сообщения
объединенного
института
ядерных
исследований
дубна

Р10-86-502

В.Б.Злоказов

О ДЕТЕРМИНИСТСКОЙ ИНТЕРПРЕТАЦИИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

1986

Введение

Метод наименьших квадратов (м.н.к.) может иметь и детерминистскую интерпретацию: погрешность $e(x_j)$ наблюдаемой кривой $y(x_j)$ мы можем считать дискретной функцией с нормой в каждом x_j, равной максимально наблюдаемому по модулю значению, и все рассуждения вести не в терминах функций распределений, ковариаций и дисперсий, а в терминах функциональных пространств и определенных в них норм.

Формализм математической статистики оправдан лишь тогда, когда есть вероятностное описание механизма образования погрешностей и есть возможность и необходимость интерпретировать результаты м.н.к.-анализа именно в статистическом смысле. Рутинные предположения о нормальном распределении погрешностей, об отсутствии корреляций между ними и т.д. в действительности являются волюнтаристской уловкой, призванной скрыть отсутствие подлинной информации об этих погрешностях. Можно предположить, что и само мышление экспериментаторов, несмотря на широкое использование статистической терминологии, имеет часто детерминистский характер: дисперсии входных и выходных данных интерпретируются чаще всего в смысле норм; такая важная характеристика данных, резко отличающая статистический формализм от детерминистского, как корреляции, не используется ни на входе в м.н.к., ни на выходе из него. Поскольку, однако, конфликтов на практике между статистической лексикой и детерминистскими по сути действиями не наблюдается, это наводит на мысль о существовании глубокой аналогии между статистической процедурой м.н.к. и детерминистской. Цель этой работы состоит в том, чтобы показать, что такое единство действительно существует и результаты м.н.к.-анализа не зависят от характера используемой в нем терминологии. При этом детерминистский м.н.к. использует более адекватный практике язык и заключает в себе меньше субъективизма, поскольку информацию о нормах погрешностей получить значительно легче и проще, чем о полной функции их распределения.

Дополнительной целью работы было исследование различных способов и форм накопления данными информации, - вопрос, часто игнорируемый в математических теориях анализа данных. В частности, информационная емкость данных экспериментальной физики определяется не так, как это принято в классическом регрессионном анализе.

§ I. Линейный детерминистский м.н.к.

В дальнейшем большие буквы с индексами будут обозначать матрицы и векторы, маленькие – их элементы, а $\hat{P}_i$ – м.н.к.-оценки вектора параметров P_i.

Итак, пусть задан Y_j, $j=1,2,\dots,m$ – наблюдаемые данные, модель которых имеет следующий вид:

$$Y_j=F_{ji}P_i + E_j \quad , \quad i=1,2,\dots,n \quad , \qquad (I.I)$$

где F_{ji} – линейная модель, P_i – вектор параметров, E_j – погрешности, норма которых в каждой точке j известна и равна максимальному (по модулю) значению. Нормировкой величин Y, F и E на $\|E_j\|$ можно добиться того, что норма погрешности данных в каждой точке будет равна единице. Далее, будем считать, что F_{ji} имеет ранг n.

Теперь м.н.к.-оценки параметров минимизируют выражение

$$(Y_j-F_{ji}P_i)'(Y_j-F_{ji}P_i) \qquad (I.2)$$

в пространстве параметров R^n. Нормальное уравнение

$$F'_{ji}Y_j = F'_{ji}F_{ji}P_i$$

дает нам эти м.н.к.-оценки

$$\hat{P}_i = (F'_{ji}F_{ji})^{-1}F'_{ji}Y_j = B_{ij}Y_j \quad , \qquad (I.3)$$

Сформулируем теперь детерминистский аналог несмещенной оценки.

Определение I. Оценка $\hat{P}_i=D_{ij}Y_j$ называется несмещенной, если ее погрешность является линейной функцией только погрешностей E_j:

$$\hat{P}_i-P_i= \Phi_{ij}E_j$$

Если $D_{ij}Y_j$ - несмещенная оценка, то в силу (I.I) имеем

$$\hat{P}_i=D_{ij}F_{ji}P_i + D_{ij}E_j .$$

Чтобы при произвольных E_j оценки $\hat{P}_i$ отличались от P_i только на $D_{ij}E_j$, необходимо и достаточно, чтобы имело место

$$D_{ij}\ F_{ji} = I_{ij} . \qquad (I.4)$$

Пользуясь этим, можно показать, что м.н.к.-оценка является несмещенной:

$$\hat{P}_i=B_{ij}Y_j=B_{ij}F_{ji}P_i+B_{ij}E_j=P_i+B_{ij}E_j \ . \qquad (I.5\ )$$

Однако, если в м.н.к. используется модель F_{ji}, тогда как адекватной данным является модель G_{ji}, м.н.к.-оценки будут смещены:

$$\hat{P}_o=(F'F)^{-1}F'(GP+E)=(F'F)^{-1}F'(FP+RP+BE)=P+RP+BE \quad ,$$

где $R_{ji}=G_{ji}-F_{ji}$. Член $R_{ji}P_i$ будет смещением, или систематической ошибкой.

Рассмотрим теперь вопрос о детерминистских погрешностях м.н.к.-оценок.

Определение 2. Погрешностью величины $\hat{P}_i$ назовем $\max\limits_i|\hat{P}_i-P_i|$ при условии, что $(E'_j,E_j)\le 1$.

Такое определение вполне соответствует правилам математики, а то, что взят шар, а не куб $|E_j|\le 1$, согласуется с подходом статистики, которая ограничивает область изменений погрешностей данных доверительным шаром, и пренебрегает такими маловероятными комбинациями E_j, как,например, наихудший расклад $E_j=\pm 1$.

Теорема I. Квадрат погрешности м.н.к.-оценки $\hat{P}_i$ равен i-му диагональному элементу обратной м.н.к-матрицы.

Доказательство. Используя (I.5), имеем

$$(\hat{P}_i-P_i)(\hat{P}_i-P_i)'=B_{ij}E_jE'_jB'_{ij} \ .$$

Возьмем $(\hat{p}_i-p_i)$. Для нее справедливо

$$(\hat{p}_i-p_i)^2 = (\sum_{j=1}^{m} b_{ij}\ e_j)^2 .$$

Решим следующую задачу на условный экстремум: найти

$$\max_{\{e_j\}}\ (\sum_{j=1}^{m} b_{ij}e_j)^2$$

при условии $\sum\limits_{j=1}^{m} e_j^2 = 1$.

Здесь поставлен знак равенства, так как выпуклая функция достигает максимума на границе области.

Составив выражение Лагранжа с множителем λ и взяв от него частные производные по e_j, получим

$$zb_{ik} = -\lambda e_k \ , \quad k=1,\dots,m, \quad z=\sum_{j=1}^{m} b_{ij}e_j \ . \qquad (I.6)$$

Пусть e_k в z те самые, которые обращают z^2 в максимум. Возведя обе части (I.6) в квадрат, сложив по k и воспользовавшись граничным условием, получим

$$z^2= \lambda^2/(\sum_{k=1}^{m} b_{ik}^2) \ . \qquad (I.7)$$

С другой стороны, умножив обе части (I.6) на b_{ik} и сложив, получим

$$z\sum_{k=1}^{m} b_{ik}^2 = -\lambda\sum_{k=1}^{m} b_{ik}e_k = -\lambda z .$$

Откуда $\lambda = -\sum_{k=1}^{m} b_{ik}^2$.

Подставив λ в (I.7), получим

$$z^2 = \sum_{k=1}^{m} b_{ik}^2 .$$

Это и есть максимум выражения $\left(\sum_{j=1}^{m} b_{ij}e_j\right)^2$ при условии $\sum_{j=1}^{m} e_j^2 = 1$. Элемент $\sum_{k=1}^{m} b_{ik}^2$ есть не что иное, как i-й диагональный элемент матрицы $B_{ij}B'_{ij}$. Но

$$B_{ij}B'_{ij} = (F'F)^{-1}F'F(F'F)^{-1} = (F'_{ij}F_{ji})^{-1} ,$$

что и требовалось доказать.

В качестве суммарной погрешности м.н.к.-оценок можно взять евклидову норму вектора $\hat{P}_i - P_i$. Поскольку она равна максимальному собственному значению матрицы $(F'_{ji}F_{ji})^{-1}$ и, следовательно, строго меньше ее следа, то на сфере $\sum_{j=1}^{m} e_j^2 = 1$ наихудший расклад м.н.к.-оценок не достигается.

Возникает вопрос: как относятся погрешности детерминистских м.н.к.-оценок к погрешностям других линейных несмещенных оценок? Ответ дает детерминистский аналог теоремы Гаусса – Маркова.

<u>Теорема 2</u>. В классе линейных несмещенных оценок погрешности м.н.к.-оценок минимальны.

<u>Доказательство</u>. Возьмем произвольную линейную оценку

$$\hat{P}_i^1 = A_{ij}Y_j .$$

Из ее несмещенности следует $A_{ij}F_{jk} = I_{ik}$, $i,k=1,\dots,n$, что означает, что ранг матрицы A должен быть равен n. Как и для м.н.к.-оценок, можно показать, что квадраты погрешностей оценок $\hat{P}^1$ будут равны диагональным элементам матрицы $(A_{ij}A'_{ij})$. Задача, следовательно, состоит в том, чтобы показать, что у матрицы $C_{ij} = A_{ij}A'_{ij} - (F'_{ji}F_{ji})^{-1}$ все диагональные элементы неотрицательны. Имеем

$$C = AA' - AF(F'F)^{-1}F'A' = A(I - F(F'F)^{-1}F')A' .$$

Достаточно показать неотрицательную определенность матрицы $Q = I - F(F'F)^{-1}F'$.

Имеет место:

1) $Q' = I - (F(F'F)^{-1}F')' = I - F(F'F)^{-1}F' = Q$,

2) $QQ = I - 2(F(F'F)^{-1}F') + (F(F'F)^{-1}F')(F(F'F)^{-1}F') = I - F(F'F)^{-1}F' = Q$.

Отсюда для произвольного X имеем

$$(QX,X) = (Q^2X,X) = (QX,Q'X) = (QX,QX) \geqslant 0 .$$

Если бы у матрицы C i-ый диагональный элемент имел бы знак минус, то на векторе E таком, что $e_j=0$ при $j \neq i$, и $e_i=1$, мы имели бы: $(CE,E) = c_{ii} < 0$, что противоречило бы неотрицательной определенности C. Это завершает доказательство.

Следующий вопрос – какие значения может принимать минимизируемый функционал (I.2) при подстановке м.н.к.-оценок вместо параметров, при условии, что модель FP действительно адекватна данным Y. Эти значения нужны нам для решения обратной задачи: убедиться, что модель FP адекватна данным Y даже при самом плохом раскладе погрешностей исходных данных.

<u>Теорема 3</u>. Если модель FP адекватна данным Y, и ошибки $|e_j| \leqslant 1$, $j=1,\dots,m$, то при $P_i = \hat{P}_i$ значение (I.2) заключено между нулем и m.

<u>Доказательство</u>. Нуль, очевидно, будет нижней гранью (I.2), достигаемой, например, при $E_j=0$. Определим максимум (I.2). Имеем

$$(Y-F\hat{P})'(Y-F\hat{P}) = (Y-FBY)'(Y-FBY) = (E-FBE)'(E-FBE) .$$

Учитывая, что $(FB)'=FB$ и $(FB)'(FB)=FB$, получаем

$$\chi^2 = (E-FBE)'(E-FBE) = (E',E) - (E',FBE) . \qquad (I.8)$$

Так как χ^2 - выпуклая функция, то она достигает максимума на границе, причем в точке, лежащей на граничной поверхности так, что через нее нельзя провести прямую, на которой эта точка была бы внутренней. Следовательно, точка максимума может иметь координаты только $e_j = \pm 1$.

Отсюда $\max\limits_{\{e_j\}} \chi^2 = m - \min\limits_{\{e_j\}=1} (E',FBE) \leqslant m$.

Более точная верхняя граница будет зависеть от модели FP. Например, если $FP=c$, c-const и m – четное число, то точной верхней границей значений (I.2) будет m. Действительно, если минимизируется $\sum_{j=1}^{m}(y_j-c)^2$ и $y_j = c+(-1)^j$, то м.н.к.-оценка $\hat{c}$ равна c и $\chi^2 = \sum_{j=1}^{m} 1 = m$. С другой стороны, пусть FB имеет диагональную структуру. Так как ранг F равен n, то равенства

$$\{FBY = \lambda Y\} \equiv \{F(F'F)^{-1}F'Y = \lambda Y\} \equiv \{F'Y = \lambda F'Y\}$$

возможны лишь при $\lambda = I$ и $\lambda = 0$. Следовательно, $F_{ji}B_{ij}$ имеет собственное значение I кратности n и 0 кратности $m-n$. Отсюда $\min\limits_{|e_j|=1}(E'FBE)=n$, и для такой модели $\max\chi^2=m-n$. Суммируя сказанное, детерминистский критерий проверки гипотез можно сформулировать так: модель FP не противоречит гипотезе о ее адекватности данным Y, если χ^2 при $P_i=\hat{P}_i$ и $|e_j|\leq 1$ не превосходит существенно число m.

Теорема 4. Если нормы погрешностей в каждой точке неизвестны, но одинаковы, то они не меньше величины $\sqrt{\chi^2/m}$ при $P_i=\hat{P}_i$, сосчитанных для единичных весов в (I.2).

Доказательство очевидно.

§ 2. Информационная емкость данных и состоятельность м.н.к.-оценок

В классическом регрессионном анализе наблюдения Y_j, $j=1,\dots,m$ ранжируются в ряд по нарастанию аргумента j, а число m в случае, например, стационарных случайных погрешностей служит информационной мерой данных, поскольку в силу эргодических свойств стационарных процессов информационная емкость траекторий таких процессов пропорциональна их длинам /I/.

С информационной емкостью данных тесно связан вопрос о состоятельности оценок, т.е. о их стремлении к точным значениям параметров при росте количества информации в данных.

Для классических регрессий условия состоятельности статистических м.н.к.-оценок /2/ переносятся на детерминистский случай без изменений.

Теорема 5. Пусть λ_k- собственные числа матрицы $F'F$. Если $\min\limits_k \lambda_k \to \infty$ при $m\to\infty$, то $\|\hat{P}-P\| \to 0$.

Доказательство следует из цепочки соотношений:

$$\|\hat{P}-P\| \leq \sum_{i=1}^{n} \max_{\{e_j\}} |\hat{p}_i - p_i| = Sp(F'F)^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} .$$

Для регрессий указанного типа м.н.к.-оценки могут быть состоятельными, а могут и не быть.

Пусть $y_j=aj+e_j$, $z_j=a\exp(-(j-t)^2)+e_j$, $j=1,\dots,m$ и $|e_j|\leq 1$. Для первой регрессии $\lambda_1=\sum\limits_{j=1}^{m} j^2$, для второй $\lambda_2=\sum\limits_{j=1}^{m}\exp(-2(j-1)^2)$. При $m\to\infty$, $\lambda_1\to\infty$, λ_2 стремится к конечному пределу. Отсюда, м.н.к.-оценка параметра a в первом случае является состоятельной, во втором – нет.

Современная экспериментальная физика изобилует примерами наблюдаемых кривых, информационная емкость которых характеризуется не столько количеством точек наблюдения, обычно определяемым разрешающей способностью регистрирующей аппаратуры и диапазоном наблюдаемых эффектов, сколько т.н. статистикой отсчетов.

Попытаемся построить обобщенное изложение теории таких регрессий. Пусть непрерывной моделью какого-либо процесса является неотрицательная функция $f(x)$. Физические наблюдения строятся так, что суммы т.н. "событий" с координатами x_j являются (с какой-то погрешностью) дискретным приближением функции $f(x)$. Можно считать, что в каждой точке x_j из N_j возможностей произойти или не произойти $f(x_j)$"событиям" реализуется доля шансов, пропорциональная N_j, т.е.

$$f(x)=k(x)N(x) . \qquad (2.I)$$

Такой формализм обобщает ряд реальных распределений вероятностей процессов физических реакций, таких, как пуассоновское, биномиальное и др. Число $N(x)$, а точнее его дискретный аналог $N(x_j)$, мы и будем называть статистикой.

В качестве моделей погрешностей $E(x_j)$ можно предложить ряд семейств функций от $N(x_j)$, обобщающих различные реальные случаи. В случае регрессий 2-го типа масштаб значений параметров в общем случае будет зависеть от статистики. Так же и погрешность в общем случае будет расти с ростом статистики. Поэтому в общем случае норма погрешности $e(x_j)$ будет иметь вид

$$\|e(x_j)\| = c(x_j)N^{\beta}(x_j) , \text{ где } \beta \geq 0 . \qquad (2.2)$$

На регрессии данного типа вся предыдущая теория м.н.к. переносится без изменений, однако состоятельность м.н.к.-оценок для таких регрессий логично определить иначе: м.н.к.-оценка состоятельна, если при $\min\limits_j N_j\to\infty$ стремятся к нулю $\|\hat{p}_i-p_i\| / \|p_i\|$.

Стремление к нулю относительной (а не абсолютной) погрешности оценки параметров введено потому, что просто $\|\hat{p}_i-p_i\|\to 0$ при росте статистики может и не иметь места в силу (2.I). Но если масштаб параметра p_k не зависит от статистики, то из $\frac{\|\hat{p}_k-p_k\|}{\|p_k\|}\to 0$ следует $\|\hat{p}_k-p_k\|\to 0$, так что данное определение состоятельности является достаточно универсальным.

Итак, пусть $Y_j=F_{ji}P_i+E_j$, причем $F_{ji}P_i \sim k_jN_j$, $\|e_j\| \sim c_jN_j^{\beta}$.

Теорема 6. М.н.к.-оценка $\hat{p}_i$ состоятельна, если $\max\limits_j\|e_j\|$ растет строго медленнее, чем $\min\limits_j N_j$ при $\min\limits_j N_j\to\infty$.

Доказательство следует из таких неравенств:

$$\left\| \frac{\hat{p}_k - p_k}{p_k} \right\| \frac{\sum_i \left\{ \left(\sum_j \frac{1}{\|e_j\|^2} F'_{ji} F_{jk} \right)^{-1} \sum_j \frac{1}{\|e_j\|^2} e_j F'_{ji} \right\}}{\sum_i \left\{ \left(\sum_j \frac{1}{\|e_j\|^2} F'_{ji} F_{jk} \right)^{-1} \sum_j \frac{1}{\|e_j\|^2} k_j N_j F_{ji} \right\}} = \frac{\sum_j a_{kj} e_j}{\sum_j a_{kj} k_j N} \leqslant$$

$$\leqslant \frac{\sum_j |a_{kj}| \; \|e_j\|}{\sum_j |a_{kj}| \; k_j N_j} \leqslant \frac{C \max_j \|e_j\|}{\min_j k_j N_j} .$$

Теорема 7. Условия $N_j / N_k \leqslant \mathrm{const}$ при любых j и k и $\beta < 1$ являются достаточными для состоятельности м.н.к.-оценок параметров.

Доказательство. Пусть $\max_j \|e_j\| = \|e_m\|$. Имеем

$$\left\| \frac{\hat{p}_k - p_k}{p_k} \right\| \leqslant \frac{C \max_j \|e_j\|}{\min_j k_j N_j} = \frac{C\, k_m N_m}{\min_j k_j N_j} \cdot \frac{C_m N_m^{\beta}}{k_m N_m} .$$

Если $\min_j N_j \to \infty$, то $N_m' \to \infty$ и тогда

$$\lim_{N_k \to \infty} \left\| \frac{\hat{p}_k - p_k}{p_k} \right\| = \lim_{N_m \to \infty} \frac{C_1 N_m^{\beta}}{N_m} = 0 .$$

Примечание I. Пусть линия физической регрессии представляет собой сумму функции пика и функции постоянного фона. Если уровни пика и фона пропорциональны сечениям соответствующих реакций, то отношение их амплитуд ограничено одной и той же константой для любой статистики. Если при этом события распределены, например, по Пуассону или γ-функции, то м.н.к.-оценки параметров при любой невырожденной параметризации состоятельны, так как в этом случае $\beta = I/2$. Контрпримером служит спектр ядерной реакции, в котором события распределены по компаунд-пуассоновскому закону, так как в этом случае $\beta = I$. Такие данные не способны накапливать информацию.

Примечание 2. Если выполнены условия теоремы 7, то асимптотика относительной погрешности оценок параметров имеет порядок $N^{\beta-1}$. Для $\beta = I/2$, это дает $N^{-\frac{1}{2}}$, что можно считать эффективной асимптотикой.

§ 3. Нелинейный м.н.к.

Нелинейность параметризации моделей, как правило, связана с их реалистичностью. Она в общем случае неустранима, широко распространена и перспективна. Однако она влечет за собой многочисленные математические усложнения, которые мы здесь попытаемся проанализировать. Итак, пусть справедливо представление

$$y(x) = f(x, P_i) + e(x) ,$$

и f зависит от параметров P_i (или от части их) нелинейно. Составим выражение

$$F(e(x), P_i) = \sum_{j=1}^{m} w(x_j) \left\{ y(x_j) - f(x_j, P_i) \right\}^2, \text{ где } w(x) = 1/\|e(x)\|^2 . \quad (3.I)$$

Значение $\hat{P}_i$, обращающее (3.I) в минимум, и будет нелинейной м.н.к.–оценкой.

Первой сложностью нелинейного м.н.к. является то, что (3.I) может иметь бесконечно много как локальных, так и глобальных минимумов, в которых значения параметров отличаются от истинных сколь угодно сильно. Но даже если этот минимум является единственным, он может не достигаться в конечных P_i. Пусть, например, $y_j = \exp(-j) + e_j$, $\|e_j\| = \exp(-j)$, а в качестве модели выбрана $\exp(-pj)$. Если все $y_j = 0$, то $F((e(x), p)) = \sum_{j=1}^{m} \exp(-2pj + 2j)$. Минимум (а точнее, нижняя грань) этого выражения равен нулю и достигается при $p = +\infty$. При этом

$$\frac{d^2 F}{\partial p^2} = \sum_j 4j^2 \exp(-2pj + 2j) > 0 \qquad \text{для любых } p .$$

Выясним, какие условия гарантируют существование единственной конечной нелинейной м.н.к.–оценки.

Выражение (3.I) определено на $E^m \otimes R^n$, где E^m – пространство векторов погрешностей E_j, R^n – пространство векторов параметров P_i. Мы можем также трактовать его как величину, определенную на R^n и зависящую от обобщенного параметра E_j, причем так, что при $\|E_j\| \to 0$ (3.I) стремится к выражению, которое при корректной параметризации является строго выпуклой функцией в некоторой окрестности истинных значений параметров P_{oi}, и имеет в этой окрестности единственный минимум. Действительно, формальное условие минимума имеет вид

$$\sum_j \frac{1}{\|e_j\|^2} \left\{ f_{jo} - f_j + e_j \right\} f'_{jk} = 0 \qquad k = 1, \ldots, n , \qquad (3.2)$$

где $f_{jo} = f(x_j, P_{oi})$, $f'_{jk} = \dfrac{\partial f(x_j, P_i)}{\partial P_k}$. При $\max_j |e_j| \to 0$ и произвольных f'_{jk} равенство возможно лишь, если $f_j \to f_{jo}$, т.е. точки минимума $P_i \to P_{oi}$. Отсюда условия, гарантирующие существование для данного класса погрешностей в некоторой окрестности вектора P_o единственной нелинейной м.н.к.–оценки, формулируется следующая теорема.

Теорема 8. Если параметризация модели $f(x)$ такова, что:

I) выражение (3.I) допускает представление

$$F(e(x), P) = F(e(x), P_o) + \left\{ F'(e(x), P_o), \Delta P \right\} + \frac{1}{2} \left\{ F''(e(x), P_o) \Delta P, \Delta P \right\} + O(\|\Delta P\|^2)$$

в прямом произведении некоторых малых окрестностей $e_o(x) \equiv 0$ и P_o;

2) оператор $F''(e(x),P_o)^{-1}$ в этих окрестностях ограничен, то нелинейная м.н.к.-оценка является единственной в окрестности P_o, конечной, и асимптотически несмещенной.

Доказательство. Пусть $\|e(x)\|$ достаточно мала и выполнены условия теоремы 8. Уравнение стационарной точки

$$F'(e(x_j),P_o)+\{F''(e(x_j),P_o),\Delta P\}+O(\|\Delta P\|)=0 .$$

В силу невырожденности матрицы F'' оно имеет единственный корень P^{*}, который с точностью до малых поправок будет равен

$$P^{*}=P_o-(F''_o)^{-1}F'_o \qquad \text{, где}$$

$$(F''_o)^{-1}=\Big(\sum_j\frac{1}{\|e_j\|^2}\{f'_{ojk}f'_{oji}+e_jf''_{ojki}\}\Big)^{-1}, \quad F'_o=\sum_j\frac{1}{\|e_j\|^2}e_jf_{oji} ,$$

$$f'_{ojk}=\frac{\partial f(x_j,P_o)}{\partial P_k}, \quad f''_{ojki}=\frac{\partial^2 f(x_j,P_o)}{\partial P_k\partial P_i} .$$

Точка минимума P^{*} может быть таковой только, если матрица $F''^{-1}(e(x),P_o)$ положительно полуопределена в окрестности P_o. Покажем, что при $\max_j|e_j|\to 0$ это будет иметь место. Имеем

$$F''^{-1}_o=(A+B)^{-1} \qquad \text{, где}$$

$$A=\sum_j\frac{1}{\|e_j\|^2}f'_{ojk}f'_{oji}, \qquad B=\sum_j\frac{1}{\|e_j\|^2}e_jf''_{ojki} .$$

Пусть $U(e_j)=A^{-1}B$. Легко видеть, что если диаметр окрестности погрешностей $e_o(x)$ уменьшить в n раз, то норма U уменьшится не менее, чем n раз. Это следует из неравенств:

$$\|U(e_j)\|\le\|A^{-1}B\|\le\|A^{-1}\|\cdot\|B\| ,$$

$$\|U(e_j/n)\|\le\frac{1}{n^2}\Big\|\sum_j\frac{1}{\|e_j\|^2}f'_{ojn}f'_{oji}\Big\|^{-1}\frac{n^2}{n}\Big\|\sum_j\frac{1}{\|e_j\|^2}e_jf''_{ojki}\Big\|=\frac{1}{n}\|A^{-1}\|\cdot\|B\| .$$

Выберем окрестность $e_o(x)$ достаточно малого диаметра, на которой $\|U\|<1$. Имеем

$$(A+B)^{-1}=(A(I+A^{-1}B))^{-1}=\Big(\sum_{n=0}^{\infty}(-1)^nU^n\Big)A^{-1}=A^{-1}+\sum_{n=1}^{\infty}(-1)^nU^nA^{-1}=A^{-1}+R .$$

Для нормы второго члена справедлива оценка

$$\|R\|\le\Big\|\sum_n(-1)^nU^n\Big\|\cdot\|A^{-1}\|=\frac{\|A^{-1}\|\cdot\|U\|}{1+\|U\|} .$$

Для малых $\{e_j\}$ или малых норм $\|B\|$ она мала, откуда

$$F''^{-1}_o\approx A^{-1}=\sum_j\frac{1}{\|e_j\|^2}f'_{ojn}f'_{oji} ,$$

где A^{-1} положительно определена.

Используя сказанное, можно записать

$$P^{*}-P_o\approx A^{-1}F'_o+RF'_o ,$$

где 2-й член будет нелинейной функцией помехи $\{e_j\}$, т.е. нелинейная м.н.к.-оценка смещена. Но тем же путем, что и выше, можно показать, что при $\max_j e_j\to 0$ $\|RF'_o\|\to 0$, что означает асимптотическую несмещенность нелинейной м.н.к.-оценки. Теорема доказана.

Доказанная теорема обнаруживает аналогию между линейным и нелинейным оцениванием в малой окрестности нулевой погрешности и истинных значений параметров. Поэтому на нелинейный случай без изменений переносятся такие две теоремы.

Теорема 9. Если нелинейная модель $f(x,P)$ адекватна данным $y(x)$, то значения функционала (3.I) при $P=\hat P$ заключены между 0 и m.

Теорема IO. Нелинейная м.н.к.-оценка $\hat P$ состоятельна, если $\max_j\|e_j\|$ растет строго медленнее, чем $\min_i N_j$ при $\min_j N_j\to\infty$.

Доказательство. Используют асимптотическую несмещенность нелинейной м.н.к.-оценки, и в остальном повторяют рассуждения линейного случая.

Заключение

Изложенные результаты показывают, что между статистическим м.н.к. и детерминистским действительно существует глубокая аналогия, а точнее сходство с точностью до деталей, и это объясняет, почему результаты вычислений при обработке данных как правило не зависят от терминологии и лексики, которые используются в этих вычислениях. В то же время детерминистский м.н.к. лучше излагает ряд некоторых аспектов оценивания параметров, таких, как, например, соотношения между погрешностями исходных данных и погрешностями оценок параметров, а понятие нормы погрешности несет в себе меньше субъективизма и волюнтаризма, чем понятие распределения погрешностей.

Литература

I. Гнеденко Б.В. Курс теории вероятностей. "Наука", М., 1965.
2. Климов Г.П. Теория вероятностей и математическая статистика. Изд-во МГУ, М., 1983.

Рукопись поступила в издательский отдел
8 сентября 1986 года.

Злоказов В.Б. Р10-86-502

О детерминистской интерпретации метода наименьших квадратов

Модель экспериментальных данных имеет обычно вид: $f(x) = m(x) + e(x)$, где $m(x)$ - истинная функция, а $e(x)$ - случайная помеха, понимаемая в смысле математической статистики. Альтернативный подход, названный в работе детерминистским, состоит в том, что $e(x)$ - случайно выбранный элемент нормированного функционального пространства с известной нормой. Этот подход не требует таких обычных субъективистских предположений о $e(x)$, как нормальность, некоррелированность и т.д., да и норму функции по выборке оценить легче, чем полную функцию распределения случайной величины. Показывается, что такой м.н.к. сохраняет полную аналогию со статистическим, в частности, сохраняют силу теорема о норме погрешности оценок, теорема Гаусса - Маркова и другие, чем и объясняется то, что на практике статистическая терминология и детерминистские по сути действия не вступают в конфликт. Также рассмотрены нестатистические подходы к трактовке понятия "информация" в физических данных.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1986

Перевод О.С.Виноградовой

Zlokazov V.B. P10-86-502

On the Deterministic Interpretation of the Least Square Method

The model of the experimental data looks normally like this: $f(x) = m(x) + e(x)$ where $m(x)$ is the true function, and $e(x)$ is a random error, interpreted from the point of view of mathematical statistics. An alternative approach, called here deterministic, consists in $e(x)$ being a randomly taken element of some normed functional space with a known norm. This approach does not imply such conventional subjective assumptions about $e(x)$ as its normality, uncorrelatedness etc., besides that the norm of an element rather than the full distribution function of an random quantity is easier to be evaluated from observations. It is shown that such L.S.M. is quite compatible with the statistical one, namely, there are still holding theorem about errors of estimates, Gauss - Markov theorem etc., which explains the absence of the controversy between statistical terminology and in fact deterministic actions in practice. Non-statistical treatment of the notion "information" of the physical data is also considered.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1986