Галкин В.Я., Меченов А.С.

О РЕГУЛЯРИЗАЦИИ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ ЗАДАЧ

Введение

При обработке результатов наблюдений в самых различных областях знаний возникает необходимость перехода от реальных экспериментальных устройств к их математическому описанию. Иными словами по вполне понятным причинам нам неизбежно приходится говорить о математической модели явления или экспериментальной ситуации. При этом чаще всего приходится иметь дело с результатами измерений, носящими случайный характер.

Наиболее употребительной моделью является линейная, связывающая интересующие исследователя характеристики явления с наблюденными величинами линейно при фиксированных состояниях исследуемой системы. Когда условное математическое ожидание наблюдений представляет собой линейную форму от искомых параметров, говорят о линейной регрессионной модели. Ее изучению посвящена обширная математическая литература [1, 13, 23, 26]. При этом основное внимание уделяется моделям полного ранга, т.е. корректно поставленным задачам, установлению статистических свойств оценок параметров и проверке адекватности моделей в различных предположениях.

Предметом нашего рассмотрения является линейная регрессионная модель неполного ранга, т.е. некорректно поставленная задача [3, 6, 7, 9]. В связи с широким использованием ЭВМ при обработке результатов наблюдений исследователь редко имеет возможность (за исключением диалоговых систем) вмешиваться в процесс численного решения задачи. Кроме того, всякое изменение модели требует дополнительной информации о явлении, которая не всегда доступна. Поэтому развиваются устойчивые методы оценки параметров и проверки адекватности в моделях неполного ранга [2, 5, 11, 16, 17, 19-22, 24-27]. В настоящей работе рассматриваются вопросы оценивания параметров в линейной регрессионной модели методами регуляризации, псевдообращения и байесовским подходом с учетом неточного задания исходных данных и корреляций. Для такой задачи строятся регуляризирующие алгоритмы, изучается их связь с псевдообращением, с байесовскими процедурами и с наилучшими линейными несмещенными оценками.

§1. Оценивание параметров в линейной регрессионной модели

Будем исходить из следующей множественной линейной модели наблюдений

$$y = X\beta + \varepsilon,\ E\varepsilon = 0,\ E\varepsilon\varepsilon' = \sigma^2 I, \tag{1.1}$$

где X — матрица порядка $n \times p$ (α строк, p столбцов) полного ранга ($rankX = r = p$) считается известной, $\beta = (\beta_1, \ldots, \beta_p)'$ — p-мерный вектор неизвестных параметров, $y = (y_1, \ldots, y_n)'$ — вектор наблюдений (случайный вектор), $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ — вектор случайных ошибок наблюдений, компоненты которого независимы и одинаково распределены,

$$E\varepsilon_i = 0, E\varepsilon_i^2 = \sigma^2, i = \overline{1,n}, \operatorname{var}\varepsilon = \sigma^2 I,$$

где $I = I_n$ — единичная матрица порядка $n \times n$.

Обычно применяемый для решения задачи (1.1) метод наименьших квадратов состоит в нахождении оценки $\hat{\beta} = \hat{\beta}(y)$ вектора β такой, что

$$\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right) = \min_{\beta} \varepsilon'\varepsilon = \min_{\beta} (y - X\beta)'(y - X\beta). \tag{1.2}$$

Необходимое условие минимума по β

$$\frac{d}{d\beta}\varepsilon'\varepsilon = 0$$

приводит к системе нормальных уравнений

$$X'X\beta = X'y. \tag{1.3}$$

Поскольку X — матрица полного ранга, т.е. $\det(X'X) \neq 0$, то

$$\hat{\beta} = (X'X)^{-1} X'y \tag{1.4}$$

и называется м.н.к.-оценкой. Ее свойства: несмещенность —

$$E\hat{\beta} = \beta,$$

ковариационная матрица —

$$\operatorname{var}\hat{\beta} = \sigma^2 (X'X)^{-1},$$

при этом несмещенной оценкой для σ^2 является

$$\hat{\sigma}^2 = \frac{\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)}{n - p}.$$

В классе несмещенных линейных оценок параметра β м.н.к.-оценка $\hat{\beta}$ обладает минимальной дисперсией (теорема Гаусса-Маркова).

В предположении нормального распределения ошибок ε статистика

$$F = \frac{n-p}{p} \frac{\left(\beta - \hat{\beta}\right)' X'X \left(\beta - \hat{\beta}\right)}{\left(y - X\hat{\beta}\right)' \left(y - X\hat{\beta}\right)}$$

имеет распределение Фишера-Снедекора с параметрами p и $n-p$. Поскольку в этом случае $\hat{\beta}$ совпадает с оценкой максимального правдоподобия, то статистика $\left(\hat{\beta}, \hat{\sigma}^2\right)$ является достаточной и полной.

Введем квадрат уклонения м.н.к.-оценки от ее математического ожидания

$$L^2 = \left(\beta - \hat{\beta}\right)' \left(\beta - \hat{\beta}\right). \tag{1.5}$$

Тогда имеем для математического ожидания

$$EL^2 = \sigma^2 tr\left(X'X\right)^{-1}, \tag{1.6}$$

т.е.

$$E\hat{\beta}'\hat{\beta} = \beta'\beta + \sigma^2 tr\left(X'X\right)^{-1},$$

и, кроме того, дисперсия

$$\operatorname{var} L^2 = 2\sigma^4 tr\left(X'X\right)^{-2}. \tag{1.7}$$

Переходя к собственным значениям матрицы $X'X$

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{\min} > 0,$$

будем иметь для (1.6) и (1.7) соответственно

$$EL^2 = \sigma^2 \sum_{i=1}^{p} \lambda_i^{-1}, \tag{1.8}$$

$$\operatorname{var} L^2 = 2\sigma^4 \sum_{i=1}^{p} \lambda_i^{-2}. \tag{1.9}$$

Ясно, что при малых собственных значениях величина EL^2 становится большой. Это равносильно плохой обусловленности матрицы $X'X$. При решении практических регрессионных задач часто предлагается в таком случае изменить структуру матрицы X или размерность вектора β (улучшить исходную модель). Такой подход — проблема моделирования в конкретной предметной области, и она нетривиальна, особенно когда описываемое моделью (1.1) явление не характеризуется точно заданным набором параметров β, т.е. исследователь находится в ситуации "черного ящика".

В таких случаях часто используют процедуру псевдообращения [1, 2, 16, 17, 25, 26]. Другой подход заключается в применении регуляризирующих алгоритмов [3-12, 14-15, 18-22, 27]. Вопросы, связанные с неточностью задания матрицы X обсуждаются в [6, 18-21,28,29]. Байесовский подход рассматривается в [5, 24].

§ 2. Построение регуляризованных оценок

Рассмотрим некорректно поставленную задачу линейной регрессии

$$y = X\beta + \varepsilon, E\varepsilon = 0, E\varepsilon\varepsilon' = \sigma^2 I, \qquad (2.1)$$

где $X'X$ — вырождена (имеет ранг $r < p$) или плохо обусловлена. Идеология метода наименьших квадратов в этом случае остается прежней

$$(y - Xb)'(y - Xb) = \min_{\beta} \varepsilon'\varepsilon = \min_{\beta}(y - X\beta)'(y - X\beta), \qquad (2.2)$$

и поскольку система нормальных уравнений (1.3) всегда разрешима приводит к бесконечному набору оценок b истинного значения параметров β.

Определение. *Нормальной оценкой* b_0 *назовем оценку, минимизирующую* $b'b$:

$$b_0'b_0 = \min b'b. \qquad (2.3)$$

Пусть β_0 — вектор параметров с минимальной нормой (нормальный) для

$$X\beta = Ey.$$

Тогда

$$Eb_0 = \beta_0 \neq \bar{\beta},$$

т.е. в общем случае математическое ожидание нормальной оценки не совпадает с истинным значением параметра $\bar{\beta}$.

Естественным образом регуляризирующий алгоритм [3, 4, 6] построения b_0 записывается как следующая задача на условный экстремум:

$$\min \beta'\beta$$

при условии, что остаточная сумма квадратов меньше или равна своему математическому ожиданию при истинном значении параметров (или сумме квадратов ошибок)

$$(y - X\beta)'(y - X\beta) \leq \|y - X\bar{\beta}\|^2 \equiv n\sigma^2. \qquad (2.4)$$

Как известно, данная постановка равносильна минимизации $\beta'\beta$ при

$$(y - X\beta)'(y - X\beta) = n\sigma^2, \tag{2.5}$$

т.е. минимизации сглаживающего функционала

$$M^\alpha[\beta] = (y - X\beta)'(y - X\beta) + \alpha\beta'\beta \tag{2.6}$$

(константа $n\sigma^2$ опущена) при условии (2.5).

Необходимое условие минимума (2.6)

$$\frac{dM^\alpha}{d\beta} = 0$$

приводит к нормальному уравнению Эйлера

$$(X'X + \alpha I_p)\beta = X'y. \tag{2.7}$$

Откуда регуляризованная оценка b^α имеет вид

$$b^\alpha = (X'X + \alpha I_p)^{-1} X'y \tag{2.8}$$

при выборе α из условия (2.5).

Из общих результатов теории регуляризирующих алгоритмов следует сходимость по вероятности

$$b^\alpha \xrightarrow{P} \beta_0,$$

если $\sigma \to 0$, а α выбрано из (2.5). Если же $\sigma = const$, то при $\alpha \to 0$ имеет место сходимость

$$b^\alpha \to b_0.$$

При практическом решении регрессионных задач значение σ^2 обычно неизвестно. Поэтому рассматриваются и другие способы построения регуляризованных оценок. Одной из них является задача отыскания квазирешения [5, 9, 10].

Определение. *Квазиоценкой* (2.1) *назовем оценку* α, *минимизирующую*

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

на компактном множестве $\beta'\beta \le R^2$, *где* R *— заданная постоянная.*

Аналогично предыдущему задача отыскания квазирешения сводится к минимизации функционала

$$(y - X\beta)'(y - X\beta) + \alpha(\beta'\beta - R^2)$$

при условии

$$\beta'\beta = R^2. \tag{2.9}$$

Это приводит к уравнению (2.7) с выбором параметра α из (2.9), что также возможно только при известном R. Однако информация об R столь же редка как и информация о σ^2. Кроме того, если $\beta_0'\beta_0 > R^2$, то b^* — смещенная оценка и

$$E\left(b^*\right)' b^* < \beta_0'\beta_0.$$

В [11] из интуитивных соображений записывалась так называемая гребневая оценка

$$\hat{\beta}^* = \left(X'X + kI\right)^{-1} X'y, \tag{2.10}$$

где $k > 0$ ищется из [5]

$$\min EL^2(k) = \min E\left(\hat{\beta}^* - \beta\right)' \left(\hat{\beta}^* - \beta\right). \tag{2.11}$$

Там же [5] доказана следующая

Теорема (существования). *Всегда существует такое* $k > 0$ *(на классе оценок* (2.10)*), что*

$$EL^2(k) < EL^2(0) = \sigma^2 \sum_{i=1}^{p} \lambda_i^{-1}.$$

Она показывает, что все оценки типа регуляризованных предпочтительнее нормальных. Как показано в [12]

$$b_0 = \sum_{i=0}^{\infty} (-1)^i \frac{\alpha^i}{i!} \frac{\partial^i b^\alpha}{\partial \alpha^i}, \tag{2.12}$$

т.е. любая регуляризованная оценка b^α является первым членом разложения нормальной оценки b_0 в функциональный ряд по параметру регуляризации α (при значении α, найденном из условий (2.5), (2.9) или (2.11)).

Заметим, что при неизвестном σ^2 несмещенной оценкой для него является

$$\hat{\sigma}^2 = \frac{\left(y - Xb_0\right)' \left(y - Xb_0\right)}{n - r},$$

вычисление которой требует знания ранга r матрицы X.

Вообще говоря, все что мы можем требовать от качества произвольной оценки b — это, чтобы она минимизировала средний квадрат уклонения относительно истинного значения $\bar{\beta}$ [13]:

$$\min EL^2 = \min E\left(b - \bar{\beta}\right)' \left(b - \bar{\beta}\right) = \min\left\{\operatorname{var} b'b + \left(Eb - \bar{\beta}\right)' \left(Eb - \bar{\beta}\right)\right\}. \tag{2.13}$$

Здесь второе слагаемое есть квадрат смещения b при оценивании $\bar{\beta}$. Оно равно нулю, если b — не смещена относительно $\bar{\beta}$, в частности, это так, когда $\bar{\beta} = \beta_0$ и $b = b_0$, т.е. b_0 является нормальной оценкой. Таким образом, некоторая оценка b, более близкая к $\bar{\beta}$, чем b_0, может иметь перед b_0 преимущество, но $\operatorname{var} b'b$ для нее всегда больше, чем $\operatorname{var} b_0'b_0$. Однако построение такого b требует дополнительной априорной информации о $\bar{\beta}$. Такой информацией может служить, например, ограниченность $\bar{\beta}$ [2], что приводит к задаче линейного программирования, или — условие $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_p$ (аналог монотонности), что, аналогично [14, 15] приводит к регуляризирущему алгоритму решения задачи (2.1).

Статистические свойства такой оценки b существенно зависят от априорной информации и в каждом конкретном случае должны изучаться индивидуально. Тогда как в предположении $\bar{\beta} = \beta_0$ статистические свойства нормальной оценки b_0 во многом аналогичны свойствам м.н.к.-оценки $\hat{\beta}$, поэтому для решения задачи (2.1) широко используется аппарат псевдообратных матриц.

§ 3. Неточно заданная или измеренная матрица *X*

Пусть задана линейная регрессионная модель

$$y = \tilde{X}\beta + \varepsilon, E\varepsilon = 0, E\varepsilon\varepsilon' = \sigma^2 I_n, \tag{3.1}$$

где X — задаваемая экспериментатором матрица ранга p, $\tilde{X} = X + \Xi$ и Ξ — матрица случайных ошибок, элементы которой ξ_{ij} — независимые одинаково распределенные случайные величины с

$$E\xi_{ij} = 0, E\xi_{ij}^2 = \rho^2.$$

Эта модель называется моделью с активными ошибками. Поскольку полная ошибка $u = \Xi\beta + \varepsilon$ зависит от искомых параметров, то метод наименьших квадратов неприменим. Если считать ошибки нормально распределенными, то применим метод максимального правдоподобия [18], откуда оценка параметров невырожденной модели имеет вид

$$\hat{b} = Arg\min_{\beta}\left(X\beta - y\right)'\left(Euu'\right)^{-1}\left(X\beta - y\right) + \ln\det Euu' + n\ln 2\pi =$$

$$= Arg\min_{\beta}\frac{\|X\beta - y\|^2}{\sigma^2 + \rho^2\|\beta\|^2} + n\ln\left(\sigma^2 + \rho^2\|\beta\|^2\right) + n\ln 2\pi$$

и, для своего вычисления, предполагает знание дисперсий ошибок.

В том случае, когда матрица X имеет неполный ранг r (а исследователь это знает, поскольку задает значения этой матрицы), матрица $\tilde{X}$ имеет полный ранг с вероятностью 1. По аналогии с (2.4) построим [18] регуляризованное приближение b_{ϖ} к вектору нормальных параметров β_0, таким образом, чтобы квадрат нормы этого приближения был минимален, когда отрицательный удвоенный логарифм функции правдоподобия меньше или равен своему математическому ожиданию ϖ^2 при истинных значениях параметров $\bar{\beta}$

$$b_{\varpi} = Arg \min_{\beta:(X\beta-y)'(Euu')^{-1}(X\beta-y)+\ln\det Euu'+n\ln 2\pi\le\varpi^2} |\beta|^2. \tag{3.2}$$

Теорема. *Устойчивое приближение к нормальному вектору параметров существует, единственно и удовлетворяет соотношению*

$$b_{\varpi} = Arg \min_{\beta:\frac{|X\beta-y|^2}{\sigma^2+\rho^2|\beta|^2}+n\ln\left(\sigma^2+\rho^2|\beta|^2\right)+n\ln 2\pi=\varpi^2} |\beta|^2 \tag{3.3}$$

и, как следствие, системе линейных алгебраических уравнений

$$\begin{cases}\left[X'X+\left(\alpha\left(\sigma^2+\rho^2|\beta_\alpha|^2\right)+\rho^2\left(n-\dfrac{|X\beta_\alpha\text{-}y|^2}{\sigma^2+\rho^2|\beta_\alpha|^2}\right)\right)\mathrm{I}\right]\beta_\alpha = X'y,\\ \dfrac{|X\beta_\alpha\text{-}y|^2}{\sigma^2+\rho^2|\beta_\alpha|^2}+n\ln\left(\sigma^2+\rho^2|\beta_\alpha|^2\right)+n\ln 2\pi=\varpi^2.\end{cases} \tag{3.4}$$

Эта система линейных алгебраических уравнений похожа на уравнение Эйлера (2.7).

Неточное измерение матрицы X существенно усложняет задачу. Пусть имеется линейная регрессионная модель

$$y = X\beta+\varepsilon, E\varepsilon=0, E\varepsilon\varepsilon'=\sigma^2 I_n, \tag{3.5}$$

где X — матрица ранга p измеряется так, что $\tilde{X}=X+\Xi$ и Ξ — матрица случайных ошибок, элементы которой ξ_{ij} — независимые одинаково распределенные случайные величины с

$$E\xi_{ij}=0, E\xi_{ij}^2=\mu^2.$$

Эту модель называют моделью с пассивными ошибками. Оценка ее параметров методом наименьших расстояний [18,28,29] имеет вид

$$\hat{b} = Arg\min_{\beta}\min_{X,\eta:X\beta=\eta}\frac{|\tilde{X}-X|^2}{\mu^2}+\frac{|\eta-y|^2}{\sigma^2} = Arg\min_{\beta}\frac{\left(\tilde{X}\beta-y\right)'\left(\tilde{X}\beta-y\right)}{\sigma^2+\mu^2\beta'\beta}, \tag{3.6}$$

где $|\tilde{X}|^2=\sum_{i=1}^{n}\sum_{j=1}^{p}\tilde{x}_{ij}^2$, и предполагает знание для своего вычисления по меньшей мере отношения дисперсий ошибок.

В том случае, когда неизвестная матрица X имеет неполный ранг r (что для исследователя в общем то остается загадкой), матрица $\tilde{X}$ имеет полный ранг с вероятностью 1 и может быть только плохо обусловлена. Аналогично [18], определим обобщенную регуляризованную оценку $\tilde{b}_{\alpha_{\sigma,\mu}}$ вектора нормальных параметров β_0 как решение вариационной задачи

$$\tilde{b}_{\alpha_{\sigma,\mu}} = Arg \min_{\beta:\frac{(\tilde{X}\beta-y)'(\tilde{X}\beta-y)}{\sigma^2+\mu^2\beta'\beta} \le \frac{\|\tilde{X}\beta-y\|^2}{\sigma^2+\mu^2\|\beta\|^2}=n} \beta'\beta, \tag{3.7}$$

которая вычисляется из системы линейных алгебраических уравнений

$$\begin{cases} \left(\tilde{X}'\tilde{X}+\left(\alpha\left(\sigma^2+\mu^2|\tilde{b}_\alpha|^2\right)-\dfrac{|\tilde{X}\tilde{b}_\alpha-y|^2}{\sigma^2+\mu^2|\tilde{b}_\alpha|^2}\right)I\right)\tilde{b}_\alpha=\tilde{X}'y, \\ \dfrac{|\tilde{X}\tilde{b}_\alpha-y|^2}{\sigma^2+\mu^2|\tilde{b}_\alpha|^2}=n, \end{cases} \equiv \begin{cases} \left(\tilde{X}'\tilde{X}+\alpha_{\sigma,\mu}I\right)\tilde{b}_{\alpha_{\sigma,\mu}}=\tilde{X}'y, \\ \dfrac{|\tilde{X}\tilde{b}_{\alpha_{\sigma,\mu}}-y|^2}{\sigma^2+\mu^2|\tilde{b}_{\alpha_{\sigma,\mu}}|^2}=n, \end{cases} \tag{3.8}$$

также похожей на уравнение Эйлера (2.7).

§ 4. Оценки псевдообращения

Рассмотрим, по-прежнему, задачу (2.1), метод наименьших квадратов для которой приводит к нормальному уравнению (1.3) с вырожденной (ранга $r<p$) или плохо обусловленной матрицей $X'X$. Бесконечный набор оценок b параметра $\bar{\beta}$ может быть записан через обобщенную обратную матрицу B^- [1,2].

Определение. *Обобщенная обратная матрица для произвольной $n\times p$ матрицы B определяется как любая матрица B^-, удовлетворяющая условию*

$$BB^-B=B. \tag{4.1}$$

Такая матрица существует всегда и не единственна. Если положить $B=X'X$, то

$$b=B^-X'y=X^-y,\left(X^-=B^-X'\right). \tag{4.2}$$

Матрица B^- удовлетворяющая, кроме (4.1), еще трем условиям

$$\begin{gathered} B^-BB^-=B^-, \\ \left(BB^-\right)'=BB^-, \\ \left(B^-B\right)'=B^-B, \end{gathered} \tag{4.3}$$

называется псевдообратной матрицей Мура-Пенроуза [2] и обозначается через B^+ (она единственна). Справедлива [1,2] следующая

Теорема. *Оценка b минимизирует (2.2) тогда и только тогда, когда b представима в форме*

$$b = X^{+}y + \left(I - X^{+}X\right)t, \tag{4.4}$$

где t — произвольный p-мерный вектор.

Поэтому нормальная оценка b_0 (2.3) записывается также, как следует из [2, 16], в виде

$$b_0 = X^{+}y \tag{4.5}$$

и представляет собой аналог м.н.к.-оценки (1.4). Однако при численном построении псевдообратной матрицы требуется или априорная информация о точном ранге матрицы X, или — решение каким-то образом проблемы ее ранжирования. Поэтому перейдем к другому подходу определения и вычисления псевдообратной матрицы.

С целью установления связи между регуляризацией и псевдообращением рассмотрим следующую вариационную задачу

$$X^{+} = \arg \min_{X^{-}:\left\|XX^{-} - I_n\right\| \le n-r} \left\|X^{-}\right\|^2, \tag{4.6}$$

где матричная и векторная нормы согласованы.

Оказывается, решение X^{+} такой задачи существует, единственно и удовлетворяет равенству

$$\left\|XX^{+} - I_n\right\| = tr\left(XX^{+} - I_n\right) = n - r. \tag{4.7}$$

Отсюда (4.6) равносильна минимизации функционала Тихонова

$$M^{\alpha}\left[X^{-}\right] = \left\|XX^{-} - I_n\right\|^2 + \alpha\left\|X^{-}\right\|^2 \tag{4.8}$$

при условии (4.7). Уравнение Эйлера для функционала (4.8)

$$\left(X'X + \alpha I_p\right)X_{\alpha}^{+} = X'. \tag{4.9}$$

Именно поэтому матрица X^{+} определяется [2, 17] как предел

$$X^{+} = \lim_{\alpha \to +0} X_{\alpha}^{+} = \lim_{\alpha \to +0}\left(X'X + \alpha I_p\right)^{-1} X', \tag{4.10}$$

что равносильно выполнению условия (4.7) для уравнения (4.9).

Аналогично (2.12) имеет место разложение в функциональный ряд по параметру:

$$X^{+} = \sum_{i=0}^{\infty} (-1)^i \frac{\alpha^i}{i!} \frac{\partial^i X_{\alpha}^{+}}{\partial \alpha^i}, \tag{4.11}$$

которое можно использовать для построения X^{+}. Кроме того, всякая регуляризованная оценка b_{α} имеет вид

$$b_{\alpha} = X_{\alpha}^{+}y \tag{4.12}$$

Для модели с активными ошибками выполняется такое же построение.

Для модели с пассивными ошибками (3.5) псевдообратная матрица Мура-Пенроуза будет являться решением следующей вариационной задачи

$$\tilde{X}^{+} = \arg \min_{\tilde{X}^{-} \left| \tilde{X}\tilde{X}^{-} - I_n \right| \leq n-p} \left| \tilde{X}^{-} \right|^2, \tag{4.13}$$

то оценка метода наименьших квадратов совпадает с нормальной и, как в (4.5),

$$\tilde{b}_0 = \tilde{X}^{+} y, \tag{4.14}$$

и, как в (2.12),

$$\tilde{b}_0 = \sum_{i=0}^{\infty} (-1)^i \frac{\alpha^i_{\sigma,\mu}}{i!} \frac{\partial^i \tilde{b}_{\alpha_{\sigma,\mu}}}{\partial \alpha^i_{\sigma,\mu}}, \tag{4.15}$$

т.е. обобщенная регуляризованная оценка (3.7) для модели пассивных ошибок матрицы (3.5) есть первый член разложения оценки $\tilde{b}_0$ в функциональный ряд.

Запишем регуляризованную псевдообратную матрицу $\tilde{X}^{+}_{\alpha}$ как решение уравнения Эйлера

$$\left(\tilde{X}'\tilde{X} + \alpha I_p \right) \tilde{X}^{+}_{\alpha} = \tilde{X}' \tag{4.16}$$

Тогда $\tilde{X}^{+}$ и $\tilde{X}^{+}_{\alpha}$ связаны соотношением

$$\tilde{X}^{+} = \sum_{i=0}^{\infty} (-1)^i \frac{\alpha^i}{i!} \frac{\partial^i \tilde{X}^{+}_{\alpha}}{\partial \alpha^i}. \tag{4.17}$$

При этом $E\tilde{b}_0 \neq \beta_0$, то есть смещена, также как и регуляризованная оценка $\tilde{b}_{\alpha_{\sigma,\mu}}$. Из общих результатов теории регуляризирующих алгоритмов следует сходимость по вероятности

$$\tilde{b}_{\alpha_{\sigma,\mu}} \xrightarrow{P} \beta_0,$$

если $\sigma, \mu \to 0$, *а* α *выбрано из (3.8).*

Отметим некоторые вычислительные аспекты метода регуляризации. Используя [22] разложение матрицы

$$X = QDR, \tag{4.18}$$

где Q, R - унитарные, D — верхняя двухдиагональная, вычисления регуляризованных оценок можно проводить с необходимой в данном случае более высокой точностью, экономией времени и памяти компьютера, особенно если применять метод квадратного корня для решения уравнений Эйлера с двухдиагональной матрицей, который хотя и менее экономичен в смысле ресурсов по отношению к методу прогонки, но более эффективен в отношении ошибок округления.

§ 5. Учет корреляций

Здесь мы будем исходить из следующего обобщения модели множественной линейной регрессии

$$y = X\beta + \varepsilon, E\varepsilon = 0, E(\varepsilon - E\varepsilon)(\varepsilon - E\varepsilon)' = \sigma^2 V, \tag{5.1}$$

где $\sigma^2 V$ — ковариационная матрица, отличная от диагональной, т.е. компоненты вектора ε коррелированы.

Если X — матрица полного ранга p, а V^{-1} — невырождена, то заменой

$$z = (U')^{-1} y,\ H = (U')^{-1} X,$$

где $U'U = V$, модель (5.1) сводится к (1.1).

Если $rankX = r < p$, а V^{-1} невырождена, то той же самой заменой модель (5.1) сводится к (2.1), и нормальная оценка псевдообращением b_0 записывается как аналог (4.5):

$$b_0 = \left((U')^{-1} X\right)^{+} (U')^{-1} y. \tag{5.2}$$

Если $rankX = r < p$ и V^{-1} — вырождена, то следуя [2, 23], введем параметрическую функцию ψ как

$$\psi = c'\beta, \tag{5.3}$$

где c — заданный вектор. Функция ψ оцениваема тогда и только тогда, когда

$$\left(X^{+}X - I\right)c = 0. \tag{5.4}$$

При таком выборе c

$$\hat{\psi} = \hat{a}'y,$$

где

$$\hat{a} = \left(I - \bar{U}^{+}U\right)\left(X^{+}\right)' c, \bar{U} = \left(I - XX^{+}\right), V = U^2. \tag{5.5}$$

Оценка $\hat{\psi}$ является наилучшей линейной оценкой для ψ, поскольку для любой другой линейной оценки ψ^* имеет место следующее неравенство

$$E(\psi^* - \psi)^2 = E(\hat{\psi} - \psi)^2.$$

Теорема (Гаусс-Марков) [2]. *Пусть*

$$\hat{b} = X^{+}\left(I - \bar{U}^{+}U\right)y = Gy. \tag{5.6}$$

Тогда имеем $E\hat{b} = X^{+}X\beta$, *и* $\hat{\psi} = c'\hat{b}$ *- наилучшая линейная несмещенная оценка для* ψ *(если* V *- не вырождена, то* $\hat{b} = b_0$ *из* (5.2)*).*

Пусть $k_1 = rankG$. Тогда статистика

$$F = \frac{n-r}{k_1} \frac{\left(G\left(\beta - \hat{b}\right)\right)' \left[G(X'X)^{+} G'\right]^{+} G\left(\beta - \hat{b}\right)}{\left(y - X\hat{b}\right)' \left(y - X\hat{b}\right)} \qquad (5.7)$$

имеет распределение Фишера-Снедекора с параметрами k_1 и $n-r$ в предположении нормальности ε.

§ 6. Связь с байесовской процедурой

Возвращаясь к модели (2.1), будем считать вектор неизвестных параметров β моментной характеристикой случайного вектора b:

$$Eb = \beta,\ E(b-\beta)(b-\beta)' = \sigma^2\alpha^{-1}I_p,\ \alpha > 0.$$

Тогда модель (2.1) запишется в виде

$$\begin{cases} y = X\beta + \varepsilon, E\varepsilon = 0, E\varepsilon\varepsilon' = \sigma^2 I_n, \\ b = \beta + \eta, E\eta = 0, E\eta\eta' = \sigma^2\alpha^{-1}I_p, \end{cases} \qquad (6.1)$$

где η — ошибка априорной информации; или, в матричной форме

$$z = H\beta + \xi,$$

$$z = \begin{bmatrix} y \\ b \end{bmatrix}, H = \begin{bmatrix} X \\ I_p \end{bmatrix}, \xi = \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}, \qquad (6.2)$$

т.е. мы находимся в рамках модели (5.1) с заведомо невырожденной ковариационной матрицей

$$E\xi\xi' = \sigma^2 V = \sigma^2 \begin{bmatrix} I_n & 0 \\ 0 & \alpha^{-1}I_p \end{bmatrix}.$$

Метод наименьших квадратов для (4.2) (байесовская процедура [24]) приводит к нормальному уравнению

$$H'H\beta = H'z, \qquad (6.3)$$

так что байесовская оценка $\hat{b}_0$ имеет вид

$$\hat{b}_0 = \left(X'X + \alpha I_p\right)^{-1}\left(X'y + \alpha b\right). \qquad (6.4)$$

Таким образом, оценка $\hat{b}_0$ при априорной информации $b \equiv 0$ идентична регуляризованной оценке b^{α} (2.8) и, следовательно, обладает всеми свойствами последней. В случае возмущенной матрицы $\tilde{X}$ обобщенные регуляризованные оценки $\tilde{b}_\alpha$ и $\tilde{b}_\rho$ также совпадают с байесовской $\hat{b}_0$. Однако при малых значениях α матрица V^{-1} становится плохообусловленной, и в этом случае для вычисления байесовской оценки $\hat{b}_0$ удобнее применять методику наилучших линейных несмещенных оценок (5.6), так как ранг V^{-1} известен и равен n.

Рассмотрим теперь байесовскую постановку с заданными ковариационными матрицами ошибок

$$\begin{cases} y = X\beta + \varepsilon, E\varepsilon = 0, E\varepsilon\varepsilon' = \sigma^2 V, \\ b^* = \beta + \eta, E\eta = 0, E\eta\eta' = \sigma^2 W, \end{cases} \tag{6.5}$$

которая сводится к модели (6.2) с $z' = (y, b^*)$ при известных

$$Eb^* = \beta,\ V = U^2,\ W = \alpha^{-1}\left[w_{ij}\right],\ \sigma^2.$$

Нормальное уравнение для такой модели

$$\left(X'V^{-1}X + W^{-1}\right)\beta = X'V^{-1}y + W^{-1}b^* \tag{6.6}$$

идентично уравнению Эйлера метода регуляризации в вариационной постановке:

$$\min\left(\beta - b^*\right)' W^{-1}\left(\beta - b^*\right) \tag{6.7}$$

при условии

$$\left(y - X\beta\right)' V^{-1}\left(y - X\beta\right) = n\sigma^2. \tag{6.8}$$

С использованием аппарата псевдообратных матриц решение задачи (6.7), (6.8) записывается [25] в виде

$$\bar{b}^* = \left[I_p - \left(PW^{-1}P\right)^+ PW^{-1}\right]\left(X'V^{-1}X\right)^+\left(X'V^{-1}y + W^{-1}\beta^*\right),$$

где обозначено

$$P = I_p - \left(UX\right)^+ UX.$$

Здесь, также как и в предыдущих параграфах, возможно исследование различных аспектов задания матриц X ,V и W.

Рассмотренные в работе постановки задач, так или иначе, приводят к нормальным оценкам вектора параметров β или в условиях более полной информации — к близким к ним регуляризованным оценкам. Псевдообращение, являясь частью регуляризирующих процедур, позволяет записать оценки b в универсальной компактной форме, что привычнее для описания их статистических свойств.

Литература

1. Рао С.Р. Линейные статистические методы и их применение. М.: Наука, 1968, с. 547.
2. Albert A. Regression and the Moore-Penrouse Pseudoinverse. Academic Press, New York-London. 1972 (Русский перевод: Липцер Р.Ш.: Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. М.: Мир. 1975).
3. Тихонов А.Н. О некорректных задачах линейной алгебры и устойчивом методе их решения. - ДАН СССР, 1965, т. 162, № 4, с. 763-766.
4. Тихонов А.Н. Об устойчивости алгоритмов для решения вырожденных систем линейных алгебраических уравнений. - ЖВМ и МФ, 1965, т. 5, №. 4, с. 718-722.
5. Hoerl A.E., Kennard R.W. Ridge regression: biased estimation for non-orthogonal problems // Technometrics. V. 12. N 1. P. 55-67. 1970.
6. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. - М.: Наука, 1979, с. 288.
7. Лаврентьев М.М. О некоторых некорректных задачах математической физики. - Новосибирск: СО АН СССР, 1962, с. 92.
8. Морозов В.А. Регулярные методы решения некорректно поставленных задач. - М.: Изд-во МГУ, 1974.
9. Иванов В.К. О линейных некорректных задачах. - ДАН СССР, 1962. т. 145, № 2, с. 270-272.
10. Иванов В.К., Васин В.В., Танана В.П. Теория линейных некорректных задач и ее приложения. - М.: Наука, 1978, с. 205.
11. Hoerl A.E. Application of Ridge analysis to regression problems // Chem. Eng. Progr. N 58. P. 54-59. 1962.
12. Меченов А.С. О разложении решения линейного уравнения 1-го рода в функциональный ряд // Вычислительные методы и программирование. XXI, М.: Изд-во МГУ, 1973, с. 164-169.
13. Кендалл М. Дж. Стюарт А. Статистические выводы и связи. - М.: Наука. 1973, с. 898.
14. Гончарский А.В., Ягола А.Г. О равномерном приближении монотонных решений некорректных задач. - ДАН СССР, 1969, т. 184, № 4, с. 771-773.
15. Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. Регуляризирующие алгоритмы и априорная информация. - М.: Наука, 1983, с. 200.
16. Peters G., Wilkinson J.H. (1970) The least squares problem and pseudo-inversion // Comput. J. N13. P. 309-316

17. Жуковский Е.Л., Липцер Р.Ш. О вычислении псевдообратных матриц. - ЖВМ и МФ, 1975, т. 15, и 2, с. 489-492.
18. Меченов А.С. О подходе максимального правдоподобия к оценке параметров линейных функциональных соотношений. //Численные методы в математической физике. Московский университет. 1996. С. 153-159.
19. Тихонов А.Н. Об одном принципе взаимности. - ДАН СССР, 1980, т. 253, № 2, с. 302-308.
20. Тихонов А.Н. О нормальных решениях приближенных систем линейных алгебраических уравнений. - ДАН СССР, 1980, т. 254, № 3, с. 549-554.
21. Тихонов А.Н. О приближенных системах линейных алгебраических уравнений. - ЖВМ и МФ, 1980, т. 20, № 6, с. 1373-1383.
22. Воеводин В.В. О методе регуляризации. - ЖВМ и МФ, 1969, т.9, N3. с. 671-673.
23. Шеффе Г. Дисперсионный анализ. - М.: Наука, 1980, с. 512.
24. Jeffreys H. Theory of Probability. -London, Oxford University Press. 1961.
25. Price C.M. The matrix pseudoinverse and minimal variance estimate // SIAM Rev., V 6. P. 115-120. 1964.
26. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980, с. 456.
27. Федотов А.М. Линейные некорректные задачи со случайными ошибками в данных. - Новосибирск: Наука, Сиб. отд., 1982, с. 189.
28. Меченов А.С. О частично приближенных системах линейных алгебраических уравнений // Ж. вычисл. матем. и матем. физ., т. 31, N6, 1991, с. 790-799.
29. Меченов А.С. Регуляризованный метод наименьших квадратов. М.: Изд-во Моск. ун-та, 1988. 96 с.