Mathematics

O. V. SHALAEVSKII

1960

SovietRxiv

Abstract

Full Text

Mathematics

O. V. SHALAEVSKII

SOME REMARKS ON THE OBSERVATION EQUATION WITH UNKNOWN WEIGHTS

(Presented by Academician A. N. Kolmogorov, July 3, 1959)

In this note several steps are taken toward the confidence estimation of the values of given linear functions of unknown, but quite definite, parameters $\xi_1, \dots, \xi_m$, connected with the measured quantity λ by the relation $\lambda = a_0 + a_1\xi_1 + \dots + a_m\xi_m$ (where $a_0, a_1, \dots, a_m$ are prescribed a priori constants), under the classical assumptions of the scheme of equations by elements; however, it is not required that the accuracies of the measurements, or their ratios, be known quantities. The exposition is given in matrix form, now generally accepted in analogous mathematical works (see, for example, ([1])).

Consider r groups of values of λ:

$$\Lambda^{(\alpha)}_{n_\alpha 1} = \|\lambda^{(\alpha)}_i\|, \quad n_\alpha \geq m, \quad \alpha = 1, \dots, r,$$

fixing for group α the matrix

$$A^{(\alpha 0)} = A^{(\alpha 0)}_{n_\alpha 1} = \|a^{(\alpha)}_{i0}\|$$

and the matrix

$$A^{(\alpha)} = A^{(\alpha)}_{n_\alpha m} = \|a^{(\alpha)}_{ij}\|$$

of rank m. Put

$$\Lambda^{(\alpha)} = A^{(\alpha 0)} + A^{(\alpha)}\Xi, \quad \Lambda = A^{(0)} + A\Xi,$$

where

$$\Lambda = \Lambda_{r1} = \|\Lambda^{(\alpha)}\|; \quad A^{(0)} = A^{(0)}_{r1} = \|A^{(\alpha 0)}\|, \quad A = A_{r1} = \|A^{(\alpha)}\|, \quad \Xi = \Xi_{m1} = \|\xi_j\|.$$

The vector Λ is subjected to measurement.

Suppose that this measurement gives the vector of observations L, whose components may be regarded as mutually independent; $l^{(\alpha)}_i$ is the result of measuring $\lambda^{(\alpha)}_i$, $i = 1, \dots, n_\alpha$, $\alpha = 1, \dots, r$, and has a normal distribution with mean $\lambda^{(\alpha)}_i$ and standard deviation σ_α. The vector of observations for $\Lambda^{(\alpha)}$ will be denoted by

$$L^{(\alpha)} = L^{(\alpha)}_{n_\alpha 1} = \|l^{(\alpha)}_i\|; \quad L = L_{r1} = \|L^{(\alpha)}\|.$$

Introduce the matrix

$$G = G_{km} = \|g_{ij}\|$$

of rank $k \le m$, and construct the vector

$$H = G\Xi$$

–the desired linear functions of the elements.

To carry out confidence estimation of the vector H in the described situation with unknown weights of the observations, it is convenient to combine the approach, due to A. Wald ([1]), to the solution of the Behrens–Fisher problem* with the device, noted by Yu. V. Linnik ([3]), for constructing confidence ellipsoids.

It is not difficult to form an expression equivalent to the initial statistic in the Behrens–Fisher problem. Namely, applying the corresponding prescription of least squares, we process the observations as if they were all of equal accuracy; this leads to the point estimate for Ξ, and substituting it in the formula for H, we obtain

$$\bar{H} = \|\eta_i\| = G(A^TA)^{-1}A^T(L - A^{(0)})$$

(as in ([1]), the assignment of the superscript T turns the given matrix into the transposed one); let

$$[vv]_\alpha = \min_{\Xi}(L^{(\alpha)} - A^{(\alpha 0)} - A^{(\alpha)}\Xi)^T(L^{(\alpha)} - A^{(\alpha 0)} - A^{(\alpha)}\Xi), \quad s_\alpha^2 = \frac{[vv]_\alpha}{n_\alpha - m},$$

$$C_\alpha = G(A^TA)^{-1}(A^{(\alpha)})^TA^{(\alpha)}(A^TA)^{-1}G^T, \quad M = C_1s_1^2u_1 + \dots + C_{rs}r^2u_r,$$

---

* This is the so-called problem of testing the hypothesis of equality of the means of two normal populations when the ratio of their variances is unknown ([4–6]).

$u_1, \dots, u_r$ are real variables; moreover, looking a little ahead, set

$$N = D_1u_1 + \dots + D_{ru}r, \qquad D_\alpha = C_\alpha\sigma_\alpha^2, \qquad T_{\alpha\beta}^{st} = \left(\frac{\partial^{s+t}|M|/\partial u_\alpha^s\partial u_\beta^t}{|M|}\right)_0;$$

the subscript zero everywhere below means that the corresponding functions are evaluated at the point $u_1 = u_2 = \dots = u_r = 1$; then the required expression will be $(\bar{H} - \bar{\bar{H}})^TM_0^{-1}(\bar{H} - \bar{\bar{H}})$.

In the limit, as $\min_\alpha n_\alpha = n \to \infty$, the quantity found has a χ^2-distribution with k degrees of freedom–the rate of convergence is of order $1/n$. But this limiting assertion can be developed in the following way.

Theorem 1. For every natural q there exists a function $V_q(c, s_1^2, \dots, s_r^2)$, $c \ge 0$, such that

$$P\left[(\bar{H}-\bar{\bar{H}})^T M_0^{-1}(\bar{H}-\bar{\bar{H}}) \leq V_q(c, s_1^2, \dots, s_r^2)\right] = P(\chi_k^2 \leq c) + R_q,$$

where χ_k^2 is a random variable having the χ^2-distribution with k degrees of freedom, $|R_q| < C/n^q$, and the constant $C = C(r, m)$.

The proof of Theorem 1 is based on the joint independence of the quantities $\bar{\eta}_1, \dots, \bar{\eta}_k, s_1^2, \dots, s_r^2$ and on the possibility of generalizing and refining the approach of A. Wald mentioned above, as well as on the possibility of generalizing the expansions of Wallace ([7]). The proof is accompanied by a method for successively determining the functions V_q in the form of finite series in powers of $\frac{1}{n_\alpha - m}$, $\alpha = 1, \dots, r$. However, as q increases, their computation very quickly becomes complicated.

It should be especially emphasized that the constant C is absolute with respect to any data of the problem except r and m, in particular with respect to the accuracies $\sigma_1^2, \dots, \sigma_r^2$. We indicate considerations by means of which this fact can be derived.

Let $N^{-1}D_\alpha = Q_\alpha$, $\alpha = 1, \dots, r$. Let us single out the class $\mathfrak{R}$ of matrices admitting the representation $Q_{\alpha_1} \cdots Q_{\alpha_s} N_0^{-1} Q_{\beta_1}^T \cdots Q_{\beta_t}^T$, where $\alpha_1, \dots, \alpha_s;\ \beta_1, \dots, \beta_t$ are some indices from the set $(1, \dots, r)$.

Lemma 1. The derivative

$$\frac{\partial^h}{\partial u_1^{h_1} \cdots \partial u_r^{h_r}} \left\{ \frac{1}{|N|^{1/2}} \int_{Y^T N^{-1} Y \leq c} \Phi(Y) e^{-\frac{1}{2} Y^T N_0^{-1} Y}\, dy_1 \cdots dy_k \right\}, \qquad h \geq 0,$$

is a finite sum composed of expressions of the form

$$\frac{1}{|N|^{1/2}} \int_{Y^T N^{-1} Y \leq c} \Psi(Y) e^{-\frac{1}{2} Y^T N_0^{-1} Y}\, dy_1 \cdots dy_k,$$

multiplied by certain constant numbers.

$\Phi(Y)$ and $\Psi(Y)$ are either products of a finite number of quadratic forms with matrices from the class $\mathfrak{R}$, or 1.

Take l points $(u_{i1}, \dots, u_{ir})$ such that $|u_{i\alpha} - 1| < \delta < 1$; $i = 1, \dots, l$; $\alpha = 1, \dots, r$; form l matrices $N_i = \sum_{\alpha=1}^{r} D_\alpha u_{i\alpha}$ and consider the form

$$K(x, x) = X^T N_0 N_l^{-1} D_{\alpha_l} N_{l-1}^{-1} D_{\alpha_{l-1}} \cdots N_2^{-1} D_{\alpha_2} N_1^{-1} N_0 X;$$

$\alpha_2, \dots, \alpha_l$ are some indices from the set $(1, \dots, r)$.

Lemma 2. The estimate $|K(x,x)| < a \cdot X^T N_0 X$ holds, with the constant $a = a(l,m,\delta)$.

Moreover, if the variables $s_1^2, \dots, s_r^2$ of the coefficient functions at the powers $\dfrac{1}{n_\alpha - m}$ for V_q are replaced by $\sigma_\alpha^2 u_\alpha$, then the new functions of $u_1, \dots, u_r$ are everywhere bounded by an absolute constant (for given r and m) multiplied by some power of the number c. Their partial derivatives and the partial derivatives of $|N|/|N_0|$ are likewise bounded in the neighborhood $|u_\alpha - 1| < \delta,\ \alpha = 1, \dots, r$.

Theorem 1 proves useful for "approximate" confidence estimation of the vector $\overline{\mathrm{H}}$.

To this end, note that the event

$$(\overline{\mathrm{H}} - \overline{\overline{\mathrm{H}}})^T M_0^{-1} (\overline{\mathrm{H}} - \overline{\overline{\mathrm{H}}}) \leqslant V_q$$

is equivalent to covering by the ellipsoid

$$(Z - \overline{\overline{\mathrm{H}}})^T M_0^{-1} (Z - \overline{\overline{\mathrm{H}}}) = V_q,$$

where $Z = Z_{k1} = \|z_i\|$ is the vector of current coordinates of the unknown point $Z = \overline{\mathrm{H}}$. We state two theorems that may have practical application.

Theorem 2. Let

$$\mu = \sum_{\alpha=1}^{r} \frac{\rho_\alpha}{n_\alpha - m} + \sum_{\alpha,\beta=1}^{r} \frac{\tau_{\alpha\beta}}{(n_\alpha - m)(n_\beta - m)}.$$

The ellipsoid

$$(Z - \overline{\mathrm{H}})^T M_0^{-1} (Z - \overline{\mathrm{H}}) = c(1 + \mu + \mu^2)$$

covers the point $Z = \overline{\mathrm{H}}$ with probability

$$1 - \varepsilon = P(\chi_{2k} \leqslant c) + R_3$$

and

$$|R_3| < \frac{C(r,m)}{n^3}.$$

In Theorem 2, ε is the confidence level and

$$\rho_\alpha = \frac{b_1}{2} T_\alpha^2 - b^2 T_\alpha;$$

$$\begin{aligned}\tau_{\alpha\beta} = \frac{c-2-k}{4}\rho_\alpha\rho_\beta - \frac{1}{2}(b_1+b_2+b_2b_3+2b_{11}-\delta_{\alpha\beta}b_{11})(T_\alpha T_{\alpha\beta^2}+T_\beta T_{\alpha^2\beta}) \\ + (3b_1+b_1b_3-b_{10}+\tfrac{3}{4}\delta_{\alpha\beta}b_{10})T_\alpha T_\beta T_{\alpha\beta} + (b_2-b_{13}+\tfrac{1}{2}\delta_{\alpha\beta}b_{13})T_{\alpha^2\beta^2} \\ - (b_1+b_{12})T^2_{\alpha\beta} + \frac{1}{8}(3b_1+3b_2+b_2b_4+4b_2b_3 \\ + b_1b_5 - 2b_{10})(T^2_\alpha T_{\beta^2} + T^2_\beta T_{\alpha^2}) \\ - \frac{1}{8}\left(15b_1+8b_1b_3+b_1b_4+b_9-\frac{1}{2}\delta_{\alpha\beta}b_9\right)T^2_\alpha T^2_\beta \\ - \frac{1}{2}\left(b_2+b_2b_5+b_{12}-\frac{3}{2}\delta_{\alpha\beta}b_{12}\right)T_{\alpha^2}T_{\beta^2} \\ + \delta_{\alpha\beta}\left\{(3b_1+b_1b_3-b_6)T^3_\alpha - (2b_2b_3+4b_1+2b_2+b_7)T_\alpha T_{\alpha^2}\right. \\ \left. +(4b_2-\tfrac{1}{3}b_8)T_{\alpha^3} - 2\rho_\alpha\right\}.\end{aligned}$$

There is no compact formula for determining the 30 components of the column vector $B = \|b_i\|$, which is explained by the unwieldiness of the expansions used in the proof. In any case,

$$B = \|f_1, f_2, f_3, f_4\| \cdot \Gamma,$$

where

$$f_i\, k(k+2)\dots(k+2i-2) = c^{\,i-1},$$

$$\Gamma = \left\|\begin{matrix} 1 & 1 & c & -2c & 0 & 1 & -2 & 4 & -15 & 3 & -1 & -1 & 1 \\ 3 & 1 & 0 & 3c & c & 2 & -4 & 4 & -27 & 5 & -2 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 5 & -6 & 4 & -45 & 9 & -2 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -105 & 15 & -3 & -3 & 1 \end{matrix}\right\|.$$

Theorem 3. The ellipsoid

$$(Z-\overline{\mathrm{H}})^T M_0^{-1}(Z-\overline{\mathrm{H}}) = c\left\{1 + \sum_{\alpha=1}^{r}\frac{\rho_\alpha}{n_\alpha - m} + \left[\sum_{\alpha=1}^{r}\frac{\rho_\alpha}{n_\alpha - m}\right]^2\right\}$$

covers the point $Z = \mathrm{H}$ with probability

$$1-\varepsilon = P(\chi^2_k \le c) + R_2 \quad \text{and} \quad |R_2| < \frac{C(r,m)}{n^2}.$$

Confidence ellipsoids constructed by Yu. V. Linnik in the case of known observation weights ([3]) are based on Fisher’ s F-distribution. All our theorems can also be based on this distribution. But in passing to Fisher’ s F-distribution the amount of computation is not reduced; the question of comparing the approximation by the χ^2-distribution with the approximation by the F-distribution remains unclear. Therefore we shall illustrate the possibility of applying the F-distribution only in a special case.

Theorem 4. Let $n_1 = n_2 = \dots = n_r = n$ and

$$\rho = \sum_{\alpha=1}^{r} \rho_\alpha - \frac{c+2-k}{2} = b_2 \sum_{\alpha \neq \beta} T_{\alpha\beta} - \frac{b_1}{2} \sum_{\alpha \neq \beta} T_\alpha T_\beta.$$

The ellipsoid

$$(Z - \overline{\mathrm{H}})^T M_0^{-1} (Z - \overline{\mathrm{H}}) = c \left[1 + \frac{\rho}{n-m} + \frac{\rho^2}{(n-m)^2} \right]$$

covers the point $Z = \mathrm{H}$ with probability

$$1 - \varepsilon = F_{k,n-m} \left(\frac{c}{n-m} \right) + R_2, \qquad |R_2| < \frac{C(r,m)}{n^2}.$$

Let us make some general remarks.

The proof of Theorem 1 requires only that the components of the given random vector Ξ ($E\Xi = \mathrm{H}$) be in normal correlation with a matrix of second moments of the form $F_1\sigma_1^2 + \dots + F_r\sigma_r^2$ ($F_1, \dots, F_r$ are positive definite), where the parameters σ_α^2 admit estimates s_α^2, distributed as $\sigma_\alpha^2 \chi_{k_\alpha}^{2(\alpha)} / k_\alpha$ (in our case the role of k_α is played by $n_\alpha - m$), the quantities $\xi_1, \dots, \xi_m, s_1^2, \dots, s_r^2$ being jointly independent.

If one introduces $\mathrm{H} = G\Xi$ and $\overline{\mathrm{H}} = G\overline{\Xi}$ ($C_\alpha = GF_\alpha G^T$), then the formulated theorems remain valid under these more general assumptions.

Theorem 1 can be used to test any simple hypothesis concerning H, in particular the hypothesis that all (or some) of the elements $\xi_1, \dots, \xi_m$ are equal to a given number. Thus, what has been presented above may be regarded as a generalization of the Behrens–Fisher problem and an approximate solution of that problem.

Finally, it is natural to pose the problem of finding a function $V(c, s_1^2, \dots, s_r^2)$ satisfying the condition

$$P\left[(\mathrm{H} - \overline{\mathrm{H}})^T M_0^{-1} (\mathrm{H} - \overline{\mathrm{H}}) \leq V(c, s_1^2, \dots, s_r^2) \right] = P(\chi_k^2 \leq c).$$

The question of the existence of such a function is open. Even the existence of an "exact" solution of the classical Behrens–Fisher problem has not been proved (((7) gives a formal solution, (2) adds little).

I express my gratitude to A. N. Kolmogorov for a number of suggestions and comments concerning the present note and further work on the subject touched upon here.

Leningrad State University
named after A. A. Zhdanov

Received
3 VII 1959

REFERENCES

1. Yu. V. Linnik, *The Method of Least Squares and the Foundations of the Mathematical-Statistical Theory of the Treatment of Observations*, 1958.

2. A. Wald, *Selected Papers in Statistics and Probability*, N. Y., 1955.

3. Yu. V. Linnik, *Theory of Probability and Its Applications*, 2, no. 3 (1957).

4. W. U. Behrens, *Landwirtsch. Jahrb.*, 68, 807 (1929).

5. R. A. Fisher, *Ann. Eugenics*, 6, 391 (1935); 9, 174 (1939); 11, 141 (1941).

6. R. A. Fisher, *Ann. Math. Statistics*, 10, 383 (1940).

7. B. L. Welch, *Biometrica*, 34, 28 (1947).

Note: Figure translations are in progress. See original paper for figures.