сообщения
объединенного
института
ядерных
исследований
дубна

E5-88-98

N.I.Chernov, Ts.I.Ioseliani*, V.S.Kurbatov,
G.A.Ososkov

EFFICIENCY OF THE ESTIMATORS OF MULTIVARIATE DISTRIBUTION PARAMETERS FROM THE ONE-DIMENSIONAL OBSERVED FREQUENCIES

*IHEP TSU, Tbilisi

1988

1. <u>Introduction</u>. The present paper continues the investigation of statistical methods in analysis of multivariate distributions which has been started in the paper [1], where the problem of parameter estimation from a grouped sample for a hypothetical distribution of a random vector was considered. The problem arised because of the condition that only one-dimensional (marginal) observed frequencies for each single variable (component of the vector) were available. This precludes the application of the classical χ^2-test for goodness-of-fit testing parameter estimation. For lack of an appropriate theory the above problem was usually being solved by a certain method based on summing up χ^2's constructed for every one-dimensional (marginal) distribution. The incorrectness of the above heuristic method was shown in [1]. The correct T_m statistic that has an asymptotic χ^2-distribution in the case of testing simple hypothesis as well as in the case of parameter estimation was proposed in that paper too.

However, the conventional parameter estimators minimizing χ^2 are asymptotically efficient among unbiased estimators from the grouped sample because they are asymptotically equivalent to the maximum likelihood estimators for multinomial distributions (see [2], p. 426). Our statistic T_m depends on incomplete grouped frequencies, so the question of the efficiency of the estimator minimizing T_m among all the unbiased estimators cannot be solved within the framework of the common theory based on Rao-Cramer inequality. Nevertheless we succeded in proving its asymptotical efficiency in the class of estimators minimizing statistics that quadratically depend on the marginal observed frequencies. This statement is proved in the present paper. As an quantitative illustration we compare the efficiency of the estimate minimizing T_m with that of the estimate minimizing the marginal χ^2's mentioned above using the numerical example from [1]. The first estimate proves to be much more efficient. For the example we have chosen a physical model - so called isobar model of pion-nucleon interactions [3] with the production of an additional pion. The model as well as the setup of the whole problem was taken from experimental high-energy physics.

2. Formulation of the problem and necessary conventions. Now we give the exact formulation of our problem preserving all the notation from the parer /1/.

Let F be the distribution of a random vector (E_1, E_2) inside a rectangle Π defined by inequalities $E_1' \leq E_1 \leq E_1''$ and $E_2' \leq E_2 \leq E_2''$. The intervals (E_1', E_1'') and (E_2', E_2'') are subdivided into m_1 and m_2 subintervals respectively creating a grid with $m_1 \times m_2$ cells inside Π. The distribution F induces the discrete distribution p_{ij} $(1 \leq i \leq m_1, 1 \leq j \leq m_2)$ of the probabilities that an observed value of the random vector belongs to the cells which we call the expected frequencies. Having a sample of size N we obtain the numbers N_{ij} of observations belonging to the cells which we call the observed frequencies. The values

$$N_{i.} = \sum_j N_{ij}\ ; \qquad p_{i.} = \sum_j p_{ij}$$

$$N_{.j} = \sum_i N_{ij}\ ; \qquad p_{.j} = \sum_i p_{ij}$$

are called the marginal frequencies (the observed ones and the expected ones). Let us introduce so-called "through" numeration for the marginal frequencies: $N_k = N_{k.}$, $p_k = p_{k.}$ for $k=1,\dots,m_1$ and $N_{k+m_1} = N_{.k}$, $p_{k+m_1} = p_{.k}$ for $k=1,\dots,m_2$ and denote $m = m_1 + m_2$.

Let us remind some results of the paper /1/ needed for the further discussion. Denote by x the random column-vector

$$x^T = \left(\frac{N_1 - Np_1}{\sqrt{Np_1}}, \frac{N_2 - Np_2}{\sqrt{Np_2}}, \dots, \frac{N_m - Np_m}{\sqrt{Np_m}} \right),$$

by u_1 and u_2 the column-vectors

$$u_1^T = (\sqrt{p_{1.}}, \sqrt{p_{2.}}, \dots, \sqrt{p_{m_1.}}), \quad u_2^T = (\sqrt{p_{.1}}, \sqrt{p_{.2}}, \dots, \sqrt{p_{.m_2}})$$

and by v_1 and v_2 their m-dimensional "extensions"

$$v_1^T = (\sqrt{p_{1.}}, \sqrt{p_{2.}}, \dots, \sqrt{p_{m_1.}}, 0, 0, \dots, 0)$$

$$v_2^T = (0, 0, \dots, 0, \sqrt{p_{.1}}, \sqrt{p_{.2}}, \dots, \sqrt{p_{.m_2}})$$

(here and on T denotes transposition). The symmetric matrix V of order $m \times m$ is defined as follows: it consists of 4 blocks:

$$V = \left| \begin{array}{c|c} V_1 & V_{12} \\ \hline V_{12}^T & V_2 \end{array} \right|,$$

where $V_1 = I_{m_1} - u_1 u_1^T$, $V_2 = I_{m_2} - u_2 u_2^T$ (here and later on I_k is the $k \times k$ unit matrix) and the block V_{12} consists of the elements

$$v_{ij} = \frac{p_{ij} - p_{i.}\, p_{.j}}{\sqrt{p_{i.}\, p_{.j}}}$$

$(1 \leq i \leq m_1, 1 \leq j \leq m_2)$.

As shown in /1/, the vector x is asymptotically normal with the mean 0 and covariance matrix V. The eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ of V are defined in such a way that $\lambda_1 = \lambda_2 = 0$ but generally $\lambda_3 > 0$ (the case $\lambda_3 = 0$ is a degenerate one and corresponds to a sort of determinate dependence between E_1 and E_2).

The above heuristic method of the sum of the marginal χ^2's uses the statistic $T = x^T x$. On the contrary, the statistic T_m proposed in /1/ has the form $T_m = x^T Q x$ where the "weight" symmetric matrix Q is the solution of the matrix equation

$$A^T Q A = I_m - c_{m_1} c_{m_1}^T - c_{m_2} c_{m_2}^T, \tag{1}$$

where c_k is the m-vector consisting of m-1 zeroes and the only one at the k-th position. In its turn, the matrix A is a low-triangular one and satisfies the matrix equation

$$AA^T = V. \tag{2}$$

The equation set (1-2) is always compatible but does not define Q matrix uniquely. Precisely, if we consider Q and V as linear operators in $\mathbb{R}^m$, Q is uniquely defined only in the (m-2)-dimensional vector subspace $L \subset \mathbb{R}^m$ which is orthogonal to the vectors v_1 and v_2 while the vectors Qv_1 and Qv_2 may take arbitrary values. If we restrict the operators Q and V to the subspace L then $Q = V^{-1}$ (L is an eigenspace for both operators). Since $x \in L$ our definition of Q matrix defines the value of $T_m = x^T Q x$ uniquely.

The distribution F and the expected frequencies p_{ij} and p_i depend on parameters $\alpha_1, \alpha_2, \dots, \alpha_s$ $(s < m-1)$. The parameter vector $\alpha = (\alpha_1, \dots, \alpha_s)$ is defined in a domain $I \subset \mathbb{R}^s$. As in /1/ we suppose that

(i) the functions $p_{ij}(\alpha_1, \dots, \alpha_s)$ have the continuous second derivatives;

(ii) the "true" value $\alpha_0 = (\alpha_1^\circ, \dots, \alpha_s^\circ)$ of the parameter vector α is an inner point of I;

(iii) $p_k(\alpha) > c^2$ for some $c > 0$ and for all k, α;

(iv) the matrix $D = (\partial p_k / \partial \alpha_r)$, $1 \leq k \leq m$, $1 \leq r \leq s$ is of rank s for all α;

(v) $\lambda_3 > c$ for all α (λ_3 is the eigenvalue of V, see above).

The latter condition is an extra one if comparing with the classic case of estimating by minimum χ^2 method (/2/, p.p. 426-427).

The last notation from the paper [1] we need is the matrix B of order $m\times s$ consisting of the elements $(1/\sqrt{p_i})\times(\partial p_i/\partial\alpha_r)$, $1\leq i\leq m$, $1\leq r\leq s$. The matrices V,Q,B and the vector x depend on the parameters $\alpha_1,\dots,\alpha_s$ but in the subsequent formulae they are considered only at the point α_o for the "true" value of the parameter vector.

3. <u>The main result</u>. As shown in [1] (see the formula (15)) the estimate $\hat\alpha=(\hat\alpha_1,\dots,\hat\alpha_s)$ of the parameters $\alpha_1,\dots,\alpha_s$ by the minimum T_m method satisfies the relation

$$\hat\alpha-\alpha_o=\frac{1}{\sqrt{N}}(B^TQB)^{-1}B^TQx+\frac{1}{\sqrt{N}}o_p(1)\ , \qquad (3)$$

where $o_p(1)$ denotes a random vector approaching 0 in probability. As noted in [1] the proof of (3) does not use the special form of Q matrix, i.e. it is valid for any symmetric positive definite matrix R smoothly depending on α. Thus, the estimate $\hat\alpha_R$ minimizing $T_R=x^TRx$ satisfies the relation

$$\hat\alpha_R-\alpha_o=\frac{1}{\sqrt{N}}(B^TRB)^{-1}B^TRx+\frac{1}{\sqrt{N}}o_p(1)\ . \qquad (4)$$

Relation (4) implies the following lemma:

<u>Lemma 1</u>. The vector $\sqrt{N}(\hat\alpha_R-\alpha_o)$ is asymptotically normal with the mean 0 and covariance matrix

$$C_R=(B^TRB)^{-1}B^TRVRB(B^TRB)^{-1}\ . \qquad (5)$$

Since $Q=V^{-1}$ in the subspace L we obtain $QV=I_m-v_1v_1^T-v_2v_2^T$ simplifying (5) for $R=Q$:

$$C_Q=(B^TQB)^{-1}\ . \qquad (6)$$

The main result of the present paper is the following theorem:

<u>Theorem 1</u>. The estimator minimizing $T_m=x^TQx$ is asymptotically efficient in the class of estimators minimizing $T_R=x^TRx$ with a symmetric positive definite matrix R smoothly depending on α. In other words the concentration ellipsoid defined by the covariance matrix C_Q lies wholly within the concentration ellipsoid defined by any covariance matrix C_R.

<u>Proof</u>. It is known (see [2], p. 300) that the concentration ellipsoid of an estimate of a s-dimensional parameter with the covariance matrix C is defined by the equation $y^TC^{-1}y=s+2$, where $y=(y_1,y_2,\dots,y_s)$ is the column-vector of independent variables. If C_1 and C_2 are two covariance matrices then the condition of the embedment of the corresponding ellipsoids of concentration (the first one is inside the second one) can be expressed in the form: $C_1^{-1}\geqslant C_2^{-1}$ (i.e. the matrix $C_1^{-1}-C_2^{-1}$ is nonnegative definite).

Therefore the theorem 1 is equivalent to the matrix inequality

$$B^TQB\geqslant B^TRB(B^TRVRB)^{-1}B^TRB \qquad (7)$$

which must be valid for any symmetric positive definite matrix R (note that the nonsingularity of the matrix B^TRVRB needs a special proof, see below).

Let us proof the inequality (7). First we show that the matrix B^TRVRB is really nonsingular. Let $y\in\mathbb{R}^s$, then $(B^TRVRBy,y)=(VRBy,RBy)$ and the latter quantity is equal to 0 only if the vector RBy is orthogonal to L. But $B\mathbb{R}^s\subset L$ (see [1]), so $(RBy,By)=0$ in contradiction to the positive definiteness of R.

The inequality (7) means that for any vector $y\in\mathbb{R}^s$

$$(B^TQBy,y)\geqslant(B^TRB(B^TRVRB)^{-1}B^TRBy,y)\ .$$

It is equivalent to the condition that for any vector $w\in B\mathbb{R}^s$

$$(Qw,w)\geqslant(RB(B^TRVRB)^{-1}B^TRw,w)\ . \qquad (8)$$

We shall prove a more strong condition that (8) holds for any vector $w\in L$. Denote $P=I_m-v_1v_1^T-v_2v_2^T$ the orthogonal projective matrix of the m-dimensional vector space onto the subspace L. Then $Pw=w$, $PB=B$, $B^TP=B^T$ and (8) can be rewritten in the form

$$(Qw,w)\geqslant(PRPB(B^TPRPVPRPB)^{-1}B^TPRPw,w)$$

or

$$(Qw,w)\geqslant(R_1B(B^TR_1VR_1B)^{-1}B^TR_1w,w)\ , \qquad (9)$$

where $R_1=PRP$ is a symmetric matrix with two null eigenvectors v_1 and v_2 and the positive eigenspace L.

The further transformations are performed only in the subspace L where Q,R_1 and V $(Q=V^{-1})$ are symmetric positive definite operators. Therefore (9) is equivalent to

$$(R_1^{-1}QR_1^{-1}w,w)\geqslant(B(B^TR_1VR_1B)^{-1}B^Tw,w)$$

or

$$((R_1VR_1)^{-1}w,w)\geqslant(B(B^TR_1VR_1B)^{-1}B^Tw,w)\ . \qquad (10)$$

The operator $G=(R_1VR_1)^{-1}$ is also symmetric and positive definite in L, thus (10) is equivalent to the following condition:

$$(Gw,w)\geqslant(B(B^TG^{-1}B)^{-1}B^Tw,w) \qquad (11)$$

for any symmetric positive definite operator G in L and for any vector $w\in L$.

The last notation from the paper [1] we need is the matrix B of order $m\times s$ consisting of the elements $(1/\sqrt{p_i})\times(\partial p_i/\partial\alpha_r)$, $1\leq i\leq m$, $1\leq r\leq s$. The matrices V,Q,B and the vector x depend on the parameters $\alpha_1,\dots,\alpha_s$ but in the subsequent formulae they are considered only at the point α_0 for the "true" value of the parameter vector.

3. <u>The main result</u>. As shown in [1] (see the formula (15)) the estimate $\hat{\alpha}=(\hat{\alpha}_1,\dots,\hat{\alpha}_s)$ of the parameters $\alpha_1,\dots,\alpha_s$ by the minimum T_m method satisfies the relation

$$\hat{\alpha}-\alpha_0=\frac{1}{\sqrt{N}}(B^TQB)^{-1}B^TQx+\frac{1}{\sqrt{N}}o_p(1)\ , \qquad (3)$$

where $o_p(1)$ denotes a random vector approaching 0 in probability. As noted in [1] the proof of (3) does not use the special form of Q matrix, i.e. it is valid for any symmetric positive definite matrix R smoothly depending on α. Thus, the estimate $\hat{\alpha}_R$ minimizing $T_R=x^TRx$ satisfies the relation

$$\hat{\alpha}_R-\alpha_0=\frac{1}{\sqrt{N}}(B^TRB)^{-1}B^TRx+\frac{1}{\sqrt{N}}o_p(1)\ . \qquad (4)$$

Relation (4) implies the following lemma:

<u>Lemma 1</u>. The vector $\sqrt{N}(\hat{\alpha}_R-\alpha_0)$ is asymptotically normal with the mean 0 and covariance matrix

$$C_R=(B^TRB)^{-1}B^TRVRB(B^TRB)^{-1}\ . \qquad (5)$$

Since $Q=V^{-1}$ in the subspace L we obtain $QV=I_m-v_1v_1^T-v_2v_2^T$ simplifying (5) for R=Q:

$$C_Q=(B^TQB)^{-1}\ . \qquad (6)$$

The main result of the present paper is the following theorem:

<u>Theorem 1</u>. The estimator minimizing $T_m=x^TQx$ is asymptotically efficient in the class of estimators minimizing $T_R=x^TRx$ with a symmetric positive definite matrix R smoothly depending on α. In other words the concentration ellipsoid defined by the covariance matrix C_Q lies wholly within the concentration ellipsoid defined by any covariance matrix C_R.

<u>Proof</u>. It is known (see [2], p. 300) that the concentration ellipsoid of an estimate of a s-dimensional parameter with the covariance matrix C is defined by the equation $y^TC^{-1}y=s+2$, where $y=(y_1,y_2,\dots,y_s)$ is the column-vector of independent variables. If C_1 and C_2 are two covariance matrices then the condition of the embedment of the corresponding ellipsoids of concentration (the first one is inside

the second one) can be expressed in the form: $C_1^{-1}\geqslant C_2^{-1}$ (i.e. the matrix $C_1^{-1}-C_2^{-1}$ is nonnegative definite).

Therefore the theorem 1 is equivalent to the matrix inequality

$$B^TQB\geqslant B^TRB(B^TRVRB)^{-1}B^TRB \qquad (7)$$

which must be valid for any symmetric positive definite matrix R (note that the nonsingularity of the matrix B^TRVRB needs a special proof, see below).

Let us proof the inequality (7). First we show that the matrix B^TRVRB is really nonsingular. Let $y\in\mathbb{R}^s$, then $(B^TRVRBy,y)=(VRBy,RBy)$ and the latter quantity is equal to 0 only if the vector RBy is orthogonal to L. But $B\mathbb{R}^s\subset L$ (see [1]), so $(RBy,By)=0$ in contradiction to the positive definiteness of R.

The inequality (7) means that for any vector $y\in\mathbb{R}^s$

$$(B^TQBy,y)\geqslant(B^TRB(B^TRVRB)^{-1}B^TRBy,y)\ .$$

It is equivalent to the condition that for any vector $w\in B\mathbb{R}^s$

$$(Qw,w)\geqslant(RB(B^TRVRB)^{-1}B^TRw,w)\ . \qquad (8)$$

We shall prove a more strong condition that (8) holds for any vector $w\in L$. Denote $P=I_m-v_1v_1^T-v_2v_2^T$ the orthogonal projective matrix of the m-dimensional vector space onto the subspace L. Then $Pw=w$, $PB=B$, $B^TP=B^T$ and (8) can be rewritten in the form

$$(Qw,w)\geqslant(PRPB(B^TPRPVPRPB)^{-1}B^TPRPw,w)$$

or

$$(Qw,w)\geqslant(R_1B(B^TR_1VR_1B)^{-1}B^TR_1w,w)\ , \qquad (9)$$

where $R_1=PRP$ is a symmetric matrix with two null eigenvectors v_1 and v_2 and the positive eigenspace L.

The further transformations are performed only in the subspace L where Q,R_1 and V ($Q=V^{-1}$) are symmetric positive definite operators. Therefore (9) is equivalent to

$$(R_1^{-1}QR_1^{-1}w,w)\geqslant(B(B^TR_1VR_1B)^{-1}B^Tw,w)$$

or

$$((R_1VR_1)^{-1}w,w)\geqslant(B(B^TR_1VR_1B)^{-1}B^Tw,w)\ . \qquad (10)$$

The operator $G=(R_1VR_1)^{-1}$ is also symmetric and positive definite in L, thus (10) is equivalent to the following condition:

$$(Gw,w)\geqslant(B(B^TG^{-1}B)^{-1}B^Tw,w) \qquad (11)$$

for any symmetric positive definite operator G in L and for any vector $w\in L$.

Let us prove (11). Consider the symmetric operator $K = G - B(B^T G^{-1} B)^{-1} B^T$ in L. Denote by L_1 and L_2 two linear subspaces in L: $L_1 = G^{-1}B\ {}^s$ and L_2 is the orthogonal complement to $B\mathbb{R}^s$. It is easy to show that the operator K equals 0 in L_1 and equals G in L_2. Thus (11) is valid for $w \in L_1$ and for $w \in L_2$.

Suppose that (11) is not valid everywhere in L, i.e. the operator K has a negative eigenvector: $Kw = \lambda w$, $\lambda < 0$. Evidently, the vector w should be orthogonal to L_1 because L_1 is the subspace of the null eigenspace of the K operator. It is easy to show that $L_1 \cap L_2 = \{0\}$ and $\dim L_1 + \dim L_2 = \dim L$, thus the w vector could be decomposed as $w = w_1 + w_2$, where $w_1 \in L_1$ and $w_2 \in L_2$. Then

$$(Kw_2, w_2) = (K(w-w_1), w_2) = (\lambda w, w_2) = \lambda (w, w-w_1) = \lambda \|w\|^2 < 0 ,$$

but on the other hand $(Kw_2, w_2) = (Gw_2, w_2) \geq 0$. The obtained contradiction proves the inequality (11) and, as a consequence, the theorem.

4. A numerical experiment. For experimentators it is of great interest to do a certain numerical comparison of the estimate minimizing T_m with the one minimizing the sum of the marginal χ^2's conventionally used in practice. To obtain numerical values of their statistical characteristics we have made a numerical experiment using the example cited in /1/: so-called isobar model of pion-nucleon interactions near Δ_{33} production threshold, namely the reaction $\pi^- p \to \pi^- \pi^+ n$. This reaction is the major one at the energies lower 1 Gev. We have taken a simplified version of this model described in /1/. The joint probability density for energies E_1, E_2 of the secondary pions is /3/

$$\frac{\partial^2 F}{\partial E_1 \partial E_2} \cong |aR_1|^2 + |bR_2|^2 + 2\mu'_o \,\mathrm{Re}\,[aR_1 b^* R_2^*] , \qquad (12)$$

where $a = -\sqrt{2/15}\, a_3 - 1/\sqrt{3}\, a_1 e^{i\varphi}$, $b = \sqrt{8/135}\, a_3 - 1/\sqrt{27}\, a_1 e^{i\varphi}$. For other quantities we have used the same conventions as in /1/:

$$R_1 = \left(\frac{\Gamma_1}{2\pi p_1'}\right)^{1/2} \frac{1}{\omega_o - \omega_{13} - \frac{1}{2} i \Gamma_1} , \quad R_2 = \left(\frac{\Gamma_1}{2\pi p_2'}\right)^{1/2} \frac{1}{\omega_o - \omega_{23} - \frac{1}{2} i \Gamma_1} ,$$

Γ_1 is a Δ_{33} width, ω_o is its mass, ω_{13} and ω_{23} are the effective masses of $\pi^- n$ and $\pi^+ n$ systems at the final state respectively, p_1', p_2' are the Δ_{33} momenta in the center-of-mass reaction.

In expression (12) the parameters a_1, a_3, φ are unknown and are to be estimated. But any one of them can be calculated from two others via the normalization condition

$$\iint \frac{\partial^2 F}{\partial E_1 \partial E_2}\, dE_1 dE_2 = 1 .$$

In fact there are only two independent parameters, let them be a_1 and a_3. We have chosen the "true" values of parameters $\overset{o}{a}_1 = 8$, $\overset{o}{a}_3 = 1$, $\overset{o}{\varphi} \approx 4.23$.

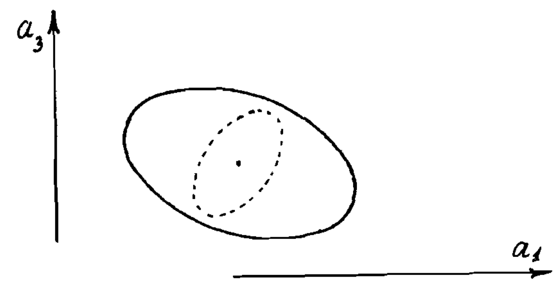
The accuracy of a joint estimator of the parameters a_1, a_3 is described by its ellipse of concentration centered at the point $(\overset{o}{a}_1, \overset{o}{a}_3)$. The equations of the concentration ellipses for our two estimators have been calculated by a computer using the formulae obtained in Sect.3. The figure represents the ellipse of concentration of the estimate minimizing the sum of the marginal χ^2's (the solid curve) and that of the estimate minimizing T_m (the dotted curve). The equations of these ellipses are

$$4.8\,(a_1 - a_1^o)^2 + 7.2\,(a_3 - a_3^o)^2 + 1.9\,(a_1 - a_1^o)(a_3 - a_3^o) = 4000$$

and

$$17.3\,(a_1 - a_1^o)^2 + 13.5\,(a_3 - a_3^o)^2 - 9.5\,(a_1 - a_1^o)(a_3 - a_3^o) = 4000$$

respectively. It is clearly seen that the proposed estimate is at least twice more accurate than the previously used one.



REFERENCES:

1. Chernov N.I., Ososkov G.A., Kurbatov V.S., JINR, E5-87-698, Dubna, 1987.
2. Cramer H. Mathematical methods of statistics. Princeton Univ. Press, Princeton, 1946.
3. Olsson M.G., Yodh G.B. Phys. Rev., 1966, 145, p. 1309.

Received by Publishing Department
on February 5, 1988.