František Rublík

On testing hypotheses approximable by cones

Mathematica Slovaca, Vol. 39 (1989), No. 2, 199--213

Persistent URL: http://dml.cz/dmlcz/130490

Math. Slovaca 39, 1989, No. 2, 199—213

ON TESTING HYPOTHESES APPROXIMABLE BY CONES

FRANTIŠEK RUBLÍK

1. Introduction

Let $(X, \mathscr{F})$ be a sample space, $\Theta \subset R^m$ be an open set and a family $\mathscr{P} = \{P_\theta;\ \theta \in \Theta\}$ of probability measures be defined by means of density functions

$$f(x, \theta) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\mu}(x),$$

where μ is a σ-finite measure on $\mathscr{F}$. We assume that the distribution of x belongs to the family $\mathscr{P}$. For every n let $x^{(n)} = (x_1, \ldots, x_n)$ denote n independent observations,

$$L(x^{(n)}, \theta) = \prod_{j=1}^{n} f(x_j, \theta) \tag{1.1}$$

be the corresponding density function and for $H \subset \Theta$ let

$$L(x^{(n)}, H) = \sup_{\theta \in H} L(x^{(n)}, \theta). \tag{1.2}$$

Let us consider testing the hypothesis H against the alternative $\Theta - H$ by means of the statistic

$$-2 \log \frac{L(x^{(n)}, H)}{L(x^{(n)}, \Theta)}, \tag{1.3}$$

where log denotes the logarithm to the base e. This statistic can be difficult to compute, partly because the function

$$\Psi(\theta^*) = \log L(x^{(n)}, \theta^*)$$

may have several maxima on H. However, in some situations the set H is convex, and if $\hat{\theta}_n$ denotes the maximum likelihood estimator MLE, then the Fisher information matrix $\mathbf{J}(\hat{\theta}_n)$ is of full rank almost everywhere. Hence almost everywhere the quadratic form

$$Q(\theta^*) = (\hat{\theta}_n - \theta^*)' \mathbf{J}(\hat{\theta}_n)(\hat{\theta}_n - \theta^*)$$

attains its minimum on the closure $\bar{H}$ in a unique point $\hat{\pi}_n^H(\hat{\theta}_n)$. We show in Theorem 2.1 that $\hat{\pi}_n^H(\hat{\theta}_n)$ is an estimator asymptotically equivalent to the restricted MLE. A statistic based on this projection estimator is used in Section 3 for the statistical quality control in the case of two normally distributed components with unknown correlation coefficient.

2. Main result

First we impose regularity conditions on the density functions and introduce several notations and notions.

Let θ be an arbitrary point in Θ.

(R1) *The function $f(x, \theta)$ is positive on X and has all partial derivatives of the third order, and they are continuous in θ.*

(R2) *There are a neighbourhood $U_\theta \subset \Theta$ of the point θ and a P_θ integrable function Q_θ such that*

$$\left| \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \log f(x, \theta) \right| \leqq Q_\theta(x)$$

for each $x \in X$, $\theta^* \in U_\theta$ and $i, j, k = 1, \ldots, m$.

(R3) *The vector*

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \left(\frac{\partial \log f(x, \theta)}{\partial \theta_1}, \ldots, \frac{\partial \log f(x, \theta)}{\partial \theta_m} \right)',$$

where the prime denotes the transpose of the vector, belongs to $L_2(P_\theta)$, and its covariance matrix $\mathbf{J}(\theta)$ is strictly positive definite.

(R4) *The equalities*

$$\int \frac{\partial}{\partial \theta_i} f(x, \theta) \, \mathrm{d}\mu(x) = 0, \quad \int \frac{\partial^2}{\partial \theta_j \partial \theta_i} f(x, \theta) \, \mathrm{d}\mu(x) = 0$$

hold for $i, j = 1, \ldots, m$.

The main assertion of this section concerns the testing of a hypothesis H. In accordance with [2] a set $H \subset \Theta$ is said to be approximable at $\theta \in H$ by a cone C if for every sequence $\{a_n\}$ of positive numbers tending to zero

$$\begin{aligned} &\sup \{v(y, c + \theta); y \in H, \|y - \theta\| \leqq a_n\} = o(a_n) \\ &\sup \{v(y, H - \theta); y \in C, \|y\| \leqq a_n\} = o(a_n). \end{aligned} \tag{2.1}$$

Here $v(y, A) = \inf\{\|y - z\|; z \in A\}$, $\|\cdot\|$ is the usual Euclidean norm on R^m and by the cone we mean any non-empty closed convex subset C of R^m such that $ay \in C$ whenever $a \geqq 0$ and $y \in C$.

We shall write in most cases $\mathbf{J}$ instead of $\mathbf{J}(\theta)$ and $P(A)$ instead of $P_\theta(A)$. Similarly, the symbols O_P, o_P relate to the measure $P = P_\theta$. Throughout this section θ will be an arbitrary point, fixed and belonging to Θ. It is shown in [6] that under the conditions (R1)—(R4) one can construct measurable mappings

$$\hat{\theta}_n : X^n \to \Theta \tag{2.2}$$

and measurable sets $A_n \subset X^n$ such that

$$\hat{\theta}_n = \theta + \mathbf{J}^{-1} n^{-1} \frac{\partial \log L(x^{(n)}, \theta)}{\partial \theta} + o_P(n^{-1/2}) \tag{2.3}$$

$$\frac{\partial \log L(x^{(n)}, \hat{\theta}_n)}{\partial \theta} = \boldsymbol{0} \quad \text{whenever } x^{(n)} \in A_n \tag{2.4}$$

$$P(A_n) \to 1 \quad \text{if} \quad n \to \infty. \tag{2.5}$$

If we denote for $y \in R^m$

$$|||y||| = (y'\mathbf{J}(\theta)y)^{1/2}, \quad |||y\hat{|||} = (y'\mathbf{J}(\hat{\theta}_n)y)^{1/2}, \tag{2.6}$$

then $|||\cdot|||$ and $|||\cdot\hat{|||}$ are norms on R^m and the following assertion holds.

Theorem 2.1. *Let $\theta \in H$. Let us assume that all the elements of the matrix $\mathbf{J}(\theta)$ possess partial derivatives of the first order continuous in θ. Let the set H be approximable at θ by a cone C, and measurable mappings $\tilde{\theta}_n^H : X^n \to H$ satisfv the conditions (cf. (1.2))*

$$\begin{aligned} &P[L(x^{(n)}, \tilde{\theta}_n^H) = L(x^{(n)}, H)] \to 1 \quad \textit{if} \quad n \to \infty \\ &\tilde{\theta}_n^H \to \theta \quad \textit{in } P_\theta \textit{ measure.} \end{aligned} \tag{2.7}$$

(I) Let $\bar{H}$ be the closure of the set H. If $\hat{\pi}_n^H(\hat{\theta}_n)$: $X^n \to \bar{H}$ are measurable mappings such that (cf. (2.6))

$$|||\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n)\hat{|||}^2 = \inf\{|||\hat{\theta}_n - \theta^*\hat{|||}^2\,;\ \theta^* \in H\} + o_P(n^{-1}), \tag{2.8}$$

then

$$\tilde{\theta}_n^H = \hat{\pi}_n^H(\hat{\theta}_n) + o_P(n^{-1/2}). \tag{2.9}$$

(II) If $P[L(x^{(n)}, \hat{\theta}_n) = L(x^{(n)}, \Theta)] \to 1$, then

$$\mathscr{L}[n(\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n))'\mathbf{J}(\hat{\theta}_n)(\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n)) \mid P_\theta] \to \mathscr{L}[g(z, C) \mid N(\boldsymbol{0}, \mathbf{J}^{-1})], \tag{2.10}$$

where $\to$ denotes the weak convergence of probability measures and the function g is defined by the formula

$$g(z, C) = \inf_{\theta^* \in C}(z - \theta^*)'\mathbf{J}(z - \theta^*). \tag{2.11}$$

Let us denote for any closed convex set $V \subset R^m$ by π^V the projection on V in the norm $|||\cdot|||$ (cf. (2.6)). Before proceeding to the proof of Theorem 2.1, we introduce the following assertion.

Lemma 2.1. *If the assumptions of the preceding theorem are fulfilled, then*

$$\tilde{\theta}_n^H = \pi^{C+\theta}(\hat{\theta}_n) + o_P(n^{-1/2}). \tag{2.12}$$

Proof. First we prove that

$$\log L(x^{(n)}, \tilde{\theta}_n^H) = \log L(x^{(n)}, \pi^{C+\theta}(\hat{\theta}_n)) + o_P(1). \tag{2.13}$$

According to the regularity conditions, (2.3) and the central limit theorem

$$\hat{\theta}_n = \theta + O_P(n^{-1/2}). \tag{2.14}$$

According to Lemma 4.1 in Appendix $|||\pi^C(y)||| \leqq |||y|||$, which together with

$$\pi^{C+\theta}(y) = \pi^C(y - \theta) + \theta \tag{2.15}$$

and with (2.14) implies the relation

$$\pi^{C+\theta}(\hat{\theta}_n) = \theta + O_P(n^{-1/2}). \tag{2.16}$$

Making use of (2.14), (2.16), Taylor's theorem, (2.4), (2.5) and the equality $E\left[\frac{\partial^2 \log f}{\partial \theta^2}\right] = -\mathbf{J}$, we get

$$\log L(x^{(n)}, \pi^{C+\theta}(\hat{\theta}_n)) = \log L(x^{(n)}, \hat{\theta}_n) - \frac{n}{2} g(\hat{\theta}_n, C + \theta) + o_P(1). \tag{2.17}$$

According to Lemma 1 in [2] under the validity of (2.7)

$$\tilde{\theta}_n^H = \theta + O_P(n^{-1/2}). \tag{2.18}$$

Hence given ε positive one can choose a positive constant M_ε such that the sets $H_n = \{\theta^* \in H;\ \|\theta^* - \theta\| < n^{-1/2} M_\varepsilon\}$ satisfy

$$\limsup_n P[\tilde{\theta}_n^H \notin H_n] < \varepsilon. \tag{2.19}$$

Combining (2.14), Taylor's theorem, (2.4), (R2) and the law of large numbers we obtain

$$\log L(x^{(n)}, H_n) = \log L(x^{(n)}, \hat{\theta}_n) - \frac{n}{2} g(\hat{\theta}_n, H_n) + o_P(1). \tag{2.20}$$

Since the sequence $\{n^{1/2}(\hat{\theta}_n - \theta)\}$ is bounded in probability, from (2.1) and

$$(z - \theta^*)'\mathbf{J}(z - \theta^*) - (z - \tilde{\theta})'\mathbf{J}(z - \tilde{\theta}) \leqq 2(z - \theta^*)'\mathbf{J}(\tilde{\theta} - \theta^*) \tag{2.21}$$

we get

$$g(\hat{\theta}_n, C + \theta) - g(\hat{\theta}_n, H_n) \leqq o_P(n^{-1}). \tag{2.22}$$

On the other hand, the inequality $|||\pi^C(y)||| \leqq |||y|||$ and the equivalence of the norms $\|\cdot\|$, $|||\cdot|||$ imply the existence of a positive constant K such that

$$g(y, C) = \inf\{\|y - \tilde{\theta}\|^2 ; \|\tilde{\theta}\| \leqq K\|y\|, \tilde{\theta} \in C\}.$$

Therefore denoting

$$B_n = \{x^{(n)} ; \|\hat{\theta}_n - \theta\| \leqq 2^{-1}K^{-1}n^{-1/2}M_\varepsilon\}$$

and taking into account (2.21) and (2.1) we see that

$$[g(\hat{\theta}_n, H_n) - g(\hat{\theta}_n, C + \theta)]\chi_{B_n} \leqq o_P(n^{-1}), \tag{2.23}$$

where χ_{B_n} is the indicator function of the set B_n. Since we can assume without loss of generality that the constant M_ε is chosen so that $\liminf_n P(B_n) > 1 - \varepsilon$, the relations (2.23), (2.22), (2.20), (2.17) and (2.19) lead to (2.13).

Further, making use of (2.14), (2.16), (2.18), Taylor's theorem, (R2), the law of large numbers, (2.4) and (2.5) we obtain

$$\begin{aligned}
&\log L(x^{(n)}, \pi^{C+\theta}(\hat{\theta}_n)) = \\
&= \log L(x^{(n)}, \tilde{\theta}_n^H) + (\pi^{C+\theta}(\hat{\theta}_n) - \tilde{\theta}_n^H)' \frac{\partial \log L(x^{(n)}, \pi^{C+\theta}(\hat{\theta}_n))}{\partial \theta} + \\
&\quad + \frac{n}{2} |||\pi^{C+\theta}(\hat{\theta}_n) - \tilde{\theta}_n^H|||^2 + o_P(1) = \\
&= \log L(x^{(n)}, \tilde{\theta}_n^H) - n(\pi^{C+\theta}(\hat{\theta}_n) - \tilde{\theta}_n^H)'\mathbf{J}(\pi^{C+\theta}(\hat{\theta}_n) - \hat{\theta}_n) + \\
&\quad + \frac{n}{2} |||\pi^{C+\theta}(\hat{\theta}_n) - \tilde{\theta}_n^H|||^2 + o_P(1).
\end{aligned} \tag{2.24}$$

Since $\tilde{\theta}_n^H \in H$, taking into account both (2.1) and (2.18) we get

$$\tilde{\theta}_n^H = \pi^{C+\theta}(\tilde{\theta}_n^H) + o_P(n^{-1/2}), \tag{2.25}$$

which together with (2.24), (2.16), (2.14) and (2.13) yields

$$-n(\pi^{C+\theta}(\hat{\theta}_n) - \pi^{C+\theta}(\tilde{\theta}_n^H))'\mathbf{J}(\pi^{C+\theta}(\hat{\theta}_n) - \hat{\theta}_n) + \frac{n}{2} |||\pi^{C+\theta}(\hat{\theta}_n) - \tilde{\theta}_n^H|||^2 = o_P(1).$$

But (2.15) holds, and according to Lemma 4.1 in the Appendix the first term of this equality is non-negative, which implies (2.12.)

Now we prove (2.9). It is clear from (2.12) that the relation

$$\pi^{C+\theta}(\hat{\theta}_n) = \hat{\pi}^H(\hat{\theta}_n) + o_P(n^{-1/2}) \tag{2.26}$$

implies (2.9). We shall prove this equality in two steps. The first one is to prove that

$$\pi^{C+\theta}(\hat{\theta}_n) = \hat{\pi}^{C+\theta}(\hat{\theta}_n) + o_P(n^{-1/2}), \tag{2.27}$$

where $\hat{\pi}^{C+\theta}$ is the projection on $C + \theta$ in the norm $|||\cdot\hat{|||}$, described in (2.6). Taking into account (2.15) and Lemma 4.1 from the Appendix we get

$$|||\pi^{C+\theta}(\hat{\theta}_n) - \hat{\pi}^{C+\theta}(\hat{\theta}_n)|||^2 \leqq |||\pi^C(\hat{\theta}_n - \theta) - \hat{\pi}^C(\hat{\theta}_n - \theta)|||^2 +$$
$$+ 2(\pi^C(\hat{\theta}_n - \theta) - \hat{\pi}^C(\hat{\theta}_n - \theta))'\mathbf{J}(\hat{\theta}_n - \theta - \pi^C(\hat{\theta}_n - \theta)).$$

Hence the equivalence of the norms $\|\cdot\|$, $|||\cdot|||$ means that the validity of (2.27) will be stablished by proving

$$|||\hat{\theta}_n - \theta - \hat{\pi}^C(\hat{\theta}_n - \theta)|||^2 - |||\hat{\theta}_n - \theta - \pi^C(\hat{\theta}_n - \theta)|||^2 = o_P(n^{-1}). \tag{2.28}$$

To do this, let us choose a positive constant Z_ε such that the sets $D_n = \{x^{(n)};\ n^{1/2}\|\hat{\theta}_n - \theta\| \leqq Z_\varepsilon\}$ satisfy the inequality

$$\liminf_n P(D_n) > 1 - \varepsilon. \tag{2.29}$$

Since the derivatives of the information matrix $\mathbf{J}$ are continuous,

$$\sup_{x^{(n)}\in D_n}\|\mathbf{J}(\hat{\theta}_n) - \mathbf{J}\| = O(n^{-1/2}). \tag{2.30}$$

Therefore if we denote for a positive define matrix $\mathfrak{J}$ by $s(\mathfrak{J})$ its smallest and by $G(\mathfrak{J})$ its greatest characteristic root, then making use of the fact that

$$|z'\mathbf{A}y| \leqq \|z\|\ \|\mathbf{A}\|\ \|y\| \tag{2.31}$$

implies $|s(\mathbf{J}) - s(\mathfrak{J})| \leqq \|\mathbf{J} - \mathfrak{J}\|$ and $|G(\mathbf{J}) - G(\mathfrak{J})| \leqq \|\mathbf{J} - \mathfrak{J}\|$, we get the existence of an integer n_0 such that

$$s = \inf_{n\geqq n_0}\inf_{x^{(n)}\in D_n} s(\mathbf{J}(\hat{\theta}_n)) \quad G = \sup_{n\geqq n_0}\sup_{x^{(n)}\in D_n} G(\mathbf{J}(\hat{\theta}_n))$$

are finite positive numbers. Hence from (2.30), (2.31) and from the inequalities

$$|||\hat{\pi}^C(y)\hat{|||} \leqq |||y\hat{|||} \quad |||\pi^C(y)||| \leqq |||y||| \tag{2.32}$$

we obtain

$$\sup_{x^{(n)}\in D_n}|\,|||\hat{\theta}_n - \theta - \hat{\pi}^C(\hat{\theta}_n - \theta)|||^2 - |||\hat{\theta}_n - \theta - \hat{\pi}^C(\hat{\theta}_n - \theta)\hat{|||}^2| = o(n^{-1}). \tag{2.33}$$

A similar process, applied to the metrics ϱ, $\hat{\varrho}$ induced by the norms $|||\cdot|||$ and $\widehat{|||\cdot|||}$ respectively, leads to

$$\sup_{x^{(n)}\in D_n}|\hat{\varrho}^2(\hat{\theta}_n - \theta, C) - \varrho^2(\hat{\theta}_n - \theta, C)| = o(n^{-1}),$$

which together with (2.33) and (2.29) yields (2.28).

Now when we know that (2.27) holds, we prove

$$\hat{\pi}^{C+\theta}(\hat{\theta}_n) = \hat{\pi}_n^H(\hat{\theta}_n) + o_P(n^{-1/2}). \tag{2.34}$$

Taking into account both (2.15) and Lemma 4.1, we get

$$\begin{aligned}|||\hat{\pi}^{C+\theta}(\hat{\theta}_n) - \hat{\pi}_n^H(\hat{\theta}_n)\widehat{|||}^2 &\leqq |||\hat{\pi}^{C+\theta}(\hat{\theta}_n) - \hat{\pi}_n^H(\hat{\theta}_n)\widehat{|||}^2 + \\ &+ 2[\hat{\pi}^{C+\theta}(\hat{\theta}_n) - \hat{\pi}^{C+\theta}(\hat{\pi}_n^H(\hat{\theta}_n))]'\mathbf{J}(\hat{\theta}_n)[\hat{\theta}_n - \hat{\pi}^{C+\theta}(\hat{\theta}_n)].\end{aligned} \tag{2.35}$$

However, (2.1) and (2.8) imply the relation

$$|||\hat{\pi}^{C+\theta}(\hat{\pi}_n^H(\hat{\theta}_n)) - \hat{\pi}_n^H(\hat{\theta}_n)\widehat{|||}^2\chi_{D_n} = o_P(n^{-1})$$

and substituting into (2.35) we see that the validity of (2.34) will be established by proving

$$\sup_{x^{(n)}\in D_n}|\hat{\varrho}^2(\hat{\theta}_n, H) - \hat{\varrho}^2(\hat{\theta}_n, C+\theta)| = o(n^{-1}). \tag{2.36}$$

Obviously, $\hat{\varrho}(\hat{\theta}_n, H) \leqq \hat{\varrho}(\hat{\theta}_n, \theta)$ implies the existence of a $O^* = O^*(n^{-1/2})$ such that for all $x^{(n)}\in D_n$ and $n \geqq n_0$

$$\hat{\varrho}(\hat{\theta}_n, H) = \hat{\varrho}(\hat{\theta}_n, H^*),$$

where $H^* = \{\theta^*\in H;\ \|\theta^* - \theta\| \leqq O^*\}$. Further, given $O(n^{-1/2})$ one can find a sequence $O^* = O^*(n^{-1/2})$ such that $\tilde{\theta}\in C$ and $\|\tilde{\theta}\| \leqq O(n^{-1/2})$ imply

$$\varrho(\tilde{\theta} + \theta, H) = \varrho(\tilde{\theta} + \theta, H^*).$$

Hence making use of (2.32), (2.21), (2.30), (2.31) and (2.1) we obtain (2.36), which completes the proof of the assertion (I).

Taking into account (2.36) and (2.16) we get

$$n|||\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n)\widehat{|||}^2 = ng(\hat{\theta}_n, C+\theta) + o_P(1), \tag{2.37}$$

which together with (2.3), the central limit theorem and the continuity of $g(\cdot, C)$ yields (2.10).

We remark that if the mappings $\hat{\pi}_n^H(\hat{\theta}_n)$ from Theorem 2.1 take their values in H, then by means of (2.9) one can prove

$$\log L(x^{(n)}, \hat{\pi}_n^H(\hat{\theta}_n)) = \log L(x^{(n)}, \tilde{\theta}_n^H) + o_P(1). \tag{2.38}$$

But (2.13) and (2.17) imply that

$$-2 \log \frac{L(x^{(n)}, \tilde{\theta}_n^H)}{L(x^{(n)}, \hat{\theta}_n)} = ng(\hat{\theta}_n, C + \theta) + o_P(1).$$

Combining this with (2.38), (2.37) and (2.10) we obtain the relation

$$\mathscr{L}\left[-2\log \frac{L(x^{(n)}, \hat{\pi}_n^H(\hat{\theta}_n))}{L(x^{(n)}, \hat{\theta}_n)} \middle| P_\theta \right] \to \mathscr{L}[g(z, C) \,|\, N(\boldsymbol{0}, \mathbf{J}^{-1})] \tag{2.39}$$

which can also serve as a basis for the construction of a test of the hypothesis H.

3. Applications

Let us denote

$$\Theta = \{(\mu_1, \mu_2, \sigma_1, \sigma_2, \varrho)'; \ \sigma_1 > 0, \ \sigma_2 > 0, \ |\varrho| < 1\}, \tag{3.1}$$

where θ' denotes the transpose of the column vector θ. For each parameter $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \varrho)'$ let $f(\cdot, \theta)$ be the density of the 2-dimensional normally distributed random variable $X = (X_1, X_2)'$, possessing the means μ_1, μ_2, the variances σ_1^2, σ_2^2 and the correlation coefficient ϱ. After some computation we obtain that the densities $\{f(\cdot, \theta); \theta \in \Theta\}$ satisfy the regularity conditions (R1)—(R4) from Section 2 and the information matrix

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1, & \mathbf{0} \\ \mathbf{0}, & \mathbf{J}_2 \end{pmatrix}, \tag{3.2}$$

where

$$\mathbf{J}_1 = \begin{pmatrix} \sigma_1^{-2}(1-\varrho^2)^{-1} & , & -\varrho[\sigma_1\sigma_2(1-\varrho^2)]^{-1} \\ -\varrho[\sigma_1\sigma_2(1-\varrho^2)]^{-1}, & & \sigma_2^{-2}(1-\varrho^2)^{-1} \end{pmatrix} \tag{3.3}$$

$$\mathbf{J}_2 = \begin{pmatrix} (2-\varrho^2)[\sigma_1^2(1-\varrho^2)]^{-1}, & -\varrho^2[\sigma_1\sigma_2(1-\varrho^2)]^{-1}, & -\varrho[\sigma_1(1-\varrho^2)]^{-1} \\ -\varrho^2[\sigma_1\sigma_2(1-\varrho^2)]^{-1}, & (2-\varrho^2)[\sigma_2^2(1-\varrho^2)]^{-1}, & -\varrho[\sigma_2(1-\varrho^2)]^{-1} \\ -\varrho[\sigma_1(1-\varrho^2)]^{-1}, & -\varrho[\sigma_2(1-\varrho^2)]^{-1}, & (1+\varrho^2)(1-\varrho^2)^{-2} \end{pmatrix}, \tag{3.4}$$

Let c be a positive constant, M_1, M_2 real numbers and

$$H = \{\theta \in \Theta; \ \mu_1 + c\sigma_1 \leqq M_1, \ \mu_2 + c\sigma_2 \leqq M_2\}. \tag{3.5}$$

Since according to [7] the probability $P_\theta(X_1 \leqq M_1, X_2 \leqq M_2)$ is a non-decreasing function of the correlation coefficient, one can easily prove the following statement.

Lemma 3.1. *(I) If c is the $(1 - \Delta/2)$-quantile of the one-dimensional standard normal distribution, then for each $\theta \in H$*

$$P_\theta(X_1 \leqq M_1, X_2 \leqq M_2) \geqq 1 - \Delta. \tag{3.6}$$

(II) The condition from (I) cannot be relaxed, i.e. if c is smaller than the $(1 - \Delta/2)$-quantile, then there is a parameter $\theta \in H$ for which (3.6) does not hold.

Thus (3.5) implies (3.6) and it is of statistical interest to test the hypothesis H under normality assumptions. To do this by means of (2.10), let for n independently observed values $x_1, \ldots, x_n$ of the vector X the symbol $\hat{\theta}_n$ denote the MLE of θ, i.e.

$$\hat{\theta}_n = (\bar{x}(1), \bar{x}(2), s_1, s_2, \hat{\varrho})', \tag{3.7}$$

where $\bar{x}(1)$, $\bar{x}(2)$ are sample means and s_1, s_2 are sample standard deviations of the corresponding coordinates, and $\hat{\varrho}$ is the sample correlation coefficient. Since according to [1] the equality (2.4) holds, $\{\hat{\theta}_n\}$ satisfy the conditions (2.2)—(2.5). If $\mathbf{J}(\theta)$ is the matrix (3.2) and $\pi^H(z, \theta)$ is the projection of z on the closure $\bar{H}$ in the norm $|||z||| = (z'\mathbf{J}(\theta)z)^{1/2}$, computed by means of the algorithm described by Lemma 4.2 from the Appendix, then the mapping π^H is continuous. Hence the mappings $\hat{\pi}_n^H(\hat{\theta}_n) = \pi^H(\hat{\theta}_n, \hat{\theta}_n)$ satisfy the assumptions of Theorem 2.1 and the following assertion holds.

Theorem 3.1. *Let the constant c in (3.5) be greater than 1. If $t > 0$, then*

$$\sup_{\theta \in H} \lim_{n \to \infty} P_\theta(T_n > t) = $$
$$= 1 - [\pi^{-1} \operatorname{arc\,tg}(\gamma^{-1}) + 2^{-1} F_1(t) + \pi^{-1} \operatorname{arc\,tg}(\gamma) F_2(t)]. \tag{3.8}$$

In this notation

$$T_n = n(\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n))' \mathbf{J}(\hat{\theta}_n)(\hat{\theta}_n - \hat{\pi}_n^H(\hat{\theta}_n)),$$

F_j is the chi-square distribution with j degrees of freedom, the function arc tg *takes its values in the interval $(-\pi/2, \pi/2)$ and*

$$\gamma = [1 + 2((c^2 + 1)^2 - 2)^{-1}]^{1/2}. \tag{3.9}$$

Moreover, if $\theta \in \Theta - H$, then

$$\lim_{n \to \infty} P_\theta[T_n > M] = 1 \tag{3.10}$$

for every real number M.

The meaning of this theorem is obvious. If $t = t(\alpha, c)$ is the number for which (3.8) equals α and if Ψ_n is the test rejecting (3.5) if $T_n > t$ and accepting H otherwise, then $\{\Psi_n\}$ are consistent tests of H of the asymptotic size α.

The proof of the theorem is an application of Theorem 2.1. To verify its assumptions, we use the following lemma, where $\lambda_1(\theta^*)$ stands for the greater and $\lambda_2(\theta^*)$ stands for the smaller characteristic root of the covariance matrix, corresponding to the parameter θ^*. We remark that the following lemma is formulated generally, not especially for the case (3.5). It will be clear from the proof that the assertions of the lemma remain valid if Θ is replaced with the parameter set of the regular m-dimensional normal distributions and m is any positive integer.

Lemma 3.2. *Let $\theta \in \Theta$ and $H = \Theta \cap S$, where S is a closed set.*

(I) There exist positive numbers $d_1 > d_2$ and a positive number d such that if

$$\begin{aligned} W_1 &= \{\theta^* \in \Theta;\ \lambda_1(\theta^*) > d_1 \text{ or } \lambda_2(\theta^*) < d_2\} \\ W_2 &= \{\theta^* \in \Theta - W_1;\ \|\mu^* - \mu\| > d\}, \end{aligned} \tag{3.11}$$

where μ^* is the vector of means corresponding to the parameter θ^*, then the random variables $L(\cdot, W_j)/L(\cdot, \theta)$ tend to zero in probability P_θ for $j = 1, 2$.

(II) If $\theta \in H$, then there exist measurable mappings $\tilde{\theta}_n^H : X^n \to H$ such that (2.7) *holds.*

Proof. (I) Making use of the law of large numbers one obtains

$$\frac{1}{n} \log L(x^{(n)}, \theta) = -\log 2\pi - 2^{-1} \log |\Sigma| - 1 + o_P(1), \tag{3.12}$$

where Σ is the covariance matrix corresponding to θ. If $\lambda_1(A) \geqq \lambda_2(A)$ denote characteristic roots of the positive definite matrix A, then according to Theorem 1.10.2 in [8]

$$\operatorname{tr}(A\Sigma^{*-1}) \geqq \sum_{j=1}^{2} \frac{\lambda_j(A)}{\lambda_j(\Sigma^*)}. \tag{3.13}$$

Hence if we put $s = \lambda_2(\theta)$ and denote $\eta(z) = \log z + s(2z)^{-1}$, then making use of the law of large numbers we get that with probability tending to 1

$$\frac{1}{n} \log L(x^{(n)}, \theta^*) \leqq -\log 2\pi - 2^{-1} \sum_{j=1}^{2} \eta(\lambda_j(\theta^*)). \tag{3.14}$$

But $\eta(z)$ attains its minimum atz $z = s/2$ and $\eta(z) \to \infty$ if $z \to 0$ or $z \to \infty$, and for this reason (3.12) and (3.14) imply the existence of d_1 and d_2. The assertion on W_2 can be proved similarly by means of $\eta(z) \geqq \eta(s/2)$.

(II) If we denote $K = H - W_1 - W_2$, then the assertion (I) implies that

$$\lim_{n \to \infty} P[L(x^{(n)}, K) = L(x^{(n)}, H)] = 1.$$

Since $\lambda_1(\theta^*) \geqq \max\{\sigma_1^{*2}, \sigma_2^{*2}\}, \lambda_2(\theta^*) \leqq \min\{\sigma_1^2, \sigma_2^2\}$ and $|\Sigma^*| = \sigma_1^{*2}\sigma_2^{*2}(1 - \varrho^{*2})$, it is obvious that the set K is compact. This according to Lemma 3.3 in Section 5.3 of [6] means that there exist measurable mappings $\tilde{\theta}_n^H : X^n \to K$ such that

$L(x^{(n)}, \tilde{\theta}_n^H) = L(x^{(n)}, K)$. Taking into account results of [9] we see that $\tilde{\theta}_n^H \to \theta$ in probability, hence (2.7) holds.

Now we may apply Theorem 2.1. Since $\mathbf{J} = \mathbf{DND}$, where

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}_1, & \mathbf{0} \\ \mathbf{0}, & \mathbf{N}_2 \end{pmatrix} \quad \mathbf{N}_1 = \begin{pmatrix} 1, & -\varrho \\ -\varrho, & 1 \end{pmatrix}$$

$$\mathbf{N}_2 = \begin{pmatrix} 2-\varrho^2, & -\varrho^2, & -\varrho \\ -\varrho^2, & 2-\varrho^2, & -\varrho \\ -\varrho, & -\varrho, & (1+\varrho^2)/(1-\varrho^2) \end{pmatrix} \tag{3.15}$$

and $\mathbf{D}$ is a diagonal matrix with the diagonal $\sigma_1^{-1}(1-\varrho^2)^{-1/2}$, $\sigma_2^{-1}(1-\varrho^2)^{-1/2}$, $\sigma_1^{-1}(1-\varrho^2)^{-1/2}$, $\sigma_2^{-1}(1-\varrho^2)^{-1/2}$, $(1-\varrho^2)^{-1/2}$ we obtain from (2.10) that for every t positive

$$\lim_{n\to\infty} P_\theta(T_n > t) = P(v^2(z, \mathbf{N}^{1/2}\mathbf{D}C) > t \mid N(\mathbf{0}, \mathbf{I}_5)). \tag{3.16}$$

In this notation v is the distance from a set in the usual Euclidean norm and C is the the cone by means of which the set H is approximable at θ. But H is approximable by R^5 if θ is an inner point of H, by the cone $\{x \in R^5;\ x_1 + c x_3 \leqq 0\}$ if $\mu_1 + c\sigma_1 = M_1$, $\mu_2 + c\sigma_2 < M_2$ and in the case $\mu_i + c\sigma_i = M_i$, $i = 1, 2$, the set H is approximable at θ by the cone

$$C = \{x \in R^5;\ x_1 + cx_3 \leqq 0,\ x_2 + cx_4 \leqq 0\}.$$

Thus using this notation, denoting the left-hand side of (3.8) by P and taking into account the equality $\mathbf{D}C = C$ we see that

$$P = \sup_{|\varrho|<1} P[v^2(z, \mathbf{N}^{1/2}C) > t \mid N(\mathbf{0}, \mathbf{I}_5)]. \tag{3.17}$$

One can easily find out that

$$\mathbf{N}_1^{1/2} = (\boldsymbol{s}_1, \boldsymbol{s}_2)\begin{pmatrix} (1+\varrho)^{1/2}, & 0 \\ 0, & (1-\varrho)^{1/2} \end{pmatrix}\begin{pmatrix} \boldsymbol{s}_1' \\ \boldsymbol{s}_2' \end{pmatrix},$$

$$\boldsymbol{s}_1 = \begin{pmatrix} 2^{-1/2} \\ -2^{-1/2} \end{pmatrix}, \quad \boldsymbol{s}_2 = \begin{pmatrix} 2^{-1/2} \\ 2^{-1/2} \end{pmatrix}. \tag{3.18}$$

Solving the equations $|\mathbf{N}_2 - \lambda \mathbf{I}_3| = 0$ we get

$$\lambda_1 = 2 \quad \lambda_2 = [2(1-\varrho^2)]^{-1}(\varepsilon + \alpha) \quad \lambda_3 = [2(1-\varrho^2)]^{-1}(\varepsilon - \alpha)$$
$$\varepsilon = 3(1-\varrho^2) + 2\varrho^4 \quad \alpha = [1 - 2\varrho^2 + 13\varrho^4 - 12\varrho^6 + 4\varrho^8]^{1/2}.$$

Since $\alpha^2 = \varepsilon^2 - 8(1-\varrho^2)^2 > 0$ if $|\varrho| < 1$, these expressions are well-defined. Let us denote

$$\boldsymbol{p}_1 = \begin{pmatrix} 2^{-1/2} \\ -2^{-1/2} \\ 0 \end{pmatrix} \quad \boldsymbol{p}_2 = \begin{pmatrix} -[(4\alpha)^{-1}(\alpha-\gamma)]^{1/2} \\ -[(4\alpha)^{-1}(\alpha-\gamma)]^{1/2} \\ [(2\alpha)^{-1}(\alpha+\gamma)]^{1/2}\operatorname{sign}\varrho \end{pmatrix}$$

$$\boldsymbol{p}_3 = \begin{pmatrix} [(4\alpha)^{-1}(\alpha + \gamma)]^{1/2} \\ [(4\alpha)^{-1}(\alpha + \gamma)]^{1/2} \\ [(2\alpha)^{-1}(\alpha - \gamma)]^{1/2} \operatorname{sign} \varrho \end{pmatrix} \tag{3.19}$$

$$\gamma = -1 + 5\varrho^2 - 2\varrho^4,$$

where $\operatorname{sign} \varrho$ equals 1 or -1 for $\varrho \geqq 0$ or $\varrho < 0$, respectively. Since $\alpha^2 - \gamma^2 > 0$ if $0 < |\varrho| < 1$, these vectors are well-defined. But $\mathbf{N}_2 \boldsymbol{p}_j = \lambda_j \boldsymbol{p}_j$ and the vectors $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3$ are orthonormal, from which

$$\mathbf{N}_2^{1/2} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) \begin{pmatrix} \lambda_1^{1/2}, & 0, & 0 \\ 0, & \lambda_2^{1/2}, & 0 \\ 0, & 0, & \lambda_3^{1/2} \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_1' \\ \boldsymbol{p}_2' \\ \boldsymbol{p}_3' \end{pmatrix}. \tag{3.20}$$

Substituting (3.20) and (3.18) into (3.17) and utilizing the fact that the orthogonal transformation

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{s}_1' & 0 \\ \boldsymbol{s}_2' & \\ & \boldsymbol{p}_1' \\ 0 & \boldsymbol{p}_2' \\ & \boldsymbol{p}_3' \end{pmatrix} \boldsymbol{z} \tag{3.21}$$

preserves $N(\boldsymbol{O}, \mathbf{I}_5)$, we get

$$P = \sup_{|\varrho| < 1} P[v^2(y, C_1) > t \mid N(\boldsymbol{O}, \mathbf{I}_5)]. \tag{3.22}$$

In this notation $C_1 = \{x \in R^5 ;\ w_1'x \leqq -|w_2'x|\}$ and

$$w_1(1) = w_1(3) = 0, \quad w_1(2) = [2(1 - \varrho)]^{-1/2}$$
$$w_1(4) = -c[(\alpha - \gamma)(1 - \varrho^2)]^{1/2}[2\alpha(\varepsilon + \alpha)]^{-1/2}$$
$$w_1(5) = c[(\alpha + \gamma)(1 - \varrho^2)]^{1/2}[2\alpha(\varepsilon - \alpha)]^{-1/2}$$
$$w_2' = ([2(1 + \varrho)]^{-1/2}, 0, c/2, 0, 0).$$

Applying to these vectors the Gram–Schmidt orthogonalization procedure one can find an orthonormal basis $\tilde{w}_1, \ldots, \tilde{w}_5$ of R^5 such that $\tilde{w}_j = \|w_j\|^{-1} w_j$ for $j = 1, 2$. Therefore if $\mathbf{W}$ is the matrix whose jth row is the transpose of $\tilde{w}_j$ for $j = 1, \ldots, 5$, then making use of the transformation $u = \mathbf{W}y$ we see that (3.22) holds with $C_1 = \{z \in R^5 ;\ \|w_1\| z_1 \leqq -\|w_2\| |z_2|\}$. But

$$\|w_2\|^2 \|w_1\|^{-2} = (2 + c^2 + c^2\varrho)(1 - \varrho)(2 + c^2 + 2\varrho + c^2\varrho^2)^{-1}$$

and since $c > 1$, this function attains its maximum for $\varrho = -c^{-2}$. This means that

$$P = P[v^2(y, D) > t \mid N(\boldsymbol{O}, \mathbf{I}_2)], \tag{3.23}$$

where $D = \{x \in R^2 ;\ x_1 \leqq -\gamma|x_2|\}$ and γ is the number (3.9). But according to

Lemma 1 in [4]

$$P[v^2(y, D) \leqq t \,|\, N(\mathbf{0}, \mathbf{I}_2)] =$$
$$= [2^{-1} - \pi^{-1} \operatorname{arc\,tan}(\gamma)] + 2^{-1} F_1(t) + [2^{-1} - \pi^{-1} \operatorname{arc\,tan}(\gamma^{-1})] F_2(t)$$

and taking into account both (3.23) and the formula $\operatorname{arc\,tan}(\gamma) + \operatorname{arc\,tan}(\gamma^{-1}) = \pi/2$ we obtain (3.8).

If $\theta \in \Theta - H$, then making use of $\hat{\theta}_n \to \theta$, $\theta \notin H$ one can easily prove (3.10).

We remark that if $0 < c \leqq 1$, then the function $\|w_2\|^2 \|w_1\|^{-2}$ attains its maximum for $\varrho = -1$. Thus in this case (3.23) and (3.9) hold with $\gamma = 2^{1/2} c^{-1}$, and (3.10) also remains valid.

Theorem 3.1 of this section deals with the quality control of two possibly correlated normally distributed components. We remark that quality control of finitely many independent normally distributed components was investigated in [4] and [5]. In these papers formulas for the MLE of the parameter under the constraints $\mu_j + c\sigma_j \leqq M_j$, $\mu_j - c\sigma_j \geqq m_j$ and the asymptotic distribution of the maximum likelihood ratio test statistic are presented.

4. Appendix.

The topic of this section are some basic properties of projections on convex sets and on finite intersection of half-spaces.

Lemma 4.1. *Let C be a closed convex subset of a Hilbert space L, and π^C be the projection on the set C.*

(I) If $z \in C$, then $(y - \pi^C(y), z - \pi^C(y)) \leqq 0$ and the distance $v(y, \pi^C(y) + a(z - \pi^C(y)))$ is increasing on the segment $\overline{z, \pi^C(y)}$ towards z.

(II) If C is a cone, then the vectors $y - \pi^C(y)$, $\pi^C(y)$ are orthogonal, $\|\pi^C(y)\| \leqq \|y\|$ and $(y - \pi^C(y), z) \leqq 0$ for each $z \in C$.

Proof. Assertion (I) can be found in [3, p. 69]. If C is a cone, then the function $g(a) = \|y - a\pi^C(y)\|^2$ attains its minimum in $a = 1$, and therefore $g'(1) = 0$. Hence $(y, \pi^C(y)) = \|\pi^C(y)\|^2$, the vectors $y - \pi^C(y)$, $\pi^C(y)$ are orthogonal and the lemma is proved.

Let $\mathbf{J}$ be a symmetric positive definite $m \times m$ matrix and R^m be the Hilbert space with the inner product $[x, y] = x'\mathbf{J}y$. If $a \in R^m$ is a non-zero vector and b is a real number, then $C = \{x \in R^m ; a'x + b \leqq 0\}$ is a convex set. It is easy to verify that the mapping

$$\pi^C(x) = \begin{cases} x - (a'\mathbf{J}^{-1}a)^{-1}(a'x + b)\,\mathbf{J}^{-1}a & x \notin C \\ x & x \in C \end{cases} \tag{4.1}$$

is the projection of x on C in the norm

$$|||x||| = (x'\mathbf{J}x)^{1/2}. \tag{4.2}$$

The formula (4.1) obviously remains valid if the inequality in C is replaced with equality to zero. If $a_1, \ldots, a_n$ are non-zero vectors, $b_1, \ldots, b_n$ are real numbers and

$$C(j) = \{x \in R^m ; a_j' x + b_j \leqq 0\} \quad C(j_1, \ldots, j_r) = \bigcap_{s=1}^{r} C(j_s),$$

then the following assertion holds.

Lemma 4.2. *If π^C denotes the projection of x on the set $C = C(12 \ldots n)$ in the norm (4.2), then*

$$\pi^C(x) = \begin{cases} \pi^{C(23 \ldots n)}(x) & \text{if } \pi^{C(23 \ldots n)}(x) \in C(1) \\ \pi^{C(13 \ldots n)}(x) & \text{if } \pi^{C(23 \ldots n)}(x) \notin C(1),\ \pi^{C(13 \ldots n)}(x) \in C(2) \\ \vdots & \\ \pi^{C(12 \ldots n-1)}(x) & \text{if } \pi^{C(1 \ldots j-1 j+1 \ldots n)}(x) \notin C(j) \quad j = 1, \ldots, n-1, \\ & \quad \pi^{C(1 \ldots n-1)}(x) \in C(n) \\ \pi^H(x) & \text{if } \pi^{C(1 \ldots j-1 j+1 \ldots n)}(x) \notin C(j) \quad j = 1, \ldots, n, \end{cases}$$

where $H = \{x \in R^m ; c_j' x + b_j = 0\ j = 1, \ldots, n\}$.

Proof. Let us assume that $\pi^{C(1 \ldots j-1 j+1 \ldots n)}(x) \notin C(j)$. Then the segment S with the end-points $\pi^C(x)$, $\pi^{C(1 \ldots j-1 j+1 \ldots n)}(x)$ is a subset of $C(1 \ldots j-1 j+1 \ldots n)$ and the distance $v(z, x)$ increases on S towards $\pi^C(x)$. If $a_j' \pi^C(x) + b_j < 0$, then the function $g(z) = a_j' z + b_j$ attains different signs at the end-points of S. This implies the existence of an internal point $z \in S$ such that $g(z) = 0$. Hence $z \in C(1 \ldots j-1 j+1 \ldots n) \cap C(j)$ and since $v(z, x) < v(\pi^C(x), x)$, we obtain a contradiction. Hence $g(\pi^C(x)) = 0$ and the lemma is proved.

Since the closure of the set (3.5) is determined by the inequalities

$$\mu_1 + c\sigma_1 - M_1 \leqq 0, \quad \mu_2 + c\sigma_2 - M_2 \leqq 0,$$
$$-\sigma_1 \leqq 0, \quad -\sigma_2 \leqq 0, \quad \varrho - 1 \leqq 0, \quad -\varrho - 1 \leqq 0$$

and the projection on the intersection of the hyperplanes determined by some of these inequalities can be computed by a derivation, the previous lemma yields a recurrent algorithm for the computation of the projection of the estimate $\hat{\theta}_n$ on the set H in the metric, induced by the Fisher information matrix $\mathbf{J}(\hat{\theta}_n)$.

REFERENCES

[1] ANDERSON, T. W.: Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York 1958.

[2] CHERNOFF, H.: On the distribution of the likelihood ratio. Ann. Math. Statist. 25, 1954, 573—578.

[3] LUENBERGER, D. G.: Optimization by Vector Space Methods. John Wiley and Sons, New York 1969.

[4] RUBLÍK, F.: On the two-sided quality control. Apl. Mat. 27. 1982, 87—95.

[5] RUBLÍK, F.: On the two-sided multiparameter control. Math. Slovaca 33, 1983, 347—355.
[6] SCHMETTERER, L.: Introduction to Mathematical Statistics. Springer-Verlag, Berlin 1974.
[7] SLEPIAN, D.: The one-sided barrier problem for Gaussian noise. Bell System Tech. J. 41, 1962, 463—501.
[8] SRIVASTAVA, M. S.—KHATRI, C. G.: An Introduction to Multivariate Statistics. North Holland, New York 1979.
[9] WALD, A.: A note on consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 1949, 595—601.

Received February 15, 1985

Ústav merania a meracej techniky
CEFV SAV
Dúbravská cesta 9
842 19 Bratislava

О ПРОВЕРКЕ ГИПОТЕЗ АППРОКСИМОВАТЕЛЬНЫХ КОНУСАМИ

František Rublík

Резюме

Пусть $\mathbf{J}(\theta)$ обозначает информационную матрицу Фишера. Если $\hat{\theta}$-оценка максимального правдоподобия неизвестного параметра θ, тогда выборочной информационной матрицей мы называем матрицу $\hat{\mathbf{J}} = \mathbf{J}(\hat{\theta})$. В статье показано, что при условиях регулярности, обеспечивающих асимптотическую нормальность оценки максимального правдоподобия, и при предположении аппроксимовательности гипотезы H конусом, оценка максимального правдоподобия параметра при условии H является асимптотически эквивалентной с проекцией ОМП на H, вычисленной по норме, порождённой выборочной информационной матрицей. Это позволяет найти асимптотическое распределение расстояния ОМП от гипотезы H в этой метрике. Это расстояние использовано для статистического контроля качества двухмерных нормальных величин с регулярной дисперсионной матрицей.