Osaka Mathematical Journal,
Vol. 2, No. 2, November, 1950

Note on the Markoff's Theorem on Least Squares [1]

By Junjiro OGAWA

The purpose of this note is to give a simple proof of the extension of the famous Markoff's theorem on least squares by J. Neyman [2] and F. N. David, which is very useful especially in the theory of sampling [3].

Theorem. Let n random variables $x_1, x_2, \dots, x_n$

(a) be independently distributed [4], and

(b) their means be linearly restricted with $s(\leq n)$ unknown parameters $p_1, p_2, \dots, p_s$ with known coefficients, i. e.

$$E(x_i)=a_{i1}p_1+a_{i2}p_2+ \dots +a_{is}p_s,\ i=1, 2, \dots, n, \tag{1}$$

where the coefficients $a_{ij}, i=1, 2, \dots, n;\ j=1, 2, \dots, s$ are known constants.

(c) The rank of the coefficient matrix

$$A=\begin{pmatrix} a_{11}a_{12} \cdots\cdots a_{1s} \\ a_{21}a_{22} \cdots\cdots a_{2s} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}a_{n2} \cdots\cdots a_{ns} \end{pmatrix} \tag{2}$$

is equal to s.

(d) Further, let the variance σ_i^2 of x_i be

$$\sigma_i^2=\frac{\sigma^2}{P_i}, \quad i=1, 2, \dots, n, \tag{3}$$

where $P_1, P_2, \dots, P_n$ are known constants and σ unknown.

If the above conditions are satisfied, then the follnwing two statements (α) and (β) hold.

(α) The best unbiased linear estimate [5] of the linear form

$$\theta=b_1p_1+b_2p_2+ \dots +b_sp_s \tag{4}$$

with known coefficients $b_1, b_2, \dots, b_s$ is

$$F=b_1p_1^0+b_2p_2^0+ \dots +b_sp_s^0, \tag{5}$$

where $p_1^0, p_2^0, \dots, p_s^0$ are the system of values of $p_1, p_2, \dots, p_s$, for which the weighted square sum

$$S=\sum_{i=1}^{n}(x_i-a_{i1}p_1-a_{i2}p_2 \dots -a_{is}p_s)^2 P_i, \tag{6}$$

is minimum for a given system of values of $x_1, x_2, \dots, x_n$. And further, (β), the unbiased estimate of the variance of F is

$$\mu^2=\frac{S_0}{n-s}\sum_{i=1}^{n}\mu_i^2/P_i, \tag{7}$$

where

$$S_0=\sum_{i=1}^{n}(x_i-a_{i1}p_1^0-a_{i2}p_2^0 \dots -a_{is}p_s^0)^2 P_i, \tag{8}$$

and

$$b_1p_1^0+b_2p_2^0+ \dots +b_sp_s^0=\sum_{i=1}^{n}\mu_i x_i. \tag{9}$$

Remark: From (3), (9), and the condition (a), the variance of F is

$$\sigma_F{}^2=\sigma^2\sum_{i=1}^{n}\mu_i^2/P_i,$$

so, to prove the statement (β), it suffices only to show that

$$E(S_0)=(n-s)\sigma^2 \quad ^{6)} \tag{10}$$

Proof: First, we shall prove (α), that is, the best unbiased linear estimate

$$F^*=d_1x_1+d_2x_2+ \dots +d_nx_n, \tag{11}$$

coincides with F given by (5). This part of the proof is nothing but rewriting of those by J. Neyman and F. N. David in vector notations, so there is nothing new. But only for the sake of completeness of the proof, we shall describe its outlines.

Now consider the following vectors of an n-dimensional Euclidean space R_n referring to a certain orthogonal coordinates system:

$$\mathfrak{d}=\left(\frac{d_1}{\sqrt{P_1}}, \frac{d_2}{\sqrt{P_2}}, \dots, \frac{d_n}{\sqrt{P_n}}\right),$$

and

$$\mathfrak{x}=(\sqrt{P_1}\,x_1, \sqrt{P_2}\,x_2, \dots, \sqrt{P_n}\,x_n),$$

then clearly

$$F^* = \mathfrak{d}\mathfrak{x}', \tag{12}$$

where the prime means the transposed vector.

Further, let

$$P=\begin{pmatrix} \sqrt{P_1} & & & 0 \\ & \sqrt{P_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{P_n} \end{pmatrix} \quad \text{and} \quad B=PA,$$

then the condition of unbiasedness for F^* is written in the form

$$\mathfrak{d}B = \mathfrak{b}, \tag{13}$$

where

$$\mathfrak{b}=(b_1, b_2, \dots, b_s).$$

The variance $\sigma_{F^*}^2$ of F^* being

$$\sigma_{F^*}^2=\sigma^2 \, \| \mathfrak{d} \|^2,$$

where $\| \mathfrak{d} \|$ denotes the absolute value of the vector $\mathfrak{d}$, so that vector $\mathfrak{d}$ is to be determined so as to minimize $\| \mathfrak{d} \|^2$ under the condition (13). If we write an undeterminate vector (the so-called "Lagrange's multiplier") by

$$\mathfrak{l}=(\lambda_1, \lambda_2, \dots, \lambda_s),$$

then the vector $\mathfrak{d}^0$ to be determined is the solution of the system of linear equations

$$\mathfrak{d} = \mathfrak{l}B'$$

and (13). Therefore, we have

$$\mathfrak{d}^0=\mathfrak{b}(B'B)^{-1}B'\ {}^{7)}. \tag{14}$$

Consequently, from the equation (11), F^* may be written in the form

$$F^*=\mathfrak{d}\mathfrak{x}'=\mathfrak{b}(B'B)^{-1}B'\mathfrak{x}'. \tag{15}$$

Compairing (15) with (9), we should have

$$\mathfrak{p}^0=(p_1^0, p_2^0, \dots, p_s^0)=\mathfrak{x}B(B'B)^{-1}\ {}^{8)}$$

and it is easily seen that $\mathfrak{p}^0$ gives the minimum value of S, i. e. S_0.

Second, we shall prove (10):

$$\begin{aligned} S_0&= \| \mathfrak{x}(E-B(B'B)^{-1}B') \|^2 \\ &=\mathfrak{x}(E-B(B'B)^{-1}B')^2\mathfrak{x}' \\ &=\mathfrak{x}(E-B(B'B)^{-1}B')\mathfrak{x}', \end{aligned}$$

because

$$B(B'B)^{-1}B'\cdot B(B'B)^{-1}B'=B(B'B)^{-1}B'.$$

Let

$$\mathfrak{p}=(p_1, p_2, \dots, p_s),$$

then

$$S_0=(\mathfrak{x}-\mathfrak{p}B')(E-B(B'B)^{-1}B')(\mathfrak{x}-\mathfrak{p}B')', \tag{14}$$

because

$$\begin{aligned}&(\mathfrak{x}-\mathfrak{p}B')(E-B(B'B)^{-1}B')(\mathfrak{x}'-B\mathfrak{p}')\\ &=\mathfrak{x}(E-B(B'B)^{-1}B')\mathfrak{x}'-\mathfrak{x}(E-B(B'B)^{-1}B')B\mathfrak{p}'\\ &-\mathfrak{p}B'(E-B(B'B)^{-1}B')\mathfrak{x}'+\mathfrak{p}B'(E-B(B'B)^{-1}B')B\mathfrak{p}',\end{aligned}$$

and

$$\mathfrak{x}(E-B(B'B)^{-1}B')B\mathfrak{p}'=\mathfrak{p}B'(E-B(B'B)^{-1}B')\mathfrak{x}'=0,$$
$$B'B(B'B)^{-1}B'B=B'B.$$

As is easily seen, the rank of the matrix $B(B'B)^{-1}B'$ is s and the matrix was idempotent, so by an appropriate orthogonal transformation of the variates vector

$$\mathfrak{x}-\mathfrak{p}B' = \mathfrak{z}Q,$$

where Q is an orthogonal matrix and $\mathfrak{z}=(z_1, z_2, \dots, z_n)$, S_0 is transformed into the following:

$$S_0=z_1^2+ \dots +z_{n-s}^2, \tag{15}$$

and from the orthogonality of Q, we have

$$E(z_i)=0,\ D^2(z_i)=\sigma^2,\ i=1, 2, \dots, n,$$

hence we have

$$E(S_0)=(n-s)\sigma^2,$$

as was to be proved.

Another proof of the equation (10) from the point of view of geometrical considerations: If we write

$$y_i=\sqrt{P_i}(x_i-a_{i1}p_1-a_{i2}p_2 \dots - a_{is}p_s),\ i=1, 2, \dots, n,$$

then, it is easily seen that

$$E(y_i)=0 \text{ and } E(y_iy_j)=\sigma^2\delta_{ij},\ i, j=1, 2, \dots, n, \tag{16}$$

where δ_{ij} denote the Kronecker's delta.

Further, putting

$$y_i^0=\sqrt{P_i}(x_i-a_{i1}p_1^0-a_{i2}p_2^0 \dots -a_{is}p_s^0),\ i=1, 2, \dots , n,$$

we consider the following $s+2$ vectors referring to a certain orthogonal coordinates system of R_n, as drawn from the origin:

$$\mathfrak{a}_j=(\sqrt{P_1}\, a_{1j}, \sqrt{P_2}\, a_{2j}, \dots , \sqrt{P_n}\, a_{nj}),\ j=1, 2, \dots , s,$$

$$\mathfrak{y}=(y_1, y_2, \dots , y_n),$$

and

$$\mathfrak{y}^0=(y_1^0, y_2^0, \dots , y_n^0).$$

Then the determination of the vector $\mathfrak{p}^0$ so as to minimize S means that

$$\mathfrak{a}_1\mathfrak{y}^{0\prime}=\mathfrak{a}_2\mathfrak{y}^{0\prime}= \dots =\mathfrak{a}_s\mathfrak{y}^{0\prime}=0 \tag{17}$$

simultaneously.

From the condition (c) of the theorem, s vectors $\mathfrak{a}_1, \mathfrak{a}_2, \dots , \mathfrak{a}_s$ are linearly independent, so they generate an s-dimensional subspace R_s of R_n, because of (17), the vector $\mathfrak{y}^0$ lies in the subspace R_{n-s} of R_n perpendicular to R_s. Therefore, we can take a new orthogonal coordinates system with the same origin in R_n, of which the first $n-s$ axes are readily in R_{n-s} and the remaining s axes are in R_s. This fact means that, if we take an appropriate orthogonal matrx C, and put

$$\mathfrak{y} = \mathfrak{z}C, \tag{18}$$

where

$$\mathfrak{z}=(z_1, z_2, \dots , z_n) \text{ and } C=(c_{ij}),$$

then it follows that

$$\mathfrak{y}^0=(z_1, z_2, \dots , z_{n-s}, \overbrace{0 \dots\dots 0}^{s})C. \tag{19}$$

From (18), getting

$$z_i=\sum_{j=1}^{n} c_{ij}y_j,\ i=1, 2, \dots , n.$$

and because of the orthognoality of C and (16), we have

$$E(z_i)=0,\ E(z_iz_j)=\sigma^2\delta_{ij},\ i, j=1, 2, \dots , n. \tag{20}$$

From (8) and (19), we have

$$S_0= \| \mathfrak{y}^0 \|^2=z_1^2+ \dots +z_{n-s}^2,$$

whence it is easily seen from (20),

$$E(S_0) = (n-s)\sigma^2,$$

as was to be proved.

(Received June 24, 1950)

References and Notes

1) The outline of the proof was published in Japanese in Kokyuroku of Inst. Statist. Math. Vol. 5, No. 1, 1949.

2) David, F. N. and Neyman, J., Extension of the Markoff Theorem on Least Squares, Statist. Res. Mem. Vol. II, 1937.

3) See for example, Neyman, J., On the Different Two Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Journ. Roy. Statist. Soc. Vol. KCVII, 1934.

4) This condition is superfluous, it readily suffices to assume that $x_1, x_2, \dots, x_n$ are mutually uncorrelated. See M. Masuyama, Note on a Markoff's Theorem on Least Squares. Kokyuroku. of the Inst. Statist. Math. Vol. 4, No. 11, 1948.

5) J. Neyman and F. N. David named the following estimate of θ the best unbiased linear:

(i) $F = d_1x_1 + d_2x_2 + \dots + d_nx_n$,

(ii) $E(F) = \theta$, and

(iii) the variance of F is minimum among variances of all the estimates satisfying the conditions (i) and (ii).

6) So, $S_0/(n-s)$ is an unbiased quadratic estimate of σ^2. P. L. Hsu investigated on the best unbiased quadratic estimate of σ^2. See, Hsu, P. L: On the Best Unbiased Quadratic Estimate of the Variance, Statist. Res. Mem. Vol. II, 1937.

7) The rank of the matrx B is clearly equal to s, so the matrix $B'B$ is positive definite, for

$$\sum_{i,j}^{s} B'B\zeta_i\zeta_j = \sum_{i,j=1}^{s} \sum_{\alpha=1}^{n} P_\alpha a_{\alpha i} a_{\alpha j} \zeta_i \zeta_j = \sum_{i=1}^{n} (\sqrt{P_i} \sum^{s} a_{\alpha i}\zeta_i)^2 \geqq 0.$$

Therefore the inverse $(B'B)^{-1}$ exists.

8) This is essentially the same as Hsu's Lemma 1 of his paper quoted in (6).