URL: `http://dml.cz/dmlcz/124426`

KYBERNETIKA — VOLUME 31 (1995), NUMBER 2, PAGES 189 – 206

ON d-OPTIMALITY OF THE LR TESTS

FRANTIŠEK RUBLÍK[1]

The lower asymptotic distributional bound of the level attained is attained in the case of the likelihood ratio statistics. The regularity conditions on which the proofs are based are verified for the non-singular normal, the multinomial and the Poisson distribution.

1. INTRODUCTION AND THE MAIN RESULTS

First we introduce notations which will be useful for describing asymptotic properties of tests of hypotheses about q statistical populations. These q populations will be supposed to have their distributions from the same family of probabilities.

Let $\{\overline{P}_\gamma;\ \gamma \in \Xi\}$ be a family of probability measures, defined on $(X, \mathcal{F})$ by means of the densities $\{f(x,\gamma);\ \gamma \in \Xi\}$ with respect to a measure ν. If we denote the q-fold products

$$S = X^\infty \times \ldots \times X^\infty\,, \quad \mathcal{S} = \mathcal{F}^\infty \times \ldots \times \mathcal{F}^\infty\,, \quad \Theta = \Xi^q \tag{1.1}$$

then for $\theta = (\theta_1, \ldots, \theta_q) \in \Theta$ the corresponding product measure

$$P_\theta = \overline{P}^\infty_{\theta_1} \times \ldots \times \overline{P}^\infty_{\theta_q}\,, \tag{1.2}$$

defined on the σ-algebra $\mathcal{S}$, describes independent sampling from the q populations $(X, \mathcal{F}, \overline{P}_{\theta_j})$, $j = 1, \ldots, q$.

Throughout the paper we shall assume that the null and the alternative hypotheses

$$\emptyset \neq \Omega_0 \subset \Omega_1 \subset \Theta \tag{1.3}$$

are tested by means of a test statistic $T_u : S \to R$, whose large values are significant (i.e. the null hypothesis Ω_0 is rejected in favor of the alternative $\Omega_1 - \Omega_0$, whenever T_u exceeds a chosen critical constant). We shall suppose that T_u depends on

$$s = \left(\{x_j^{(1)}\}_{j=1}^\infty, \ldots, \{x_j^{(q)}\}_{j=1}^\infty\right) \in S$$

through

$$x^{(u)} = \left(y(1, n_u^{(1)}), \ldots, y(q, n_u^{(q)})\right) \tag{1.4}$$

[1]This research was supported by the Slovak Academy of Sciences under Grant No. 999366.

only, where

$$y\left(j, n_u^{(j)}\right) = \left(x_1^{(j)}, \dots, x_{n_u^{(j)}}^{(j)}\right) \tag{1.5}$$

is a sample from the jth population. To establish a bound for the asymptotic distribution of the $\log L_u(s)$, where the level attained $L_u(s) = \sup\{ P_\theta[T_u \geq T_u(s)];\ \theta \in \Omega_0 \}$, we impose these conditions.

(C1). In the notation

$$n_u = \sum_{j=1}^{q} n_u^{(j)}, \quad p_u^{(j)} = n_u^{(j)}/n_u \tag{1.6}$$

the relations

$$\lim_{u\to\infty} n_u = +\infty, \quad \lim_{u\to\infty} p_u^{(j)} = p_j \in (0, 1), \quad j = 1, \dots, q \tag{1.7}$$

hold.

(C2). The measurable space $(X, \mathcal{F}) = (R^m, \mathcal{B}^m)$, the dominating measure ν is not suppported on a flat, the parameter set

$$\Xi = \left\{ \gamma \in R^m;\ \int e^{\gamma' x}\, d\nu(x) < +\infty \right\} \tag{1.8}$$

is open, and the densities are determined by the formula

$$f(x, \gamma) = \frac{d\overline{P}_\gamma}{d\nu}(x) = e^{\gamma' x - C(\gamma)}, \tag{1.9}$$

where prime denotes transposition of the vector, and

$$C(\gamma) = \log \int e^{\gamma' x}\, d\nu(x). \tag{1.10}$$

Before proceeding to the further text we remark, that one of the consequences of (C2) according to Lemma 2.1 in [2] is that $\overline{P}_\gamma \neq \overline{P}_{\gamma^*}$ whenever the parameters $\gamma \neq \gamma^*$.

Let us denote

$$\mathcal{P} = \{ p \in R^q;\ \sum_{j=1}^{q} p_j = 1 \text{ and } p_j > 0 \text{ for all } j \} \tag{1.11}$$

and for $\theta, \theta^* \in \Theta$, $p \in \mathcal{P}$ put

$$K(\theta, \theta^*, p) = \sum_{j=1}^{q} p_j K(\theta_j, \theta_j^*), \tag{1.12}$$

$$K(\theta, \Omega_0, p) = \inf\{ K(\theta, \theta^*, p);\ \theta^* \in \Omega_0 \}, \tag{1.13}$$

where $K(\theta_j, \theta_j^*) = \int \log\left(\frac{f(x,\theta_j)}{f(x,\theta_j^*)}\right) f(x, \theta_j)\, d\nu(x)$ is the Kullback–Leibler information number.

Lemma 1.1. If the assumptions (C1), (C2) are fulfilled and $\theta \in \Omega_1 - \overline{\Omega}_0$, where the bar denotes the closure of the set, then for every parameters $\eta_u \in \Omega_0$ such that

$$K(\theta, \Omega_0, p_u) = K(\theta, \eta_u, p_u) + o(1/n_u)\,, \tag{1.14}$$

and for every real number t

$$\limsup_{u\to\infty} P_\theta \left[\frac{\log L_u(s) + n_u K(\theta, \Omega_0, p_u)}{\sqrt{n_u}\sigma_u} < t \right] \le \Phi(t)\,, \tag{1.15}$$

where in the notation (1.6)

$$p_u = (p_u^{(1)}, \ldots, p_u^{(q)})\,, \tag{1.16}$$

$$\sigma_u^2 = \sum_{j=1}^q p_u^{(j)} \sigma^2(\Pi_j(\theta), \Pi_j(\eta_u))\,, \quad \sigma^2(\gamma, \gamma^*) = \operatorname{Var}\left[\log \frac{f(x,\gamma)}{f(x,\gamma^*)} \,|\, \overline{P}_\gamma \right]\,, \tag{1.17}$$

$$\Pi_j((\theta_1, \ldots, \theta_q)) = \theta_j\,, \tag{1.18}$$

and Φ is distribution function of the $N(0,1)$ distribution.

If the set Ω_0 is closed in Θ, i.e. if

$$\Omega_0 = \Theta \cap C \quad \text{where} \quad C \subset R^{mq} \text{ is a closed set,} \tag{1.19}$$

then in accordance with Lemma 1.1 and the terminology accepted in [1] we shall say that the statistics T_u are d-*optimal (distributionally optimal) for testing* Ω_0 *against* $\Omega_1 - \Omega_0$, if for each $\theta \in \Omega_1 - \Omega_0$, every real number t and $\eta_u \in \Omega_0$ satisfying (1.14)

$$\lim_{u\to\infty} P_\theta \left[\frac{\log L_u(s) + n_u K(\theta, \Omega_0, p_u)}{\sqrt{n_u}\sigma_u} < t \right] = \Phi(t) \tag{1.20}$$

whenever (C1) holds. In the one-sample case investigated in [3] obviously $K(\theta, \Omega_0, p_u) = \inf\{\, K(\theta, \theta^*); \theta^* \in \Omega_0\}$. However, if $q > 1$, then $K(\theta, \Omega_0, p_u)$ cannot be in (1.20) replaced with its limiting value $K(\theta, \Omega_0, p)$, because the left-hand side in (1.20) could be zero and the set of d-optimal statistics would be empty for such hypotheses.

In considerations concerning the likelihood ratio test statistic we shall use for $\Omega \subset \Theta$ throughout the paper the notation

$$L(x^{(u)}, \Omega) = \sup \left\{ \prod_{j=1}^q \prod_{i=1}^{n_u^{(j)}} f(x_i^{(j)}, \theta_j);\ \theta = (\theta_1, \ldots, \theta_q) \in \Omega \right\}\,. \tag{1.21}$$

In proving d-optimality of the LRT statistics we shall use also the following condition.

(C3). If (C1) holds, then there exist measurable functions $h_u : (0, +\infty) \longrightarrow R$ such that for every real $t \geq 0$

$$\sup \left\{ P_\theta \left[2 \log \frac{L(x^{(u)}, \Theta)}{L(x^{(u)}, \theta)} \geq t \right] ; \theta \in \Theta \right\} \leq \exp \left[\frac{-t}{2} + h_u(t) \right] \tag{1.22}$$

where

$$\lim_{u \to \infty} \frac{h_u(t_u)}{\sqrt{n_u}} = 0 \tag{1.23}$$

for every sequence $\{t_u\}_{u=1}^{\infty}$ of non-negative numbers satisfying the inequality

$$\limsup_{u \to \infty} \frac{t_u}{n_u} < +\infty \,. \tag{1.24}$$

If both (1.3) and (1.19) hold, then we shall say that the set Ω_0 is Ω_1 *informationally regular* (or briefly, Ω_1 IR), if for each $p \in \mathcal{P}$ and $\theta \in \Omega_1$ there exists a unique point $\eta = \eta(\theta, p) \in \Omega_0$ such that in the notation (1.13) the equality

$$K(\theta, \Omega_0, p) = K(\theta, \eta(\theta, p), p) \tag{1.25}$$

holds.

Theorem 1.1. Let us assume, that the assumptions (C1), (C2) hold, the set Ω_0 from (1.3) is Ω_1 IR,

$$T_u = 2 \log \frac{L(x^{(u)}, \Omega_1)}{L(x^{(u)}, \Omega_0)} \tag{1.26}$$

and $\theta \in \Omega_1 - \Omega_0$.

(I) In the notation (1.13), (1.16)–(1.18) and (1.2)

$$\mathcal{L} \left[\frac{\frac{1}{2} T_u - n_u K(\theta, \Omega_0, p_u)}{\sqrt{n_u} \sigma_u} \,\Big|\, P_\theta \right] \longrightarrow N(0, 1) \tag{1.27}$$

in the sense of the weak convergence of probability measures.

(II) If also (C3) is fulfilled, then (1.20) holds and the statistics (1.26) are d-optimal for testing Ω_0 against $\Omega_1 - \Omega_0$.

This theorem is a q-sample version of Proposition 2.8 in [1], whose assumptions are of the asymptotic nature, and require verification for every particular hypothesis, which mainly in the q-sample case could be complicated. In contrast with this, Theorem 1.1 provides us with an apriori knowledge, that for the given exponential family the statistics (1.26) are d-optimal in the case of the IR hypotheses. As it is well known, and explained also in considerations concerning (2.43) in [11], under validity of (C2)

$$\log L(x^{(u)}, \theta) = G^{(u)}(x^{(u)}) - n_u K(\hat{\theta}, \theta, p_u) \tag{1.28}$$

provided that the unrestricted MLE $\hat{\theta}$ exists. Hence the MLE of the unknown parameter from Ω_0 is the value minimizing $K(\hat{\theta}, \cdot, p_u)$ on Ω_0, and the IR hypotheses may be interpreted as the ones for which the restricted MLE is uniquely determined. This suggests that the assumption of being informationally regular is not very restrictive.

2. PROOFS OF THE ASSERTIONS FROM SECTION 1

Lemma 2.1. Let the conditions (C1), (C2) hold and parameters θ, η_u satisfy the assumptions of Lemma 1.1. There exists a compact set $\Gamma \subset \Theta$ such that

$$\eta_u \in \Gamma \text{ for all } u \tag{2.1}$$

and if we put (cf. (1.21))

$$R_u(s) = \log \frac{L(x^{(u)}, \theta)}{L(x^{(u)}, \eta_u)} \tag{2.2}$$

then in the notation (1.13) and (1.16) - (1.18) for $u \to \infty$

$$\mathcal{L}\left[\frac{R_u(s) - n_u K(\theta, \Omega_0, p_u)}{\sqrt{n_u}\sigma_u} \mid P_\theta\right] \longrightarrow N(0,1)\,. \tag{2.3}$$

Moreover, if the set Ω_0 is Ω_1 IR, then (cf. (1.25))

$$\lim_{u \to \infty} \eta_u = \eta(\theta, p)\,. \tag{2.4}$$

P r o o f. First we shall assume that $\eta_u \to \eta$ for $u \to \infty$. Utilizing the Tchebychev inequality and continuity of $\sigma(\gamma, .)$ we obtain that

$$R_u(s) - n_u K(\theta, \Omega_0, p_u) = o_P(n_u^{1/2}) + \sum_j{}^* Z_{u,j} \tag{2.5}$$

with (cf. (1.18), (1.5))

$$Z_{u,j} = \log \frac{L(y(j, n_u^{(j)}), \Pi_j(\theta))}{L(y(j, n_u^{(j)}), \Pi_j(\eta_u))} - n_u^{(j)} K(\Pi_j(\theta), \Pi_j(\eta_u))\,,$$

where $\sum^*$ denotes the sum over the indices j for which the inequality $\Pi_j(\theta) \neq \Pi_j(\eta)$ holds. Since ν is not supported on a flat, according to Lemma 2.1 in [2] the number $\sigma^2(\gamma, \gamma^*)$ in (1.17) is positive if $\gamma \neq \gamma^*$. Furthermore, the parameter set Ξ of this exponential family is open, which together with $\eta_u \to \eta$ enables us to apply the Lindeberg theorem on $\left[n_u^{(j)} \sigma^2(\Pi_j(\theta), \Pi_j(\eta_u))\right]^{-1/2} Z_{u,j}$. Thus (2.3) follows from (2.5) and from the fact that σ_u^2 tends to the positive real number $\sigma^2 = \sum_j p_j \sigma^2(\Pi_j(\theta), \Pi_j(\eta))$.

Let us not assume that the sequence $\{\eta_u\}$ is convergent. Let $\tilde{\theta} \in \Omega_0$ be a fixed parameter and

$$\Gamma_j = \{\theta^* \in \Xi;\ K(\Pi_j(\theta), \Pi_j(\theta^*) \leq c_j\}\,, \quad c_j = \frac{1}{p_j}\left(\sum_{i=1}^{q} K(\Pi_i(\theta), \Pi_i(\tilde{\theta})) + 1\right)\,,$$

where $p = (p_1, \ldots, p_q)'$ is vector of the limiting values from (1.7). According to Lemma 2.2 in [11] the set Γ_j is compact. Since non-negativity of $K(\cdot,\cdot)$ together with (1.14) and (1.7) imply that $\eta_u \in \tilde{\Gamma} = \Gamma_1 \times \ldots \times \Gamma_q$ for all u sufficiently large, (2.1) is proved. Since $\tilde{\Gamma}$ is compact, from the previous part of the proof we obtain that for every subsequence $\{u_j\}$ of positive integers there exists a subsubsequence $\{u_{j_i}\}$ for which (2.3) holds, and the convergence (2.3) is proved.

Since the set $\tilde{\Gamma}$ is compact, each subsequence of $\{\eta_u\}$ contains a convergent subsubsequence. But if the set Ω_0 is Ω_1 IR, continuity of $K(\cdot, \Omega_0, \cdot)$ proved in Lemma 2.3(III) in [11] implies that this subsubsequence converges to $\eta(\theta, p)$, and (2.4) is proved. □

Proof of Lemma 1.1. The proof can be performed analogously as the proof of Theorem 2.1 in [3], where only the case $q = 1$ is considered and an assumption of existence of a minimizing point is imposed. Let us denote for $\varepsilon > 0$

$$A_u(\varepsilon) = \{s \in S;\ \log L_u(s) + R_u(s) < -\varepsilon\sqrt{n_u}\}\ .$$

Since (2.3) holds, given $\delta > 0$ we can find a number M such that the sets

$$B_u = \{s \in S;\ |R_u(s) - n_u K(\theta, \eta_u, p_u)| \le M\sqrt{n_u}\sigma_u\}$$

satisfy the inequality $1 - P_\theta(B_u) < \delta$. Hence following the lines of the proof of Theorem 2.1 in [3], p. 387, we can prove that

$$\lim_{u\to\infty} P_\theta(\,A_u(\varepsilon)\,) = 0$$

which together with Lemma 2.1 implies (1.15). □

Lemma 2.2. Let the assumptions (C1), (C2) hold. If the set Ω_0 is Ω_1 IR, $\theta \in \Omega_1$ and the parameters $\eta_u \in \Omega_0$ satisfy (1.14), then (cf. (1.21), (1.2))

$$n_u^{-1/2} \log \frac{L(x^{(u)}, \Omega_0)}{L(x^{(u)}, \eta_u)} = o_P(1), \tag{2.6}$$

where $P = P_\theta$.

Proof. Let us denote

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

where $\overline{x} = \frac{1}{n}\sum_{j=1}^n x_j$, and

$$B(\nu) = \{\xi(\gamma);\ \gamma \in \Xi\}\,, \quad \xi(\gamma) = \int x f(x,\gamma)\,\mathrm{d}\nu(x)\,.$$

As it is explained in the proof of Theorem 1.2 in [11] on p. 61, the sets A_n, $B(\nu)$ are open, the mapping ξ posseses an inversion ξ^{-1} and both ξ and ξ^{-1} have continuous derivatives of the first order. Moreover, if we denote for $(x_1, \ldots, x_n) \in A_n$

$$\hat{\theta}_n(x_1, \ldots, x_n) = \xi^{-1}(\overline{x})\,, \tag{2.7}$$

and put $g_n(x_1,\ldots,x_n) = n(\overline{x}_n{}'\hat{\theta}_n - C(\hat{\theta}_n))$, then for each parameter $\gamma \in \Xi$ on A_n

$$\log L(x_1,\ldots,x_n,\gamma) = g_n(x_1,\ldots,x_n) - nK(\hat{\theta}_n,\gamma)\,, \tag{2.8}$$

where $K(\gamma^*,\gamma) = (\gamma^*-\gamma)'\mathrm{E}_{\gamma^*}(x) - C(\gamma^*) + C(\gamma)$ is the Kullback–Leibler information quantity; non-negativity of $K(\cdot,\cdot)$ and (2.8) imply that $\hat{\theta}_n$ is the unrestricted MLE. Since the set $B(\nu)$ is open, making use of the law of large numbers we get that for each $\gamma \in \Xi$

$$\lim_{n\to\infty} P_\gamma(A_n) = 1\,, \quad \hat{\theta}_n \to \gamma \text{ a.c. } \overline{P}_\gamma^\infty\,. \tag{2.9}$$

According to the central limit theorem the random variables $\sqrt{n}(\overline{x}-\xi(\gamma))$ are bounded in the probability $\overline{P}_\gamma^\infty$. Combining this with the Taylor theorem, (2.7), the fact that the set $B(\nu)$ is open and ξ^{-1} has on $B(\nu)$ continuous partial derivatives, we get that

$$\hat{\theta}_n = \gamma + O_P(n^{-1/2})\,, \quad P = \overline{P}_\gamma^\infty\,. \tag{2.10}$$

Further, $\Omega_0 = \Xi^q \cap C$, where $C \subset R^{mq}$ is a closed set. This according to Theorem 1.2 in [11] means that there exist measurable mappings

$$\tilde{\theta}_u : D_u = A_{n_u^{(1)}} \times \ldots \times A_{n_u^{(q)}} \longrightarrow \Omega_0 \tag{2.11}$$

such that on D_u the equality

$$L\left(x^{(u)},\Omega_0\right) = L\left(x^{(u)},\tilde{\theta}_u(x^{(u)})\right) \tag{2.12}$$

holds, and in the notation $H = \{\theta^* \in \Omega_0;\, K(\theta,\theta^*,p) = K(\theta,\Omega_0,p)\}$ the random variables $\rho(\tilde{\theta}_u,H)$ tend to zero in probability P_θ. Since the set Ω_0 is Ω_1 IR, the set H consists of the unique point $\eta(\theta,p)$, and taking into account the first relation in (2.9) we see that

$$\tilde{\theta}_u = \eta(\theta,p) + o_P(1)\,, \quad P = P_\theta\,. \tag{2.13}$$

If we denote for $x^{(u)} \in D_u$

$$\hat{\theta}_{(u)}(x^{(u)}) = \left(\hat{\theta}_{n_u^{(1)}}(y(1,n_u^{(1)})),\ldots,\hat{\theta}_{n_u^{(q)}}(y(q,n_u^{(q)}))\right)\,, \tag{2.14}$$

then taking into account (2.8), (2.12) and the first equality in (2.9) we see that in the notation (1.18)

$$\log\frac{L(x^{(u)},\Omega_0)}{L(x^{(u)},\eta_u)} = \sum_{j=1}^{q} n_u^{(j)}\left[K(\Pi_j(\hat{\theta}_{(u)}),\Pi_j(\eta_u)) - K(\Pi_j(\hat{\theta}_{(u)}),\Pi_j(\tilde{\theta}_u))\right] + o_P(1)\,. \tag{2.15}$$

Let us define the function $\psi : \Xi\times\Xi\times\Xi \longrightarrow R^1$ by the formula

$$\psi(\hat{\theta},\theta^*,\theta^{**}) = K(\hat{\theta},\theta^*) - K(\hat{\theta},\theta^{**})\,. \tag{2.16}$$

Since (C2) holds, the set Ξ is open, and according to Lemma 7, chapter II in [9] also convex. If $\gamma \in \Xi$, then according to Theorem 9, chapter II in [9] derivatives of all orders of $\int e^{\gamma' x}\,\mathrm{d}\nu(x)$ may be computed by differentiating under the integration sign, and therefore $K(\gamma^*,\gamma)$ has on Ξ continuous partial derivatives of the first order. Thus applying the Taylor theorem on (2.16) we get that

$$\psi\left(\Pi_j(\hat\theta_{(u)}),\Pi_j(\eta_u),\Pi_j(\tilde\theta_u)\right)=\psi\left(\Pi_j(\theta),\Pi_j(\eta_u),\Pi_j(\tilde\theta_u)\right)+\mathrm{d}_u\,, \tag{2.17}$$

where

$$\mathrm{d}_u=\sum_{i=1}^{m}\frac{\partial\psi\Big(\alpha\Pi_j(\hat\theta_{(u)})+(1-\alpha)\Pi_j(\theta),\Pi_j(\eta_u),\Pi_j(\tilde\theta_u)\Big)}{\partial\Big(\alpha\Pi_j(\hat\theta_{(u)})+(1-\alpha)\Pi_j(\theta)\Big)_i}\Big(\Pi_j(\hat\theta_{(u)})-\Pi_j(\theta)\Big)_i\,. \tag{2.18}$$

Since the first derivatives are continuous, from (2.9), (2.4), (2.13) and (2.16) we get that

$$\frac{\partial\psi\Big(\alpha\Pi_j(\hat\theta_{(u)})+(1-\alpha)\Pi_j(\theta),\Pi_j(\eta_u),\Pi_j(\tilde\theta_u)\Big)}{\partial\Big(\alpha\Pi_j(\hat\theta_{(u)})+(1-\alpha)\Pi_j(\theta)\Big)_i}=$$

$$=\frac{\partial\psi\Big(\Pi_j(\theta),\Pi_j(\eta(\theta,p)),\Pi_j(\eta(\theta,p))\Big)}{\partial\Pi_j(\theta)_i}+o_P(1)=o_P(1)\;, \tag{2.19}$$

where $P=P_\theta$. Taking into account (2.18), (2.19), (2.10) and (C1) we see, that the absolute value of the remainder term

$$|d_u|\le o_P(1)\|\Pi_j(\hat\theta_{(u)})-\Pi_j(\theta)\|=o_P(1)O_P((n_u^{(j)})^{-1/2})=o_P(n_u^{-1/2})\,.$$

Hence (2.15)–(2.17) imply that

$$0\le n_u^{-1/2}\log\frac{L(x^{(u)},\Omega_0)}{L(x^{(u)},\eta_u)}=n_u^{1/2}\left[K(\theta,\Omega_0,p_u)-K(\theta,\tilde\theta_u,p_u)\right]+o_P(1)\le o_P(1)\,, \tag{2.20}$$

where the last inequality follows from the fact that $\tilde\theta_u\in\Omega_0$. Validity of (2.20) means that (2.6) is proved. □

Proof of Theorem 1.1. (I) Making use of Lemma 2.2 and the inequality

$$0\le\log\frac{L(x^{(u)},\Omega_1)}{L(x^{(u)},\theta)}\le\log\frac{L(x^{(u)},\Theta)}{L(x^{(u)},\theta)}$$

we see that (cf. (2.2))

$$(2n_u^{1/2})^{-1}T_u=n_u^{-1/2}R_u+o_P(1)\,,$$

and (1.27) follows from Lemma 2.1 and the inequality

$$\liminf_{u\to\infty}\sigma_u>0$$

which holds owing to (2.4) and $\theta\in\Theta-\Omega_0$.

(II) If $\theta_0\in\Omega_0$, then

$$T_u\le 2\log\frac{L(x^{(u)},\Theta)}{L(x^{(u)},\theta_0)}\,. \tag{2.21}$$

Since (C2) holds, the measure ν is not supported on a flat, which according to Lemma 2.1 in [2] means that the probabilities $\{\overline{P}_\gamma;\ \gamma \in \Xi\}$ are mutually different. Thus the Kullback–Leibler information quantity $K(\hat{\theta}_n, \gamma) \geq 0$ and the equality sign holds if and only if $\gamma = \hat{\theta}_n$. This together with (2.8) and (2.9) means, that for $\gamma \in \Xi$ almost everywhere P_γ

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{L(x_1,\dots,x_n,\Xi)}{L(x_1,\dots,x_n,\gamma)} = \lim_{n\to\infty} K(\hat{\theta}_n,\gamma) = 0\,. \tag{2.22}$$

Combining (2.21) with (C3) we obtain that

$$\log L_u(s) \leq -\frac{T_u(s)}{2} + h_u(T_u(s)) = -\frac{T_u(s)}{2} + n_u^{1/2} o_P(1) \tag{2.23}$$

where the last equality follows from (C3), (2.21), (2.22) and the law of large numbers. The equality (2.23) together with (1.27) and Lemma 1 yield (1.20). □

3. APPLICATION TO THE NORMAL DISTRIBUTION

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

$$\Xi = \{\gamma = (\mu', \sigma')' \in R^m;\ \mu \in R^k,\ \sigma \in R^a \text{ and } V(\sigma) \text{ is positive definite}\} \tag{3.1}$$

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

Theorem 3.1. Let us assume that $\Theta = \Xi^q$, the set Ω_0 from (1.3) satisfies (1.19) and $\theta \in \Omega_1 - \Omega_0$.

(I) The relations (1.7) imply (1.15).

(II) If Ω_0 is Ω_1 IR and if $\{T_u\}$ are the statistics (1.26), then under validity of (1.7) the convergence (1.27), (1.20) occurs and the statistics (1.26) are d-optimal for testing Ω_0 against $\Omega_1 - \Omega_0$.

Proof. If we put for $x \in R^k$ and $\gamma = (\mu',\sigma')' \in \Xi$

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

and analogously

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

then

$$e(\gamma)'T(x) = -\frac{1}{2}(x-\mu)'\,V(\sigma)^{-1}(x-\mu) + \frac{1}{2}\mu' V(\sigma)^{-1}\mu\,, \tag{3.2}$$

$f(x,\gamma) = \exp[e(\gamma)'T(x) - D(\gamma)]$ and e, e^{-1} are continuous mappings of Ξ onto Ξ. If the measure ν is on $\mathcal{B}^m$ defined by the formula $\nu(A) = \mu_L(T^{-1}A)$, where μ_L is the Lebesque measure on $(R^k, \mathcal{B}^k)$, and if $\overline{P}_{e(\gamma)}$ is the measure on $\mathcal{B}^m$ defined by means of the density

$$\tilde{f}(y, e(\gamma)) = \exp[e(\gamma)'y - C(e(\gamma))]$$

with respect to ν, then for any normal distributions P_γ, P_{γ^*} (with γ, γ^* from (3.1))

$$K(P_\gamma, P_{\gamma^*}) = K\left(\overline{P}_{e(\gamma)}, \overline{P}_{e(\gamma^*)}\right),$$

$$\mathrm{Var}\left(\log\frac{\tilde{f}(y,e(\gamma))}{\tilde{f}(y,e(\gamma^*))}\Big|\overline{P}_{e(\gamma)}\right) = \mathrm{Var}\Big((e(\gamma)-e(\gamma^*))'T(x)\Big|\overline{P}_\gamma\Big) = \mathrm{Var}\left(\log\frac{f(x,\gamma)}{f(x,\gamma^*)}\Big|\overline{P}_\gamma\right).$$

Obviously $\{T^{-1}(B);\, B \in \mathcal{B}^m\} = \mathcal{B}^k$, which implies that for every measurable function $M_u : R^{kn_u} \to R^1$ there exists a measurable function $\tilde{M}_u : R^{mn_u} \to R^1$ such that $M_u(x^{(u)}) = \tilde{M}_u(t^{(u)})$, with $t_i^{(j)} = T(x_i^{(j)})$ for $j = 1,\dots,q$, $i = 1,\dots,n_j^{(u)}$. Thus every test on parameters of the normal distribution can be identified with a test on parameters of the exponential family with the density $\tilde{f}$. Since according to Lemma 2.4 in [11] the measure ν is not supported on a flat and the natural set of parameters (1.8) coincides with the set (3.1) which is open, Theorem 3.1 will follow from Lemma 1.1 and Theorem 1.1, if we prove that (C3) holds. One can prove this by referring to Lemma 4.4 in [7]. Since this technical report may be not available to the reader, we prefer to prove the following lemma, from which (C3) obviously follows.

Lemma 3.1. For $\gamma = (\mu', \sigma')' \in \Xi$ let $f(x,\gamma)$ denote density of the normal distribution $N(\mu, V(\sigma))$. Let $n_u^{(j)}$ denote size of sample from the normal $N(\mu_j, V(\sigma_j))$ population, the real number $c > 0$, and in the notation

$$\hat{n}_u = \min\{n_u^{(1)}, \dots, n_u^{(q)}\} \tag{3.3}$$

the inequality

$$k + c < \hat{n}_u \tag{3.4}$$

holds. There exist numbers $h_u = h(n_u^{(1)}, \dots, n_u^{(q)}, k, c)$ such that under validity of (C1)

$$h_u = O(\log n_u) \tag{3.5}$$

and for every $\theta \in \Xi^q$ and $t > 0$ in the notation (1.21)

$$P_\theta\left[\log\frac{L(x^{(u)},\Theta)}{L(x^{(u)},\theta)} \ge t\right] \le \exp\left[-t + H_u(t)\right], \tag{3.6}$$

where

$$H_u(t) = \frac{k+c}{\hat{n}_u}\,t + h_u. \tag{3.7}$$

Proof. If $\gamma = (\mu', \sigma')'$, $\gamma^* = (\mu^{*\prime}, \sigma^{*\prime})'$ belong to (3.1), then the Kullback–Leibler information quantity

$$K(\gamma,\gamma^*) = \frac{1}{2}(\mu-\mu^*)'V(\sigma^*)^{-1}(\mu-\mu^*) + \frac{1}{2}\mathrm{tr}\left[V(\sigma)V(\sigma^*)^{-1}\right] - \frac{1}{2}\log\frac{|V(\sigma)|}{|V(\sigma^*)|} - \frac{k}{2}. \tag{3.8}$$

Hence if

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})', \quad \overline{x} = \sum_{i=1}^{n} x_i , \tag{3.9}$$

and $A_n = \{(x_1, \ldots, x_n) \in (R^k)^n ; |\hat{\Sigma}| > 0\}$, then one can easily verify that there exists a function $g_n : A_n \to R^1$ such that on A_n for each γ from (3.1)

$$\log \prod_{i=1}^{n} f(x_i, \gamma) = g_n(x_1, \ldots, x_n) - nK(\hat{\theta}_n, \gamma) , \tag{3.10}$$

where $\hat{\theta}_n$ is the parameter corresponding to the normal $N(\overline{x}, \hat{\Sigma})$ distribution. We shall proceed similarly as in the proof of Theorem 2.1 in [5]. Since for $\gamma = (\mu', \sigma')' \in \Xi$ and $n \geq k+1$ the equality $P_\gamma(A_n) = 1$ holds, and

$$\mathcal{L}[\, K(\hat{\theta}_n, \gamma) | \overline{P}_\gamma \,] = \mathcal{L}[\, K(\hat{\theta}_n, \vartheta^*) | \overline{P}_{\vartheta^*} \,]$$

where ϑ^* is the parameter corresponding to the normal $N(0, I_k)$ distribution, denoting $\vartheta = (\vartheta^*, \ldots, \vartheta^*) \in \Theta$ and utilizing the notations (2.14), (1.16) we see that for $\hat{n}_u \geq k+1$ and any positive real number $\tilde{t}$

$$P_\theta \left[\log \frac{L(x^{(u)}, \Theta)}{L(x^{(u)}, \theta)} \geq t \right] = P_\vartheta [n_u K(\hat{\theta}_{(u)}, \vartheta, p_u) \geq t] \leq$$

$$\leq \exp\left(-\tilde{t}t + \sum_{j=1}^{q} \varphi_{n_u^{(j)}}(\tilde{t} n_u^{(j)}) \right) , \tag{3.11}$$

where

$$\varphi_n(z) = \log \mathrm{E}_{\vartheta^*} \left[\exp(zK(\hat{\theta}_n, \vartheta^*)) \right] .$$

Employing the Bartlett decomposition of the Wishart matrix, described in [8], p. 55, and performing all necessary integrations, we get that for $z < n - k$

$$\begin{aligned} \varphi_n(z) \;=\; & -\frac{k}{2}\log\left(1 - \frac{z}{n}\right) + \frac{zk}{2}\log n + \left[\frac{k(k+1)}{4} - \frac{nk}{2}\right]\log 2 - \\ & - \sum_{i=1}^{k} \log\Gamma\left(\frac{n-i}{2}\right) + \sum_{i=1}^{k} \frac{z+i-n}{2} \log\left(\frac{1}{2} - \frac{z}{2n}\right) + \\ & + \sum_{i=1}^{k} \log\Gamma\left(\frac{n-i-z}{2}\right) - \frac{(k-1)k}{4}\log\left(1 - \frac{z}{n}\right) - \frac{kz}{2} , \end{aligned} \tag{3.12}$$

where Γ denotes the usual gamma function. According to the Stirling formula for logarithm of the gamma function (cf. (12.5.3) in [4])

$$\log\Gamma(x) = \left(x - \frac{1}{2}\right)\log x - x + \frac{1}{2}\log(2\pi) + o(1) , \tag{3.13}$$

where $\lim_{x\to\infty} o(1) = 0$. Combining (3.12) and (3.13) we obtain that for $\tilde{t} = 1 - (\hat{n}_u)^{-1}(k+c)$ under validity of (C1)

$$\varphi_{n_u^{(j)}}(\tilde{t}n_u^{(j)}) = O(\log n_u),$$

which together with (3.11) means that the lemma is true. □

In the following considerations we shall drop for $\gamma = (\mu', \sigma')' \in \Xi$ the notation $V(\sigma)$, and covariance matrix of the normal distribution with the density $f(x,\gamma)$ we shall denote simply by $V(\gamma)$.

Example 1. Testing the equality $\mu = \mu_0$. Let $\mu_0 \in R^k$ be a fixed vector and

$$\Omega_0 = \{\gamma \in \Xi;\ \mathrm{E}(x|\overline{P}_\gamma) = \mu_0\}. \tag{3.14}$$

If $\gamma \in \Xi$ and $\gamma^* \in \Omega_0$, then

$$K(\gamma,\gamma^*) = K\left[N(0,A), N(0,V(\gamma^*))\right] + \frac{1}{2}\log\left(|A|/|V(\gamma)|\right)$$

where $A = V(\gamma) + (\mu-\mu_0)(\mu-\mu_0)'$. Thus $K(\gamma,.)$ is on Ω_0 minimized at the unique parameter η corresponding to $N(\mu_0, A)$. If

$$S = \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' = n\hat{\Sigma} \tag{3.15}$$

and $T_n^2 = n(n-1)(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$ is the Hotelling statistic with $(n-1)$ degrees of freedom, then according to [12], p. 111

$$2\log\frac{L(x_1,\ldots,x_n,\Xi)}{L(x_1,\ldots,x_n,\Omega_0)} = n\log\left(1+\frac{1}{n-1}T_n^2\right). \tag{3.16}$$

Since increasing transformations preserve the level attained, from Theorem 3.1 we obtain that the Hotelling F-test based on $\frac{n(n-k)}{k}(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$ is d-optimal for testing the hypothesis $\mu = \mu_0$.

Example 2. Testing sphericity of the covariance matrix. Let

$$\Omega_0 = \{\gamma \in \Xi;\ \text{there exist } \sigma > 0 \text{ such that } V(\gamma) = \sigma^2 I_k\} \tag{3.17}$$

where I_k is the identity matrix. If $\gamma \in \Xi$ and $\gamma^* \in \Omega_0$, then

$$K(\gamma,\gamma^*) = (2\sigma^{*2})^{-1}\left[||\mu-\mu^*||^2 + \mathrm{tr}(V(\gamma))\right] + \frac{1}{2}\log\sigma^{*2k}|V^{-1}(\gamma)| - \frac{k}{2}.$$

Thus $K(\gamma,.)$ is on Ω_0 minimized at the unique parameter η corresponding to the $N(\mu, k^{-1}\mathrm{tr}(V(\gamma))I_k)$ distribution. From Theorem 3.1 we therefore obtain that the statistics (cf. (3.15))

$$T_n = 2\log\frac{L(x_1,\ldots,x_n,\Xi)}{L(x_1,\ldots,x_n,\Omega_0)} = n\log\left(\left[\mathrm{tr}(k^{-1}S)\right]^k/|S|\right) \tag{3.18}$$

are d-optimal for testing the sphericity hypothesis (3.17).

Example 3. Testing independence of sets of variates. Let $x' = (x_1', \ldots, x_r')$ be a partitioning of the vector $x \in R^k$. Since x is supposed to be normally distributed, independence of these subvectors corresponds to the hypothesis

$$\Omega_0 = \{\gamma \in \Xi;\ V_{ij}(\gamma) = 0 \text{ for all } i \neq j\} \tag{3.19}$$

where $V_{ij}(\gamma) = \operatorname{cov}(x_i, x_j | \overline{P}_\gamma)$. If $\gamma \in \Xi$ and $\gamma^* \in \Omega_0$, then

$$K(\gamma, \gamma^*) = \sum_{i=1}^{r} K\left[N(\mu_i, V_{ii}(\gamma)), N(\mu_i^*, V_{ii}(\gamma^*))\right] + \frac{1}{2}\log\left(|V^{-1}(\gamma)| \prod_{i=1}^{r} |V_{ii}(\gamma)|\right).$$

Thus $K(\gamma, .)$ is on Ω_0 minimized at the unique parameter corresponding to the normal $N(\mu, V^*)$ distribution, where $V_{ij}^* = 0$ if $i \neq j$ and $V_{ii}^* = V_{ii}(\gamma)$. Hence denoting the ith block of the matrix $\hat{\Sigma}$ from (3.15) by $\hat{\Sigma}_{ii}$ and taking into account Theorem 3.1 we see that the statistics

$$T_n = 2\log\frac{L(x_1, \ldots, x_n, \Xi)}{L(x_1, \ldots, x_n, \Omega_0)} = n\log\left[\Big(\prod_{i=1}^{r} |\hat{\Sigma}_{ii}|\Big)\Big/|\hat{\Sigma}|\right] \tag{3.20}$$

are d-optimal for testing the hypothesis of independence (3.19).

In the following two examples we assume that $q > 1$ and $\Theta = \Xi^q$.

Example 4. Testing equality of means. Let

$$\Omega_0 = \{\theta = (\theta_1, \ldots, \theta_q) \in \Theta;\ \mathrm{E}(x|\theta_1) = \ldots = E(x|\theta_q),\ \ V(\theta_1) = \ldots = V(\theta_q)\} \tag{3.21}$$

be the hypothesis that the means of the q normal populations are equal (and the usual assumption of equality of the covariance matrices is imposed). Let us denote

$$\Omega_1 = \{\theta = (\theta_1, \ldots, \theta_q) \in \Theta;\ V(\theta_1) = \ldots = V(\theta_q)\} \tag{3.22}$$

the alternative hypothesis which places no restriction on the means, but still assumes the equality of the covariances.

If $\theta \in \Omega_1$, $\theta^* \in \Omega_0$ and $p \in \mathcal{P}$, then in the notation $V = V(\theta_j)$, $V^* = V(\theta_j^*)$, $\mu^* = \mathrm{E}(x|\theta_j^*)$ and $A = \sum_j p_j(\mu_j - \mu^*)(\mu_j - \mu^*)' + V$ we get

$$K(\theta, \theta^*, p) = K\left[N(0, A), N(0, V^*)\right] + \frac{1}{2}\log\left(|A|/|V|\right).$$

Thus $K(\theta, ., p)$ is on Ω_0 minimized at the unique point $\eta = (\eta_1, \ldots, \eta_q)$, where $\eta_1 = \ldots = \eta_q$ correspond to the $N(\tilde{\mu}, \sum_j p_j(\mu_j - \tilde{\mu})(\mu_j - \tilde{\mu})' + V)$ distribution, with $\tilde{\mu} = \sum_j p_j\mu_j$, and Ω_0 is Ω_1 IR. If

$$\bar{x}_j,\ \hat{\Sigma}_j \tag{3.23}$$

denote the sample mean and the sample covariance matrix constructed from the sample drawn from the jth population, then making use of (1.28) we get after some computation that

$$T_u = 2\log\frac{L(x^{(u)},\Omega_1)}{L(x^{(u)},\Omega_0)} = 2n_u\Big(K(\hat\theta_{(u)},\Omega_0,p_u) - K(\hat\theta_{(u)},\Omega_1,p_u)\Big)$$
$$= -n_u\log\Lambda\,,\ \ \Lambda = \frac{|A|}{|A+B|} \tag{3.24}$$

where $A = \sum_j n_u^{(j)}\hat\Sigma_j$, $B = \sum_j n_u^{(j)}(\bar x_j - \hat\mu)(\bar x_j - \hat\mu)'$, $\hat\mu = \sum_j p_u^{(j)}\bar x_j$. From Theorem 3.1 we obtain that the statistics (3.24) are d-optimal for testing (3.21) against (3.22) (this d-optimality of course applies also to the Wilks statistic Λ with the level attained defined in this special case by the formula $L(s) = P[\Lambda \le \Lambda(s)]$).

Example 5. Testing equality of covariances. Let us denote

$$\Omega_0 = \{\theta = (\theta_1,\dots,\theta_q)\in\Theta;\, V(\theta_1) = \ldots = V(\theta_q)\} \tag{3.25}$$

the hypothesis that the covariance matrices of the q normal populations are equal. If $\theta\in\Theta$, $\theta^*\in\Omega_0$ and $p\in\mathcal{P}$, then in the notation $V(\theta_j) = V_j$, $V(\theta_1^*) = \ldots = V(\theta_q^*) = V^*$, $A = \sum_j p_j V_j$

$$K(\theta,\theta^*,p) =$$
$$= \frac{1}{2}\sum_j p_j(\mu_j - \mu_j^*)'V^{*-1}(\mu_j - \mu_j^*) + K\big[N(0,A),N(0,V^*)\big] + \frac{1}{2}\sum_j p_j\log\Big(|A|/|V_j|\Big)\,.$$

Thus the set Ω_0 is Θ IR and

$$K(\theta,\Omega_0,p) = \frac{1}{2}\sum_{j=1}^{q} p_j\log\Big(|A|/|V_j|\Big)\,. \tag{3.26}$$

In the notation (3.23) and $S_j = n_u^{(j)}\hat\Sigma_j$, $S = \sum_{j=1}^q S_j$

$$T_u = 2\log\frac{L(x^{(u)},\Theta)}{L(x^{(u)},\Omega_0)} = \log\tilde T_u\,,\qquad \tilde T_u = \left|\frac{1}{n_u}S\right|^{n_u}\Big/\prod_{j=1}^{q}\left|\frac{1}{n_u^{(j)}}S_j\right|^{n_u^{(j)}}\,. \tag{3.27}$$

As pointed out in [12], p.225, to obtain an unbiased test, instead of $\tilde T_u$ the modified statistic

$$T_u^* = \left|\frac{1}{n_u - q}S\right|^{n_u - q}\Big/\prod_{j=1}^{q}\left|\frac{1}{n_u^{(j)} - 1}S_j\right|^{n_u^{(j)}-1} \tag{3.28}$$

is used. We shall prove d-optimality of the statistic T_u^*.

Let (1.7) hold. There exist an index u_0 and a positive constant c such that $P_\theta\big[\tilde T_u/T_u^* \ge c\big] = 1$ for all $u \ge u_0$ and $\theta\in\Theta$. Thus if L_u^* is the level attained by T_u^* , then for $u \ge u_0$

$$L_u^*(s) \le \sup\left\{ P_{\theta^*}\Big[2\log\frac{L(x^{(u)},\Theta)}{L(x^{(u)},\theta^*)} \ge \log T_u^*(s) + \log c\Big]\,;\theta^*\in\Theta\right\}.$$

Let $\theta \in \Theta - \Omega_0$. From (3.26) and (3.28) we obtain that $n_u^{-1} \log T_u^* \longrightarrow 2K(\theta, \Omega_0, p)$ a.e. Since the ratio $\tilde{T}_u / T_u^*$ converges a.e. to a positive constant, and according to the proof of Theorem 3.1 the condition $(C3)$ is fulfilled,

$$n_u^{-1/2} \log L_u^*(s) \leq -(2n_u^{1/2})^{-1} T_u(s) + o_P(1),$$

where $P = P_\theta$. This together with (1.27) and Lemma 1 yields d-optimality of T_u^* for testing the hypothesis (3.25).

4. APPLICATION TO THE MULTINOMIAL DISTRIBUTION

Let $X = \{1, \ldots, k\}$ be a finite set,

$$\Xi = \left\{ (p_1, \ldots, p_{k-1})' \in R^{k-1};\ \min_i p_i > 0,\ \sum_{i=1}^{k-1} p_i < 1 \right\} \tag{4.1}$$

and

$$f(x,p) = p_x, \quad p_k = 1 - \sum_{j=1}^{k-1} p_j \tag{4.2}$$

denotes a density with respect to the counting measure μ on $(X, 2^X)$.

Theorem 4.1. Let us assume that $\Theta = \Xi^q$, the set Ω_0 from (1.3) satisfies (1.19) and $\theta \in \Omega_1 - \Omega_0$.

(I) The relations (1.7) imply (1.15).

(II) If Ω_0 is Ω_1 IR and if $\{T_u\}$ are the statistics (1.26), then under validity of (1.7) the convergence (1.27) and (1.20) occurs and the statistics (1.26) are d-optimal for testing Ω_0 against $\Omega_1 - \Omega_0$.

Proof. We shall proceed similarly as in proof of Theorem 3.1. After the identification

$$p \longrightarrow \left(\log \frac{p_1}{p_k}, \ldots, \log \frac{p_{k-1}}{p_k} \right)' \tag{4.3}$$

with the exponential family (1.9), where $\nu(A) = \text{card}\left[A \cap \{\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_{k-1}\}\right]$ and $\mathbf{0} = (0, \ldots, 0)'$, $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)'$ belong to R^{k-1}, we see that the set (1.8) of natural parameters $\Xi = R^{k-1}$, and the axiom (C2) is fulfilled. Further, let n_x denote the number of occurrences of x in $(x_1, \ldots, x_n) \in X^n$ and

$$\hat{\theta}_n = (\hat{p}_1, \ldots, \hat{p}_{k-1})', \quad \hat{p}_x = \frac{n_x}{n}. \tag{4.4}$$

Making use of the first equality in (4.2), the relation (2.4) in [6] and proceeding as in the proof of the inequality (2.10) in [6], we obtain that in the notation (2.14) and (4.4) for each $\theta \in \Theta$ and set $A \subset \overline{\Theta}$ (where $\overline{\Theta}$ denotes closure of Θ in the usual topology)

$$P_\theta(\hat{\theta}_{(u)} \in A) \leq \exp(-n_u K(A, \theta, p_u) + O(\log n_u)). \tag{4.5}$$

Since (3.10) holds also in this case, in the notation $B_u = \{\theta^* \in \Theta;\ n_u K(\theta^*, \theta, p_u) \geq t\}$

$$P_\theta \left(\log \frac{L(x^{(u)}, \Theta)}{L(x^{(u)}, \theta)} \geq t \right) = P_\theta(\hat{\theta}_{(u)} \in B_u)$$

and (C3) follows from (4.5). Thus the assumptions of Lemma 1.1 and Theorem 1.1 are fulfilled, and the assertion is proved. □

Example 6. Testing the simple hypothesis. Let $p_0 = (p_1^{(0)}, \ldots, p_{k-1}^{(0)})'$ be a fixed point from (4.1). From Theorem 4.1 we obtain that in the notation (4.4) the statistics

$$T_n = 2 \log \frac{L(x_1, \ldots, x_n, \Xi)}{L(x_1, \ldots, x_n, p_0)} = 2n \sum_{i=1}^{k} \hat{p}_i \log \frac{\hat{p}_i}{p_i^{(0)}} \tag{4.6}$$

where

$$0 \log x = 0 \tag{4.7}$$

are d-optimal for testing the hypothesis $p = p_0$ against $p \neq p_0$.

Example 7. Testing independence in contingency tables. Let in accordance with (4.1)

$$\Xi = \{(p_{11}, \ldots, p_{1s}, \ldots, p_{r1}, \ldots, p_{rs-1})' \in R^{rs-1};\ \min_{i,j} p_{i,j} > 0,\ \sum p_{ij} < 1\} \tag{4.8}$$

be the parametric set of the $r \times s$ contingency tables. Let

$$p_{i.} = \sum_{j=1}^{s} p_{ij}\,, \quad p_{.j} = \sum_{i=1}^{r} p_{ij}$$

where the number p_{rs} is defined analogously as in (4.2). Then

$$\Omega_0 = \{p \in \Xi;\ p_{ij} = p_{i.}p_{.j} \text{ for all } i, j\} \tag{4.9}$$

is the hypothesis, that the row and the column variables are stochastically independent. If the parameter $p \in \Xi$ is fixed, then making use of the Lagrange method of multipliers we find out that the parameter from (4.9) with $\tilde{p}_{ij} = p_{i.}p_{.j}$ for all i, j, is the unique parameter minimizing on Ω_0 the Kullback–Leibler information quantity $K(p, .)$. Thus the set Ω_0 is Ξ IR and according to Theorem 4.1 the likelihood ratio statistics (cf. (4.4) and (4.7))

$$T_n = 2 \log \frac{L(x_1, \ldots, x_n, \Xi)}{L(x_1, \ldots, x_n, \Omega_0)} = 2n \sum_{i=1}^{r} \sum_{j=1}^{s} \hat{p}_{ij} \log \frac{\hat{p}_{ij}}{\hat{p}_{i.}\hat{p}_{.j}} \tag{4.10}$$

are d-optimal for testing Ω_0 against $\Xi - \Omega_0$. We remark, that for $s = r = 2$ this d-optimality is proved in [1], p. 17.

Example 8. Testing equality of parameters of q multinomial populations. Let an integer $q > 1$ and $\Theta = \Xi^q$, where Ξ is the set (4.1). Let us denote

$$\Omega_0 = \{ \theta = (\theta_1, \ldots, \theta_q) \in \Theta; \theta_1 = \ldots = \theta_q \} \tag{4.11}$$

the hypothesis that the parameters of the q multinomial populations are the same. If $\theta \in \Theta$ and $p \in \mathcal{P}$ (cf. (1.11)) are fixed, then making use of the Lagrange method of multipliers we find out that $\eta = (\gamma, \ldots, \gamma)'$, where $\gamma = \sum_{j=1}^{q} p_j \theta_j$, is the unique point from (4.11) minimizing on Ω_0 the Kullback–Leibler information quantity $K(\theta, ., p)$. Thus the set Ω_0 is Θ IR, and if $x^{(u)}$ is the vector of samples (1.4) and o_{ji} denotes the number of occurrences of the element i in the sample $y(j, n_u^{(j)})$ from the jth population, then the likelihood ratio statistics

$$T_u(x^{(u)}) = 2 \log \frac{L(x^{(u)}, \Theta)}{L(x^{(u)}, \Omega_0)} = 2 \sum_{j=1}^{q} \sum_{i=1}^{k} o_{ji} \log \frac{o_{ji} o_{..}}{o_{j.} o_{.i}} \tag{4.12}$$

are according to Theorem 4.1 d-optimal for testing (4.11) against $\Theta - \Omega_0$.

5. APPLICATION TO THE POISSON DISTRIBUTION

Let $X = \{0, 1, 2, \ldots\}$,

$$\Xi = (0, +\infty) \tag{5.1}$$

and

$$f(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \tag{5.2}$$

be density of the Poisson distribution $\overline{P}_\lambda$ with respect to the counting measure μ on $(X, 2^X)$.

Theorem 5.1. Let us assume that $\Theta = \Xi^q$, the set Ω_0 from (1.3) satisfies (1.19) and $\theta \in \Omega_1 - \Omega_0$.

(I) The relations (1.7) imply (1.15).

(II) If Ω_0 is Ω_1 IR and $\{T_u\}$ are the statistics (1.26), then under validity of (1.7) the convergence (1.27) and (1.20) occurs and the statistics (1.26) are d-optimal for testing Ω_0 against $\Omega_1 - \Omega_0$.

Proof. We shall proceed similarly as in proof of Theorem 3.1. After the identification $\lambda \rightarrow \log \lambda$ with the exponential family (1.9), where $\nu(A) = \sum_{j=0}^{\infty} \chi_A(j)/j!$ with χ_A denoting the indicator function of the set A, we see that the set (1.8) of natural parameters $\Xi = R^1$ and the axiom (C2) is fulfilled. Since validity of the condition $(C3)$ follows from Lemma 4.3 in [7], and can be verified also by means of the relation (6.22) in [10], the assumptions of Lemma 1.1 and Theorem 1.1 are fulfilled, and the assertion is proved. □

Example 9. Testing equality of means. Let an integer $q > 1$ and $\Theta = \Xi^q$, where Ξ is the set (5.1). Let us denote

$$\Omega_0 = \{\, \theta = (\lambda_1, \ldots, \lambda_q) \in \Theta;\ \lambda_1 = \ldots = \lambda_q \,\} \tag{5.3}$$

the hypothesis that the parameters of the q Poisson populations are equal. If $\theta \in \Theta$ and $p \in \mathcal{P}$ are fixed, then $\eta = (\lambda^*, \ldots, \lambda^*)$, $\lambda^* = \sum p_j \lambda_j$, is the unique point from Ω_0, minimizing $K(\theta, ., p)$ and Ω_0 is $\Theta\ IR$. Hence according to Theorem 5.1 the LR test statistics

$$T_u = 2 \log \frac{L(x^{(u)}, \Theta)}{L(x^{(u)}, \Omega_0)} = 2 \sum_{j=1}^{q} n_u^{(j)} \left[\lambda^* - \hat{\lambda}_j + \hat{\lambda}_j \log \frac{\hat{\lambda}_j}{\lambda^*} \right], \tag{5.4}$$

where

$$\hat{\lambda}_j = (n_u^{(j)})^{-1} \sum_{i=1}^{n_u^{(j)}} x_i^{(j)}, \quad \lambda^* = \sum_{j=1}^{q} p_u^{(j)} \hat{\lambda}_j ,$$

are d-optimal for testing the hypothesis (5.3).

(Received February 23, 1993.)

REFERENCES

[1] R. R. Bahadur, T. K. Chandra and D. Lambert: Some further properties of likelihood ratios on general sample spaces. In: Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, 1981.

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

[3] R. H. Berk: Stochastic bounds for attained levels. J. Multivariate Anal. *14* (1984), 376–389.

[4] H. Cramér: Mathematical Methods of Statistics. Princeton University Press, Princeton 1946.

[5] D. G. Herr: Asymptotically optimal tests for multivariate normal distributions. Ann. Math. Statist. *38* (1967), 1829–1844.

[6] W. Hoeffding: Asymptotically optimal tests for multinomial distributions. Ann. Math. Statist. *36* (1965), 369–408.

[7] S. Kourouklis: Hodges-Lehmann optimality of the likelihood ratio test in regular expo nential families of distributions. Technical Report 73. The Pennsylvania State University, Dept. of Statistics, 1987.

[8] A. M. Kshirsagar: Multivariate Analysis. Marcel Dekker, New York 1972.

[9] E. L. Lehmann: Testing Statistical Hypotheses. Wiley, New York 1959.

[10] F. Rublík: On optimality of the LR tests in the sense of exact slopes. Part II. Application to individual distributions. Kybernetika *25* (1989), 117–135.

[11] F. Rublík: On consistency of the MLE. Kybernetika *31* (1995), 45–64.

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

RNDr. František Rublík, CSc., Institute of Measurement of SAS, Dúbravská cesta 9, 842 19 Bratislava. Slovak Republic.