EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 21. KÕIDE
FÜÜSIKA * MATEMAATIKA. 1972, NR. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 21
ФИЗИКА * МАТЕМАТИКА. 1972, № 1

https://doi.org/10.3176/phys.math.1972.1.08

УДК 519.281

Я. КУКС, В. ОЛЬМАН

МИНИМАКСНАЯ ЛИНЕЙНАЯ ОЦЕНКА КОЭФФИЦИЕНТОВ РЕГРЕССИИ

1. Рассмотрим проблему оценивания неизвестного m-мерного вектор-параметра θ в модели

$$E\{\boldsymbol{Y}\mid\theta\}=\boldsymbol{X}\theta, \qquad \operatorname{Cov}\{\boldsymbol{Y}\mid\theta\}=\sigma^2\boldsymbol{I}_n, \tag{1}$$

где $E\{\boldsymbol{Y}|\theta\}$ и $\operatorname{Cov}\{\boldsymbol{Y}|\theta\}$ обозначают соответственно математическое ожидание и ковариационную матрицу случайного n-мерного вектора наблюдений $\boldsymbol{Y}$ при фиксированном значении параметра θ; $\boldsymbol{X}$ — известная матрица порядка $n\times m$; $\boldsymbol{I}_n$ — единичная матрица порядка $n\times n$ и σ^2 — положительный коэффициент. В последующем, если не оговорено противное, будем считать коэффициент σ^2 известным. Пусть $\hat\theta$ — статистика, оценивающая параметр θ. В качестве критерия оценки широко используется функция риска

$$R(\hat\theta,\theta,\boldsymbol{P})=E\{(\hat\theta-\theta)^{\mathrm{T}}\boldsymbol{P}(\hat\theta-\theta)\mid\theta\}, \tag{2}$$

где $\boldsymbol{P}$ — симметричная неотрицательно определенная (н. о. о.) матрица порядка $m\times m$. В случае оценивания функции $\theta^{\mathrm{T}}\boldsymbol{p}$ статистикой $\widehat{\theta^{\mathrm{T}}\boldsymbol{p}}$, где $\boldsymbol{p}$ — ненулевой m-мерный вектор, естественно пользоваться функцией риска

$$R_1(\theta,\theta,\boldsymbol{p})=E\{[(\hat\theta-\theta)^{\mathrm{T}}\boldsymbol{p}]^2\mid\theta\}, \tag{3}$$

которая является частным видом функции (2) при $\boldsymbol{P}=\boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}$. При наличии дополнительной информации о том, что истинный параметр θ принадлежит некоторому подмножеству Ω m-мерного евклидова пространства, рассматривается критерий оценки [1]

$$Q(\hat\theta,\Omega,\boldsymbol{P})=\sup_{\theta\in\Omega}R(\hat\theta,\theta,\boldsymbol{P}). \tag{4}$$

Если $\hat\theta$ — линейная по наблюдениям $\boldsymbol{Y}$ оценка, т. е. $\hat\theta=\boldsymbol{T}\boldsymbol{Y}$, где $\boldsymbol{T}$ — матрица порядка $m\times n$, то функцию риска (3) можно записать в виде $r(\boldsymbol{T},\theta,\boldsymbol{p})=R_1(\boldsymbol{T}\boldsymbol{Y},\theta,\boldsymbol{p})$. Пусть Ω является эллипсоидом с центром в нуле, тогда частным видом критерия (4) является

$$q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})=\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}r(\boldsymbol{T},\theta,\boldsymbol{p}), \tag{5}$$

где $\boldsymbol{A}$ — симметричная н. о. о. матрица порядка $m\times m$.

Известно, что в классе линейных несмещенных оценок оценка наименьших квадратов (ОНК) минимизирует функцию риска (2) для всех θ при произвольной н. о. о. матрице $\boldsymbol{P}$. В настоящей статье, пренебрегая условием несмещенности, решается следующая

Задача. Определить линейную по наблюдениям $\boldsymbol{Y}$ оценку $\hat{\theta}=\boldsymbol{T}\boldsymbol{Y}$ параметра θ, при которой критерий оценки $q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})$ примет минимальное значение.

Показывается, что к поставленной основной задаче сводится несколько более общая задача оптимального выбора вектор-константы θ^* и матрицы $\mathbf{T}$ при условии, что оценка ищется в виде $\hat{\theta}=\theta^*+\mathbf{T}\mathbf{Y}$ и ограничения на параметр θ задаются неравенством $(\theta-\theta_0)^{\mathrm{T}}\mathbf{A}(\theta-\theta_0)\leqslant 1$. Оценка, оптимальная по критерию (5), сравнивается с ОНК. Определяется число ϱ^2, при котором в области параметров θ, задаваемой неравенством $\theta^{\mathrm{T}}\mathbf{A}\theta\leqslant\varrho^2$, полученная оценка не хуже в смысле критерия (3), чем ОНК при любом векторе $\mathbf{p}$, а следовательно, и в смысле критерия (2) при любой н. о. о. матрице $\mathbf{P}$, если при построении оптимальной оценки вместо коэффициента σ^2 использовалась некоторая величина s^2. Отметим, что оптимальная оценка представима в независимом от вектора $\mathbf{p}$ виде. В работах [2-4] получены оценки из различных иных соображений, совпадающие по виду с оптимальной по критерию (5) оценкой.

2. Приступим к решению поставленной задачи. При условии линейности оценки $\hat{\theta}$ функция риска (3) примет вид

$$r(\boldsymbol{T},\theta,\boldsymbol{p})=E\{[(\boldsymbol{T}\boldsymbol{Y}-\theta)^{\mathrm{T}}\boldsymbol{p}]^2\,|\,\theta\}=\boldsymbol{p}^{\mathrm{T}}E\{(\boldsymbol{T}\boldsymbol{Y}-\theta)(\boldsymbol{T}\boldsymbol{Y}-\theta)^{\mathrm{T}}\,|\,\theta\}\boldsymbol{p}.$$

Используя определение модели (1), после несложных преобразований имеем

$$\begin{aligned}r(\boldsymbol{T},\theta,\boldsymbol{p})&=\boldsymbol{p}^{\mathrm{T}}\{\sigma^2\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}+(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})\theta\theta^{\mathrm{T}}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})^{\mathrm{T}}\}\boldsymbol{p}=\\&=\sigma^2\boldsymbol{p}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p}+[\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})\theta]^2\end{aligned}\tag{6}$$

и критерий (5) получаем в виде

$$q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})=\sigma^2\boldsymbol{p}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p}+\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}[\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})\theta]^2.\tag{7}$$

Если $\det\boldsymbol{A}\neq 0$, то с помощью неравенства Коши—Шварца получаем

$$\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}[\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})\theta]^2=\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})\boldsymbol{A}^{-1}(\boldsymbol{I}_m-\boldsymbol{T}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{p}.\tag{8}$$

Подставляя (8) в (7), имеем

$$\begin{aligned}q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})&=\sigma^2(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p})^{\mathrm{T}}(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p})+\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{p}-2\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p})+\\&+(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p})^{\mathrm{T}}\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p}).\end{aligned}\tag{9}$$

Строго выпуклая относительно вектора $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p}$ функция (9) имеет единственный минимум в точке

$$\boldsymbol{T}^{\mathrm{T}}\boldsymbol{p}=(\sigma^2\boldsymbol{I}_n+\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{p},\tag{10}$$

и, следовательно, $\min\limits_{\boldsymbol{T}}q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})$ достигается при

$$\boldsymbol{T}_{\sigma^2}=\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}}(\sigma^2\boldsymbol{I}_n+\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})^{-1}.\tag{11}$$

Нетрудно видеть, что $\boldsymbol{T}_{\sigma^2}$ — единственная не зависящая от вектора $\boldsymbol{p}$ матрица, при которой критерий $q(\boldsymbol{T},\boldsymbol{A},\boldsymbol{p})$ принимает минимальное значение, если $\det\boldsymbol{A}\neq 0$. Умножая в равенстве (11) обе части слева на матрицу $\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, имеем

$$\begin{aligned}(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{T}_{\sigma^2}&=(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}}(\sigma^2\boldsymbol{I}_n+\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})^{-1}=\\&=\sigma^2\boldsymbol{X}^{\mathrm{T}}(\sigma^2\boldsymbol{I}_n+\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})^{-1}+\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})(\sigma^2\boldsymbol{I}_n+\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^{\mathrm{T}})^{-1}=\boldsymbol{X}^{\mathrm{T}},\end{aligned}$$

что дает

$$T_{\sigma^2} = (\sigma^2 A + X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}. \tag{12}$$

Поскольку матрица (12) оптимальна по критерию (5), если $\det A \neq 0$, то можно предположить, что она оптимальна и в случае, если условие $\det A \neq 0$ заменить условием $\det(\sigma^2 A + X^{\mathrm{T}} X) \neq 0$. Проверим это.

Если $\det(\sigma^2 A + X^{\mathrm{T}} X) \neq 0$, то

$$I_m - T_{\sigma^2} X = (\sigma^2 A + X^{\mathrm{T}} X)^{-1} (\sigma^2 A + X^{\mathrm{T}} X) - (\sigma^2 A + X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} X = \\ = \sigma^2 (\sigma^2 A + X^{\mathrm{T}} X)^{-1} A. \tag{13}$$

Используя (13), с помощью неравенства Коши—Шварца получаем

$$\sup_{\theta^{\mathrm{T}} A \theta \leqslant 1} [p^{\mathrm{T}} (I_m - T_{\sigma^2} X) \theta]^2 = \sup_{\theta^{\mathrm{T}} A \theta \leqslant 1} \sigma^4 [p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} A \theta]^2 = \\ = \sigma^4 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} A (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p. \tag{14}$$

Подставляя (12) и (14) в (7), имеем

$$q(T_{\sigma^2}, A, p) = \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} X (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p + \\ + \sigma^4 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} A (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p = \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p. \tag{15}$$

По доказанному выше при $\det A \neq 0$ выражение (15) дает минимальное значение критерия $q(T, A, p)$. Если показать, что при $\det A = 0$

$$\min_T q(T, A, p) \geqslant \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p, \tag{16}$$

то матрица T_{σ^2} будет оптимальной и в этом случае. Допустим противное, т. е. при некотором $\varepsilon > 0$

$$\min_T q(T, A, p) \leqslant \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p - \varepsilon. \tag{17}$$

Но тогда можно выбрать такую положительно определенную матрицу Δ, что

$$\sigma^2 p^{\mathrm{T}} [\sigma^2 (A + \Delta) + X^{\mathrm{T}} X]^{-1} p > \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p - \varepsilon, \tag{18}$$

где в силу того, что $\det(A + \Delta) \neq 0$,

$$\sigma^2 p^{\mathrm{T}} [\sigma^2 (A + \Delta) + X^{\mathrm{T}} X]^{-1} p = \min_T q(T, A + \Delta, p). \tag{19}$$

Из (17)—(19) следует

$$\min_T q(T, A, p) < \min_T q(T, A + \Delta, p). \tag{20}$$

Так как из $\theta^{\mathrm{T}} (A + \Delta) \theta \leqslant 1$ следует $\theta^{\mathrm{T}} A \theta \leqslant 1$, то при любой матрице T

$$q(T, A, p) \geqslant q(T, A + \Delta, p),$$

что противоречит (20), и, следовательно, (17) не имеет места.

В результате нами доказана следующая

Теорема 1. *Если* $\det(\sigma^2 A + X^{\mathrm{T}} X) \neq 0$, *то*

$$\min_T \sup_{\theta^{\mathrm{T}} A \theta \leqslant 1} r(T, \theta, p) = \sigma^2 p^{\mathrm{T}} (\sigma^2 A + X^{\mathrm{T}} X)^{-1} p \tag{21}$$

и минимум достигается при матрице

$$T_{\sigma^2} = (\sigma^2 A + X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}. \tag{22}$$

Нетрудно доказывается и обобщение теоремы 1, а именно

Теорема 2. *Если вектор $\boldsymbol{p}$ принадлежит линейному подпространству, порожденному вектор-столбцами матрицы $\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, то*

$$\min_{\boldsymbol{T}}\ \sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1} r(\boldsymbol{T},\theta,\boldsymbol{p})=\sigma^2\boldsymbol{p}^{\mathrm{T}}(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-}\boldsymbol{p}$$

и минимум достигается при матрице

$$\boldsymbol{T}_{\sigma^2}=(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-}\boldsymbol{X}^{\mathrm{T}},$$

где $(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-}$ *— любая обобщенная обратная к* $\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ *матрица* [5].

Если условие теоремы 2 не выполнено, то задача оптимизации оценки по критерию (5) не имеет смысла, поскольку в этом случае не существует матрицы $\boldsymbol{T}$, при которой критерий (5) принимает конечное значение.

Замечание. Пусть вектор $\boldsymbol{p}$ принадлежит линейному подпространству, порожденному вектор-столбцами матрицы $\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. Тогда

$$\min_{\boldsymbol{T}}\ \sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1} r(\boldsymbol{T},\theta,\boldsymbol{p})=\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}\ \min_{\boldsymbol{T}} r(\boldsymbol{T},\theta,\boldsymbol{p}). \tag{23}$$

Действительно, минимизируя выражение (6), имеем

$$\min_{\boldsymbol{T}} r(\boldsymbol{T},\theta,\boldsymbol{p})=\sigma^2(\theta^{\mathrm{T}}\boldsymbol{p})^2/(\sigma^2+\theta^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\theta). \tag{24}$$

Поскольку производная этой функции по θ равна нулю только при $\theta^{\mathrm{T}}\boldsymbol{p}=0$, то при ограничениях на θ в виде замкнутой ограниченной области супремум функции достигается на границе и при $\varepsilon>0$ справедливо

$$\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}\ \min_{\boldsymbol{T}} r(\boldsymbol{T},\theta,\boldsymbol{p})\geqslant \sup_{\theta^{\mathrm{T}}(\boldsymbol{A}+\varepsilon\boldsymbol{I}_m)\theta\leqslant 1}\sigma^2(\theta^{\mathrm{T}}\boldsymbol{p})^2/(\sigma^2+\theta^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\theta)=$$
$$=\sup_{\theta^{\mathrm{T}}(\boldsymbol{A}+\varepsilon\boldsymbol{I}_m)\theta=1}\sigma^2(\theta^{\mathrm{T}}\boldsymbol{p})^2/[\theta^{\mathrm{T}}(\sigma^2\boldsymbol{A}+\sigma^2\varepsilon\boldsymbol{I}_m+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\theta]=$$
$$=\sigma^2\boldsymbol{p}^{\mathrm{T}}(\sigma^2\boldsymbol{A}+\sigma^2\varepsilon\boldsymbol{I}_m+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-}\boldsymbol{p}.$$

Последнее выражение непрерывно по ε в точке 0, следовательно,

$$\sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1}\ \min_{\boldsymbol{T}} r(\boldsymbol{T},\theta,\boldsymbol{p})\geqslant\sigma^2\boldsymbol{p}^{\mathrm{T}}(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-}\boldsymbol{p}=\min_{\boldsymbol{T}}\ \sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta\leqslant 1} r(\boldsymbol{T},\theta,\boldsymbol{p}).$$

Обратное неравенство известно из теории игр, и тем самым равенство (23) доказано.

3. В предыдущем пункте решена задача определения линейной оценки параметра θ, если ограничения на параметр заданы в виде эллипсоида с центром в нуле. Покажем, что расширяя ограничения на параметр θ до класса эллипсоидов с произвольным центром θ_0, а также рассматривая более широкий класс оценок, представимых в виде суммы константы и линейной статистики, с помощью решенной задачи просто решается следующая

Задача. Пусть параметр θ оценивается статистикой $\hat{\theta}=\theta^*+\boldsymbol{T}\boldsymbol{Y}$. Определить вектор θ^*_O и матрицу $\boldsymbol{T}_O$, при которых критерий оценки

$$\sup_{(\theta-\theta_0)^{\mathrm{T}}\boldsymbol{A}(\theta-\theta_0)\leqslant 1} r(\theta^*,\boldsymbol{T},\theta,\boldsymbol{p})=$$

$$= \sup_{(\theta-\theta_0)^{\mathrm{T}} A (\theta-\theta_0) \leqslant 1} E\{[(\theta^* + TY - \theta)^{\mathrm{T}} p]^2 \mid \theta\}$$

примет минимальное значение.

Введем обозначение $Z = Y - X\theta_0$. Тогда

$$E\{Z \mid \theta\} = X(\theta - \theta_0) \quad \text{и} \quad \operatorname{Cov}\{Z \mid \theta\} = \sigma^2 I_n.$$

Несложными преобразованиями критерий оценки приводится к виду

$$\sup_{(\theta-\theta_0)^{\mathrm{T}} A (\theta-\theta_0) \leqslant 1} r(\theta^*, T, \theta, p) =$$

$$= \sup_{(\theta-\theta_0)^{\mathrm{T}} A (\theta-\theta_0) \leqslant 1} \{E\{[(TZ - (\theta - \theta_0))^{\mathrm{T}} p]^2 \mid \theta\} -$$

$$- 2(\theta - \theta_0)^{\mathrm{T}} (I_m - TX)^{\mathrm{T}} p p^{\mathrm{T}} (\theta^* - (I_m - TX)\theta_0) + [(\theta^* - (I_m - TX)\theta_0)^{\mathrm{T}} p]^2\},$$

откуда следует, что

$$\theta_O^* = (I_m - T_O X)\theta_0.$$

Оптимальная матрица T_O получается минимизацией первого слагаемого в выражении критерия и дается теоремой 2, т. е.

$$T_O = (\sigma^2 A + X^{\mathrm{T}} X)^{-} X^{\mathrm{T}},$$

если вектор p принадлежит линейному подпространству, порожденному вектор-столбцами матрицы $\sigma^2 A + X^{\mathrm{T}} X$ (в противном случае задача не имеет смысла).

Таким образом, оптимальной является следующая статистика:

$$\hat{\theta} = \theta_0 - (\sigma^2 A + X^{\mathrm{T}} X)^{-} X^{\mathrm{T}} X \theta_0 + (\sigma^2 A + X^{\mathrm{T}} X)^{-} X^{\mathrm{T}} Y.$$

4. При практическом оценивании параметра θ неизвестные в общем случае истинные коэффициенты σ^2 и ϱ^2 (последний определяет эллипсоид ограничений на параметр θ в виде $\theta^{\mathrm{T}} A \theta \leqslant \varrho^2$) как правило могут быть благодаря априорной информации оценены соответственно значениями $s^2 > 0$ и $r^2 > 0$. Не умаляя общности, можно принять $r^2 = 1$. В описанной ситуации естественно пользоваться оценкой $\theta_{s^2} = T_{s^2} Y$, где $T_{s^2} = (s^2 A + X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}$ (в предположении, что $\det(s^2 A + X^{\mathrm{T}} X) \neq 0$). Исходя из этого интересно знать, как ведет себя супремум функции риска при произвольных значениях σ^2 и ϱ^2, быть может не совпадающих с их оценками. В результате рассуждений, аналогичных пункту 2, оказывается, что

$$\sup_{\theta^{\mathrm{T}} A \theta \leqslant \varrho^2} r(T_{s^2}, \theta, p) =$$

$$= \sigma^2 p^{\mathrm{T}} (s^2 A + X^{\mathrm{T}} X)^{-1} (\varrho^2 s^4 / \sigma^2 A + X^{\mathrm{T}} X)(s^2 A + X^{\mathrm{T}} X)^{-1} p. \tag{25}$$

Сравним полученное значение критерия оценки с ее значением при оценивании параметра θ методом наименьших квадратов. В случае нулевой матрицы A, т. е. при отсутствии ограничений, минимаксная линейная оценка совпадает с ОНК. Интерпретация ОНК как минимаксной была дана также Г. А. Бернардом [6]. Рассмотрим случай ненулевой матрицы $\boldsymbol{A}$. Для простоты изложения будем считать $\det(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) \neq 0$. Тогда ОНК соответствует матрица $\boldsymbol{T}_{\text{ОНК}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ и

$$r(\boldsymbol{T}_{\text{ОНК}}, \theta, \boldsymbol{p}) = \sigma^2 \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{p}. \tag{26}$$

Будем считать оценку $\hat{\theta}_{s^2}$ не хуже ОНК в области параметров θ, определяемой неравенством $\theta^{\mathrm{T}}\boldsymbol{A}\theta \leqslant \varrho^2$, если при каждом $\boldsymbol{p}$

$$r(\boldsymbol{T}_{\text{ОНК}}, \theta, \boldsymbol{p}) - \sup_{\theta^{\mathrm{T}}\boldsymbol{A}\theta \leqslant \varrho^2} r(\boldsymbol{T}_{s^2}, \theta, \boldsymbol{p}) \geqslant 0. \tag{27}$$

Неравенство (27) эквивалентно требованию неотрицательной определенности любой из следующих матриц:

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} - (s^2\boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(\varrho^2 s^4/\sigma^2 \boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})(s^2\boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1},$$
$$(s^2\boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(s^2\boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) - (\varrho^2 s^4/\sigma^2 \boldsymbol{A} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}),$$
$$s^2\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{A} + (2 - \varrho^2 s^2/\sigma^2)\boldsymbol{A},$$
$$s^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2} + (2 - \varrho^2 s^2/\sigma^2)(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}. \tag{28}$$

Условие неотрицательной определенности матрицы (28) эквивалентно неравенству

$$\min_{i \in \{1, \ldots, m\}} [s^2\lambda_i^2 + (2 - \varrho^2 s^2/\sigma^2)\lambda_i] \geqslant 0,$$

где $\lambda_i (i = 1, \ldots, m)$ — собственные числа матрицы $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}$.

Следовательно, для того чтобы оценка $\hat{\theta}_{s^2} = \boldsymbol{T}_{s^2}\boldsymbol{Y}$ была не хуже, чем ОНК, необходимо и достаточно выполнения условия

$$\varrho^2 \leqslant 2\sigma^2/s^2 + \sigma^2\lambda_+, \tag{29}$$

где λ_+ — минимальное положительное собственное число матрицы $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1/2}$.

Из неравенства (29) видно, что чем меньше выбранное значение s^2 в оценке $\hat{\theta}_{s^2}$, тем больше область в пространстве параметров θ, где оценка $\hat{\theta}_{s^2}$ не хуже, чем ОНК, но выигрыш при близких к нулю значениях $\hat{\theta}$ будет меньше, что следует из выражения (25). Оценка $\hat{\theta}_{s^2}$ является оптимальной в смысле минимизации супремума функции риска $r(\boldsymbol{T}, \theta, \boldsymbol{p})$ по области θ, задаваемой неравенством $\theta^{\mathrm{T}}\boldsymbol{A}\theta \leqslant \sigma^2/s^2$.

Пример. Требуется оценить функцию $f(t) = \theta_1 + \theta_2 t$ вещественного аргумента t на основе равноточных измерений y_1 и y_2 с некоррелированной аддитивной помехой в точках $t = 1$ и $t = -1$, т. е.

$$y_1 = f(1) + \varepsilon_1, \quad y_2 = f(-1) + \varepsilon_2, \quad E(\varepsilon_i) = 0, \quad E(\varepsilon_i\varepsilon_j) = \sigma^2\delta_{ij},$$

где δ_{ij} — символ Кронекера. Это можно записать и в следующем виде:

$$f(t) = \theta^{\mathrm{T}}\boldsymbol{p}(t), \quad E\{\boldsymbol{Y} \mid \theta\} = \boldsymbol{X}\theta \quad \text{и} \quad \mathrm{Cov}\{\boldsymbol{Y} \mid \theta\} = \sigma^2\boldsymbol{I}_2,$$

где

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad \boldsymbol{p}(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{p}^{\mathrm{T}}(1) \\ \boldsymbol{p}^{\mathrm{T}}(-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Оптимизируем оценку для множества таких прямых $f(t)$, у которых разность действительных значений в точках измерения не превышает по абсолютной величине значения $\Delta > 0$, т. е.

$$(2\theta_2)^2 \leqslant \Delta^2 \quad \text{или} \quad \theta^{\mathrm{T}}\boldsymbol{A}\theta \leqslant 1, \quad \text{где} \quad \boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 0 & 4\Delta^{-2} \end{pmatrix}.$$

В качестве критерия оценки $\hat{f}(t)=\hat{\theta}^{\mathsf{T}}\boldsymbol{p}(t)$ рассмотрим функцию

$$q(\hat{f},\boldsymbol{A},t)=\sup_{\theta^{\mathsf{T}}\boldsymbol{A}\theta\leqslant 1} E\{[(\hat{\theta}-\theta)^{\mathsf{T}}\boldsymbol{p}]^2\mid\theta\},\tag{30}$$

которая при ОНК примет не зависимый от $\boldsymbol{A}$ вид

$$q(\hat{f}_{\text{онк}},t)=\sigma^2\boldsymbol{p}^{\mathsf{T}}(t)\,(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{p}(t).\tag{31}$$

Оптимальной по критерию (30) линейной оценкой при произвольном значении аргумента t согласно теореме 1 является

$$\hat{f}_{\sigma^2}=\hat{\theta}_{\sigma^2}^{\mathsf{T}}\boldsymbol{p}(t),\quad \hat{\theta}_{\sigma^2}=(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y},$$

при которой критерий (30) примет значение

$$q(\hat{f}_{\sigma^2},\boldsymbol{A},t)=\sigma^2\boldsymbol{p}^{\mathsf{T}}(t)\,(\sigma^2\boldsymbol{A}+\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{p}(t).\tag{32}$$

Подставляя в (31) и (32) матрицы $\boldsymbol{A}$ и $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}$, получаем

$$q(\hat{f}_{\text{онк}},t)=\sigma^2(1/2+t^2/2),\quad q(\hat{f}_{\sigma^2},\boldsymbol{A},t)=\sigma^2[1/2+t^2/(2+4\sigma^2/\Delta^2)].$$

Поскольку на отрезке между точками измерений эти функции достигают максимального значения при $t=\pm 1$, то выигрыш, получаемый при использовании оптимальной оценки, естественно определять функцией

$$\eta(\sigma/\Delta)=[q(\hat{f}_{\text{онк}},1)-q(\hat{f}_{\sigma^2},\boldsymbol{A},1)]/q(\hat{f}_{\text{онк}},1)=1/2-1/(2+4\sigma^2/\Delta^2),$$

некоторые значения которой следующие:

$$\eta(1)=0{,}333,\quad \eta(1/2)=0{,}167,\quad \eta(1/3)=0{,}090,\quad \eta(1/4)=0{,}056,$$
$$\eta(1/5)=0{,}037.$$

ЛИТЕРАТУРА

1. Ченцов Н. Н., Теория вероятн. и ее применения, XII, 619 (1967).
2. Raiffa H., Schlaifer R., Applied statistical decision theory, Boston, 1961.
3. Theil H., J. Amer. Statist. Assoc., 58, 401 (1963).
4. Hoerl A. E., Kennard R. W., Technometrics, 12, 55 (1970).
5. Рао С. Р., Линейные статистические методы и их применения, М., 1968.
6. Barnard G. A., J. Roy. Statist. Soc., Ser. B, 25, 124 (1963).

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
15/III 1971

I. KUKS, V. OLMAN

REGRESSIOONIKOEFITSIENTIDE LINEAARNE MINIMAKSHINNANG

Leitakse regressioonmudeli kordajatele ruutkaofunktsiooni matemaatilise ootuse supreemumit miniseeriv lineaarne hinnang lisatingimusel, et kordajate vektor kuulub antud nulltsentriga ellipsoidi.

I. KUKS, V. OLMAN

A MINIMAX LINEAR ESTIMATOR OF REGRESSION COEFFICIENTS

A linear estimator of coefficients in a regression model is found, which minimizes the supremum of an averaged quadratic loss function if the coefficients' vector belongs to a given ellipsoid with center in zero.