

Глава 6 *Метод наименьших квадратов (МНК)*

22. Средняя квадратическая аппроксимация функций

Научно-исследовательская практика часто нуждается в замене табличных функций, полученных экспериментально или из астрономических наблюдений, аналитическим выражением. Именно так решаются задачи интерполирования, численного дифференцирования и интегрирования. Однако, наличие погрешностей в этих данных не позволяет подобрать аналитические выражения, дающие график строго по экспериментальным точкам (например, при “точечном интерполировании” или при построении интерполяционных сплайнов). Уменьшить влияние этих погрешностей можно, если пользоваться для аппроксимации (приближения) наблюдаемых зависимостей одной величины от другой более гладкой кривой, чем это дает опыт.

Допустим, что реально существует зависимость переменной Y от t , но эта зависимость имеет m степеней свободы, т.е. содержит m неизвестных параметров, которые нужно подобрать. Предположим, что наблюдения Y дают ряд приближенных значений y_1, y_2, \dots, y_n , полученных при n различных значениях t . Неизвестные параметры обозначим через x_1, x_2, \dots, x_m .

Итак, задача состоит в том, чтобы определить оценки параметров $x_j, j = \overline{1, m}$ по данному наблюдениями ряду $y_k, k = \overline{1, n}$. Запишем постановку задачи в математических формулах

$$y_k = \varphi(t_k; x_1, x_2, \dots, x_m), \quad k = \overline{1, n}.$$

При $m > n$ эта задача невыполнима, так как имеет бесчисленное множество решений. При $m = n$ система может иметь однозначное решение, а при $m < n$ - число уравнений больше числа неизвестных и уравнения, вообще говоря, несовместны. Именно к этому случаю применяется метод наименьших квадратов (МНК), которому посвящается данная глава.

Допустим, что каким-либо способом мы выбрали набор неизвестных параметров: $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$. Подставим их в известную функцию зависимости y от t , получим оценки y_k :

$$\hat{y}_k = \varphi(t_k; \hat{x}_1, \hat{x}_2, \dots, \hat{x}_m), \quad k = \overline{1, n}.$$

Из-за ошибок наблюдений \hat{y}_k - вычисленное значение - не будет совпадать с наблюдаемым y_k . Набор исходных данных у астрономов принято обозначать буквой O (observatio, лат.), а вычисленных по формуле - буквой C (calculatio, лат.). Наилучшей аппроксимацией будет та, которая удовлетворяет условию

$$\|O - C\| = \min,$$

где двойными скобками обозначена норма - в определенном смысле среднее расстояние между O и C .

Как мы уже говорили, число параметров функции есть ее число степеней свободы. С помощью этих параметров кривая - геометрический образ функции - может менять свою форму и положение на плоскости. Для того, чтобы провести прямую, аппроксимирующую линейный тренд (ход) наблюдений, необходимо ввести два параметра, определяющие параллельный перенос и вращение прямой. Чтобы определить параболу второго порядка нужно иметь три параметра и т.д. Синусоида имеет также три степени свободы: положение осевой линии, амплитуду и фазу. Для определения орбит в задачах небесной механики необходимо определить шесть элементов - это шесть параметров. Для их определения необходимо мини-

мум три пары координат α и δ , но для уменьшения влияния погрешностей наблюдений, число пар координат должно быть существенно больше.

В 1792 году Лаплас в качестве нормы отклонений *O-C* принял сумму абсолютных значений остаточных разностей

$$L_1 = \sum_{k=1}^n |y_k - \hat{y}_k|.$$

Эта функция не дифференцируема и поиск минимальных значений L_1 затруднителен.

В 1905 году в книге “Новые методы определения орбит комет” Лежандр предложил в качестве меры точности приближения теоретической кривой к наблюдениям принять квадратичный критерий

$$L_2 = \sum (y_k - \hat{y}_k)^2.$$

В этом случае легко определить абсолютный минимум L_2 как функции параметров x_1, x_2, \dots, x_m . Условие минимума приводит к m уравнениям

$$\frac{\partial L_2(x_1, x_2, \dots, x_m)}{\partial x_j} = 0, \quad j = \overline{1, m}.$$

Эти уравнения получили название нормальных, а сам метод был назван методом наименьших квадратов (МНК).

Принцип минимизации суммы квадратов отклонений (*O-C*) называют иногда принципом Лежандра. Правда, немецкий ученый Гаусс этот принцип использовал раньше Лежандра на 10 лет. Поэтому было бы более правильным называть его принципом Гаусса-Лежандра. Позже он получил строгое математическое обоснование в теории вероятностей и математической статистики.

23. Метод наименьших квадратов (МНК) с независимыми равноточными наблюдениями

23.1 Применение МНК к линейным функциям

Рассмотрим случай зависимости наблюдаемой векторной величины $\mathbf{l} = (l_1, l_2, \dots, l_n)^T$ от искомых параметров $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$

$$\mathbf{l} = \mathbf{Ax} + \mathbf{v},$$

число строк матрицы \mathbf{A} определяется числом наблюдений, то есть числом моментов времени дискретного аргумента $l = l_k, k=1, 2, \dots, n$, а число столбцов – число неизвестных параметров m , \mathbf{v} - вектор остаточных разностей, которые в русской литературе иногда называют *невязками*. Решить эти уравнения нельзя, даже в случае, когда все невязки равны нулю и уравнение принимает вид $\mathbf{l} = \mathbf{Ax}$, так как матрица \mathbf{A} – не прямоугольная. Она не обратимая. Тем не менее в математике используется формальная запись решения такого урав-

нения: $\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{l}$, где матрица \mathbf{A}^+ носит название псевдообратной матрицы. Однако, нельзя определить однозначно псевдообратную матрицу. Один из способов – применение принципа Гаусса- Лежандра, то есть с помощью минимизации критерия L_2 .

Критерий L_2 в векторно-матричном виде будет выглядеть следующим образом

$$L_2 = \sum_{k=1}^n v_k^2 = \mathbf{v}^T \mathbf{v} = (\mathbf{1} - \mathbf{A}\hat{\mathbf{x}})^T (\mathbf{1} - \mathbf{A}\hat{\mathbf{x}}) = \min$$

Здесь $\hat{\mathbf{x}}$ - оценка вектора искомых параметров.

Критерий L_2 есть скалярная функция m аргументов, каждый из которых есть компонента вектора искомых параметров. Для достижения абсолютного минимума все первые производные L_2 по компонентам вектора \mathbf{x} должны обращаться в нуль

$$\frac{\partial L_2}{\partial x_j} = 0, \quad j = 1, 2, \dots, m$$

Таким образом, получим m уравнений для определения m неизвестных параметров. Эти уравнения получили название нормальных уравнений.

Распишем уравнения подробнее. Для этого нужно расписать критерий L_2 .

$$L_2 = \mathbf{l}^T \mathbf{l} - \mathbf{l}^T \mathbf{A} \mathbf{x} - \mathbf{x}^T \mathbf{A}^T \mathbf{l} + \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}$$

В координатах полученная формула будет выглядеть так:

$$L_2 = \sum_{k=1}^n [l_k^2 - 2l_k \sum_{j=1}^m A_{k,j} x_j + \sum_{i=1}^m \sum_{j=1}^m A_{k,i} A_{j,k} x_i x_j]$$

Теперь полученную скалярную функцию нам нужно продифференцировать по векторному аргументу \mathbf{x} , а результат запишем в виде матрицы-столбца. Установим сначала правило дифференцирования скалярной функции по вектору.

Пусть $z = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$. Продифференцируем данное скалярное произведение по вектору \mathbf{x} (то есть по всем компонентам этого вектора), а результат расположим столбцом. Получим

$$\frac{\partial z}{\partial x_1} = y_1$$

$$\frac{\partial z}{\partial x_2} = y_2$$

.....

$$\frac{\partial z}{\partial x_n} = y_n$$

Другими словами $\frac{\partial z}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}^T \mathbf{y})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{y}^T \mathbf{x})}{\partial \mathbf{x}} = \mathbf{y}$. Итак, если выполняется дифференцирование

скалярного произведения двух векторов по одному из этих векторов, то результатом будет второй вектор, если он записан в виде вектора-столбца, иначе его необходимо транспонировать. Вернёмся теперь к критерию Гаусса-Лежандра L_2 , записанному в матричной форме.

Первое слагаемое не содержит \mathbf{x} , поэтому дифференцирование даёт нуль. Второе и третье слагаемые можно представить как скалярные произведения двух векторов, а последнее слагаемое – билинейная форма. Это слагаемое нужно дифференцировать дважды по вектору \mathbf{x} . Получим

$$\frac{\partial L_2}{\partial \mathbf{x}} = -(\mathbf{1}^T \mathbf{A})^T - \mathbf{A}^T \mathbf{1} + \mathbf{A}^T \mathbf{A} \mathbf{x} + (\mathbf{x}^T \mathbf{A}^T \mathbf{A})^T = -2\mathbf{A}^T \mathbf{1} + 2\mathbf{A}^T \mathbf{A} \mathbf{x}$$

Оценку компонент вектора параметров \mathbf{x} найдём, приравнявая первую производную L_2 по составляющим вектора \mathbf{x} нулю. Получим нормальные уравнения, которые в матричной форме выглядят так

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{1}$$

Главная особенность нормальных уравнений заключается в том, что матрица коэффициентов этих уравнений симметрическая: её транспонирование не меняет её:

$$(\mathbf{A}^T \mathbf{A})^T = \mathbf{A}^T \mathbf{A}$$

Элементы этой матрицы симметричны относительно диагонали. Кроме того, если матрица нормальных уравнений неособенная, и определитель этой матрицы не равен нулю, то обратная матрица существует. Поэтому МНК-оценка вектора параметров нашей системы имеет вид

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{1}$$

Очевидно, что для матрицы \mathbf{A} псевдообратной, в данном случае, будет матрица

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

Вернёмся снова к исходным уравнениям

$$\mathbf{1} = \mathbf{A} \hat{\mathbf{x}} + \mathbf{v}$$

Умножим слева обе части этого уравнения на \mathbf{A}^T . Тогда

$$\mathbf{A}^T \mathbf{1} = \mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} + \mathbf{A}^T \mathbf{v}.$$

Но так как наш набор параметров определён с помощью решения нормальных уравнений, отсюда следует, что принцип Гаусса-Лежандра эквивалентен условию, которым подчиняются остаточные разности (невязки)

$$\mathbf{A}^T \mathbf{v} = \mathbf{0}$$

Определить МНК-оценки вектора параметров, конечно, важно, но это только полдела. Остаётся определить и средние квадратические ошибки этих оценок. Вначале мы договорились, что будем рассматривать случай равноточных и независимых наблюдений. Ковариационная матрица погрешностей наблюдений в этом случае имеет вид

$$\mathbf{C}_l = \sigma_0^2 \mathbf{E}_n,$$

где \mathbf{E}_n - единичная матрица размера $(n \times n)$. По определению ковариационная матрица ошибок наблюдений имеет вид

$$\mathbf{C}_l = \langle \Delta \mathbf{l} (\Delta \mathbf{l})^T \rangle$$

Для того, чтобы записать ошибку наблюдения, необходимо ввести понятие точных наблюдений, не содержащих ошибки. По-видимому, вектор точных наблюдений будет связан с вектором параметров системой тех же линейных уравнений

$$\mathbf{L} = \mathbf{A} \mathbf{X}$$

Погрешность МНК-оценки для вектора параметров будет

$$\Delta \mathbf{x} = \hat{\mathbf{x}} - \mathbf{X}, \text{ где}$$

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{l}$$

С другой стороны, $\mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} \mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{L}$, следовательно

$$\Delta \mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T (\mathbf{l} - \mathbf{L}) = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \Delta \mathbf{l}$$

Вычислим теперь ковариационную матрицу погрешностей оценивания. Напомним, что треугольные скобки означают вычисление математического ожидания (среднего значения). Эта операция относится только к случайным числам, в нашем случае – к погрешностям наблюдения и оценивания.

$$\mathbf{C}_x = \langle \Delta \mathbf{x} (\Delta \mathbf{x})^T \rangle = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \langle \Delta \mathbf{l} (\Delta \mathbf{l})^T \rangle \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}, \text{ но}$$

$$\langle \Delta \mathbf{l} (\Delta \mathbf{l})^T \rangle = \mathbf{C}_l = \sigma_0^2 \mathbf{E}_n, \text{ поэтому}$$

$$\mathbf{C}_x = \sigma_0^2 (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{E}_n \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} = \sigma_0^2 (\mathbf{A}^T \mathbf{A})^{-1}.$$

Из обработки ряда неравноточных наблюдений (см. главу 5) следует, что дисперсия погрешности отдельного наблюдения равна так называемой дисперсии ошибки единицы веса, разделённая на вес этого наблюдения. В нашем случае мы получили, что ковариационная матрица вектора ошибок оценивания параметров при независимых и равноточных наблюдениях равна дисперсии единицы веса умноженная на обратную матрицу нормальных уравне-

ний. Таким образом, матрица нормальных уравнений является весовой матрицей для иско-
мых параметров.

Ещё одно важное замечание. Ковариационная матрица ошибок МНК-оценивания не явля-
ется диагональной, хотя наблюдения мы предполагали независимыми. Следовательно, по-
грешности МНК-оценивания не являются, в общем случае, независимыми.

Приведём простой пример. Допустим, что требуется из линейных уравнений определить
три параметра параметра x , y и z . Эти параметры связаны с наблюдениями линейными урав-
нениями:

$$l_k = a_k x + b_k y + c_k z + \Delta l_k \quad k=1,2,3 \dots n$$

Используя принцип Гаусса – Лежандра,

$$L_2 = \sum_{k=1}^n [l_k - (a_k x + b_k y + c_k z)]^2.$$

Построим систему нормальных уравнений

$$\begin{aligned} \frac{\partial L_2}{\partial x} &= \frac{\partial}{\partial x} \sum (a_k x + b_k y + c_k z - l_k)^2 = 0, \\ \frac{\partial L_2}{\partial y} &= \frac{\partial}{\partial y} \sum (a_k x + b_k y + c_k z - l_k)^2 = 0, \\ \frac{\partial L_2}{\partial z} &= \frac{\partial}{\partial z} \sum (a_k x + b_k y + c_k z - l_k)^2 = 0. \end{aligned}$$

В данном случае матрица исходных уравнений имеет вид

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ \dots & \dots & \dots \\ a_n & b_n & c_n \end{pmatrix},$$

а нормальные уравнения

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{l}$$

расписываются следующим образом

$$\begin{aligned} \sum a_k^2 \cdot x + \sum a_k b_k \cdot y + \sum a_k c_k \cdot z &= \sum a_k l_k \\ \sum a_k b_k \cdot x + \sum b_k^2 \cdot y + \sum b_k c_k \cdot z &= \sum b_k l_k, \\ \sum a_k c_k \cdot x + \sum b_k c_k \cdot y + \sum c_k^2 \cdot z &= \sum c_k l_k. \end{aligned}$$

Имеем систему, состоящую из трех линейных уравнений с тремя неизвестными, которую
легко тем или иным способом решить.

Для обозначения сумм произведений или квадратов Гаусс предложил применять пря-
мые скобки следующим образом

$$\sum_{k=1}^n a_k b_k = [ab], \quad \sum_{k=1}^n a_k c_k = [ac], \quad \sum_{k=1}^n a_k^2 = [aa] \quad \sum_{k=1}^n a_k^2 = [aa] \quad \text{и т.д.}$$

В этих обозначениях нормальные уравнения примут вид

$$[aa]x + [ab]y + [ac]z = [al],$$

$$[ba]x + [bb]y + [bc]z = [bl],$$

$$[ca]x + [cb]y + [cc]z = [cl].$$

Основное свойство нормальных уравнений - симметричность матрицы системы:

$$[ab] = [ba], \quad [ac] = [ca], \quad [bc] = [cb].$$

Решить эту систему можно используя, например, формулы Крамера, или один из матричных способов.

Поскольку суммы $[al]$, $[bl]$, $[cl]$ содержат ошибки наблюдений $\Delta l_1, \Delta l_2, \dots, \Delta l_n$, то вычисленные из нормальных уравнений неизвестные x, y, z также будут содержать ошибки. Следовательно, нашим решением будет лишь оценки неизвестных $\hat{x}, \hat{y}, \hat{z}$. Запишем нормальные уравнения в матричной форме

$$\begin{pmatrix} [aa] & [ab] & [ac] \\ [ba] & [bb] & [bc] \\ [ca] & [cb] & [cc] \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} [al] \\ [bl] \\ [cl] \end{pmatrix}.$$

Очевидно, что погрешности неизвестных подчиняются той же системе уравнений

$$\begin{pmatrix} [aa] & [ab] & [ac] \\ [ba] & [bb] & [bc] \\ [ca] & [cb] & [cc] \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} [a\Delta l] \\ [b\Delta l] \\ [c\Delta l] \end{pmatrix},$$

где элементы матрицы-столбца в правой части уравнений нужно понимать так

$$[a\Delta l] = \sum_{k=1}^n a_k \Delta l_k, \quad [b\Delta l] = \sum_{k=1}^n b_k \Delta l_k, \quad [c\Delta l] = \sum_{k=1}^n c_k \Delta l_k.$$

Обозначим матрицу нормальных уравнений для краткости буквой Φ . Получим

$$\Phi = \begin{pmatrix} [aa] & [ab] & [ac] \\ [ba] & [bb] & [bc] \\ [ca] & [cb] & [cc] \end{pmatrix},$$

Теперь ковариационная матрица погрешностей оценивания параметров, согласно изложенному выше, будет иметь вид

$$C_x = \sigma_0^2 \Phi^{-1}$$

Введём обозначение

$$Q_x = \Phi^{-1},$$

тогда

$$\mathbf{Q}_x = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}.$$

Теперь образуем ковариационную матрицу ошибок оценок неизвестных:

$$\mathbf{C} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{pmatrix},$$

где $\sigma_x^2, \sigma_y^2, \sigma_z^2$ - дисперсии ошибок неизвестных, $\sigma_{xy}, \sigma_{xz}, \sigma_{yz}$ - ковариации этих неизвестных. Сразу же отметим, что если исходные данные представляют собой независимые наблюдения, то ошибки неизвестных не будут независимыми и их ковариации, вообще говоря, не равны нулю.

Отсюда следует, что

$$\begin{aligned} \sigma_x^2 &= \sigma_0^2 q_{11} \\ \sigma_y^2 &= \sigma_0^2 q_{22} \\ \sigma_z^2 &= \sigma_0^2 q_{33} \end{aligned}$$

Процедурой вычисления средних квадратических погрешностей МНК – оценок параметров, как правило, и заканчивается этот метод. Но в некоторых случаях может быть понадобится и ковариации погрешностей оценивания. Поступая только что изложенным методом, вычислим и ковариации

$$\begin{aligned} \sigma_{xy} &= \sigma_0^2 q_{12} \\ \sigma_{yx} &= \sigma_0^2 q_{23} \\ \sigma_{zy} &= \sigma_0^2 q_{31} \end{aligned}$$

Постоянные q_{11}, q_{22}, q_{33} называют также весовыми коэффициентами (в отличие от весов неизвестных p_x, p_y, p_z). Чтобы получить веса неизвестных, нужно найти обратную величину

$$p_x = 1/q_{11}, \quad p_y = 1/q_{22}, \quad p_z = 1/q_{33}.$$

В подавляющем большинстве случаев нам нужно определить средние квадратические ошибки МНК-оценок неизвестных. В этом случае вычисляют СКО “единицы веса” по формуле, которую мы введем несколько позже:

$$\varepsilon_0 = \sqrt{\frac{\sum v_k^2}{n - m}},$$

где v_k - остаточные разности: $v_k = l_k - (a_k x + b_k y + c_k z)$, а m - число неизвестных (у нас $m=3$).

Величина ε_0 , как правило, заменяет стандартное отклонение σ_0 и является его несмещенной оценкой. Поэтому СКО неизвестных определяют по формулам

$$\varepsilon_x = \varepsilon_0 \sqrt{q_{11}} = \frac{\varepsilon_0}{\sqrt{p_x}}, \quad \varepsilon_y = \varepsilon_0 \sqrt{q_{22}} = \frac{\varepsilon_0}{\sqrt{p_y}}, \quad \varepsilon_z = \varepsilon_0 \sqrt{q_{33}} = \frac{\varepsilon_0}{\sqrt{p_z}}.$$

Для системы трех уравнений веса неизвестных легко вычислить, образуя матрицу Φ . Диагональный элемент q_{11} обратной матрицы Φ^{-1} равен отношению определителя матрицы, полученной вычеркиванием первой строки и первого столбца, к определителю системы нормальных уравнений D

$$q_{11} = \frac{1}{D} \begin{vmatrix} [bb] & [bc] \\ [cb] & [cc] \end{vmatrix} = \frac{[bb] \cdot [cc] - [bc]^2}{D},$$

Аналогично

$$q_{22} = \frac{1}{D} \begin{vmatrix} [aa] & [ac] \\ [ca] & [cc] \end{vmatrix} = \frac{[aa] \cdot [cc] - [ac]^2}{D},$$

$$q_{33} = \frac{1}{D} \begin{vmatrix} [aa] & [ab] \\ [ba] & [bb] \end{vmatrix} = \frac{[aa] \cdot [bb] - [ab]^2}{D}.$$

Веса p_x, p_y, p_z - соответственно - отношения определителя системы к определителю матрицы, полученной вычеркиванием соответствующих строки и столбца, где расположен диагональный элемент обратной матрицы нормальных уравнений.

Численный пример.

Решим систему уравнений $l_k = a_k x + b_k y + c_k z + \Delta l_k$, $k = 1, 2, 3, \dots, 8$. Численные значения a_k, b_k, c_k, l_k сведены в таблицу

a_k	b_k	c_k	l_k	\tilde{l}_k	v_k
1	0	2	7	6.8	+0.2
0	3	-2	1	0.8	+0.2
-1	2	0	3	3.9	-0.9
2	-1	1	2	+ 2.2	-0.2
3	2	-2	1	0.6	+0.4
-2	-1	3	6	5.4	+0.6
0	3	-2	1	0.8	+0.2
4	1	0	5	5.2	-0.2

Составим нормальные уравнения. Первый элемент матрицы нормальных уравнений есть сумма квадратов элементов первого столбца (a_k), второй элемент первой строки есть сумма произведений элементов второго и первого столбцов $a_k \cdot b_k$, и т.д. Получим

$$\begin{pmatrix} 35 & 8 & -8 \\ 8 & 29 & -20 \\ -8 & -20 & 26 \end{pmatrix} \cdot \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \\ 28 \end{pmatrix}.$$

Решим полученную систему методом обращения матриц:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} 35 & 8 & -8 \\ 8 & 29 & -20 \\ -8 & -20 & 26 \end{pmatrix}^{-1} \begin{pmatrix} 19 \\ 11 \\ 28 \end{pmatrix} = \frac{1}{11430} \begin{pmatrix} 354 & -48 & 72 \\ -48 & 846 & 636 \\ 72 & 636 & 951 \end{pmatrix} \cdot \begin{pmatrix} 19 \\ 11 \\ 28 \end{pmatrix} = \begin{pmatrix} 0.72 \\ 2.29 \\ 3.06 \end{pmatrix}.$$

В пятом столбце таблицы приведены вычисленные значения \hat{l}_k , а в шестом - разности $l_k - \hat{l}_k$. Минимальная сумма квадратов разностей равна

$$\sum v_k^2 = 1.53, \text{ т.о. } \varepsilon_0^2 = \frac{1.53}{8-3} = 0.306.$$

Определим веса неизвестных

$$p_x = \frac{11430}{354} = 32.3, \quad p_y = \frac{11430}{846} = 13.5, \quad p_z = \frac{11430}{951} = 12.0.$$

Отсюда

$$\varepsilon_x = \sqrt{\frac{0.306}{32.3}} = 0.10, \quad \varepsilon_y = \sqrt{\frac{0.306}{13.5}} = 0.15, \quad \varepsilon_z = \sqrt{\frac{0.306}{12.0}} = 0.16.$$

Решение следует записать в виде

$$x=0.72 \pm 0.10, \quad y=2.29 \pm 0.15, \quad z=3.06 \pm 0.16.$$

23.2 МНК при неравноточных наблюдениях

В методе наименьших квадратов для неравноточных наблюдений нужно ввести веса и включить квадраты невязок с весами в критерий L_2 :

$$L_2 = \sum_{k=1}^n p_k v_k^2 = \sum_{k=1}^n p_k (l_k - a_{k1}\tilde{x}_1 - a_{k2}\tilde{x}_2 - \dots - a_{km}\tilde{x}_m)^2 \text{ или в матричной форме}$$

$$L_2 = \mathbf{v}^T \mathbf{P} \mathbf{v}.$$

Здесь p_k – веса отдельных наблюдений, а \mathbf{P} - диагональная матрица, составленная из весов наблюдений

$$\mathbf{P} = \text{diag } p_k \quad k = 1, 2, \dots, n.$$

Для того, чтобы определить МНК-оценку вектора x нужно найти минимум квадратичной формы L_2

$$\frac{\partial L_2}{\partial \mathbf{x}} = 0,$$

где дифференцирование по вектору \mathbf{x} означает дифференцирование скалярной функции по каждой из составляющих этого вектора. Мы с вами уже сталкивались со случаем, когда нужно было дифференцировать скалярную функцию по векторному аргументу. Дифференцируя функцию

$$L_2(\mathbf{x}) = (\mathbf{I}^T - \mathbf{x}^T \mathbf{A}^T) \mathbf{P}(\mathbf{I} - \mathbf{A}\mathbf{x})$$

по вектору искомых параметров, и приравнявая производные нулю, получим систему нормальных уравнений

$$\mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{P} \mathbf{l}$$

Отсюда МНК-оценка искомых параметров при неравноточных наблюдениях

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l} .$$

Покажем, что это оценка несмещенная. Напомним, что оценка приближённой величины называется несмещённой, если её математическое ожидание равно точному значению этой величины. Пусть \mathbf{X} - точные значения вектора параметров, а \mathbf{L} - наблюдения, не содержащие ошибок. Тогда, очевидно, должно выполняться строгое равенство

$$\mathbf{L} = \mathbf{A}\mathbf{X} .$$

Умножим слева обе части равенства на $\mathbf{A}^T \mathbf{P}$, получим

$$\mathbf{A}^T \mathbf{P} \mathbf{L} = \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{X} .$$

Вычислим точное значение вектора \mathbf{X} :

$$\mathbf{X} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{L} .$$

Учитывая формулу для МНК-оценки $\hat{\mathbf{x}}$, получим

$$\hat{\mathbf{x}} - \mathbf{X} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}(\mathbf{l} - \mathbf{L}) .$$

Если \mathbf{l} содержит только случайные ошибки, т.е.

$$\langle \mathbf{l} - \mathbf{L} \rangle = 0 ,$$

то выполняется также равенство

$$\langle \hat{\mathbf{x}} - \mathbf{X} \rangle = 0 .$$

Другими словами, $\langle \hat{\mathbf{x}} \rangle = \mathbf{X}$.

Вывод: МНК-оценка $\hat{\mathbf{x}}$ является несмещенной при любой матрице весов наблюдений \mathbf{P} .

23.3 Ковариационная матрица МНК-оценки

Вычислим теперь ковариационную матрицу ошибок неизвестных параметров. Для упрощения записи снова обозначим матрицу нормальных уравнений буквой Φ . Теперь она имеет вид

$$\Phi = \mathbf{A}^T \mathbf{P} \mathbf{A} .$$

Вектор ошибок Δx можно записать так:

$$\hat{\mathbf{x}} - \mathbf{X} = \Phi^{-1} \mathbf{A}^T \mathbf{P}(\mathbf{I} - \mathbf{L}),$$

$$\Delta \mathbf{x} = \Phi^{-1} \mathbf{A}^T \mathbf{P} \Delta \mathbf{l}.$$

По определению, ковариационная матрица ошибок равна

$$\mathbf{C}_x = \langle \Delta \mathbf{x} \Delta \mathbf{x}^T \rangle = \Phi^{-1} \mathbf{A}^T \mathbf{P} \langle \Delta \mathbf{l} \Delta \mathbf{l}^T \rangle \mathbf{P} \mathbf{A} \Phi^{-1} = \Phi^{-1} \mathbf{A}^T \mathbf{P} \mathbf{C}_1 \mathbf{P} \mathbf{A} \Phi^{-1}$$

Здесь мы учли, что P - диагональная матрица и $P=P^T$.

Ковариационная матрица неравноточных независимых наблюдений имеет диагональный вид

$$\mathbf{C}_1 = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

Умножим эту матрицу на P , например, слева. Получим

$$\mathbf{P} \mathbf{C}_1 = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} = \begin{pmatrix} p_1 \sigma_1^2 & 0 & \dots & 0 \\ 0 & p_2 \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n \sigma_n^2 \end{pmatrix}.$$

Но $p_k \sigma_k^2 = \sigma_0^2$, $k = \overline{1, n}$ есть “дисперсия единицы веса”. Следовательно,

$$\mathbf{P} \mathbf{C}_1 = \begin{pmatrix} \sigma_0^2 & 0 & \dots & 0 \\ 0 & \sigma_0^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_0^2 \end{pmatrix} = \sigma_0^2 \mathbf{E}_n.$$

Вернемся к формуле, определяющей ковариационную матрицу вектора погрешностей МНК-оценивания

$$\mathbf{C}_x = \Phi^{-1} \mathbf{A}^T \mathbf{P} \mathbf{C}_1 \mathbf{P} \mathbf{A} \Phi^{-1} = \sigma_0^2 \Phi^{-1} \mathbf{A}^T \mathbf{E}_n \mathbf{P} \mathbf{A} \Phi^{-1} = \sigma_0^2 \Phi^{-1}.$$

Мы вновь получили тот же результат, что и при равноточных наблюдениях (см. 6.2.2): ковариационная матрица неизвестных с точностью до множителя равна обратной матрице нормальных уравнений. В данном случае σ_0^2 - есть дисперсия ошибки наблюдений, вес которому мы приписали равным единице (дисперсия единицы веса).

Ковариационная матрица ошибок неизвестных содержит исчерпывающую информацию о точности определения каждого параметра из x_1, x_2, \dots, x_n и ковариации их ошибок. Остается неопределенной величина σ_0^2 . Ее, как и в случае одного ряда неравноточных наблюдений, можно априорно задать или вычислить по результатам эксперимента.

23.4 Апостериорная оценка дисперсии единицы веса

Теперь получим формулу, позволяющую вычислить несмещенную оценку дисперсии единицы веса по данным наблюдений. Поступим также, как и в случае определения дисперсии единицы веса при неравноточных наблюдениях.

Рассмотрим взвешенную сумму квадратов остаточных разностей

$$L_2 = \sum_{k=1}^n p_k v_k^2 = (\mathbf{1} - \mathbf{A}\hat{\mathbf{x}})^T \mathbf{P}(\mathbf{1} - \mathbf{A}\hat{\mathbf{x}}).$$

Поскольку $\hat{\mathbf{x}}$ является несмещенной оценкой, то справедливо равенство

$$L_2 = (\Delta\mathbf{l} - \mathbf{A} \cdot \Delta\mathbf{x})^T \mathbf{P}(\Delta\mathbf{l} - \mathbf{A} \cdot \Delta\mathbf{x}).$$

Вычислим среднее значение для L_2

$$\langle L_2 \rangle = \langle (\Delta\mathbf{l} - \mathbf{A} \cdot \Delta\mathbf{x})^T \mathbf{P}(\Delta\mathbf{l} - \mathbf{A} \cdot \Delta\mathbf{x}) \rangle = \langle \Delta\mathbf{l}^T \mathbf{P} \Delta\mathbf{l} - \Delta\mathbf{l}^T \mathbf{P} \mathbf{A} \Delta\mathbf{x} - \Delta\mathbf{x}^T \mathbf{A}^T \mathbf{P} \Delta\mathbf{l} + \Delta\mathbf{x}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \Delta\mathbf{x} \rangle.$$

Выполним преобразования в каждом слагаемом:

$$\begin{aligned} \langle \Delta\mathbf{l}^T \mathbf{P} \Delta\mathbf{l} \rangle &= \langle (\Delta l_1 \quad \Delta l_2 \quad \dots \quad \Delta l_n) \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n \end{pmatrix} \begin{pmatrix} \Delta l_1 \\ \Delta l_2 \\ \dots \\ \Delta l_n \end{pmatrix} \rangle = \\ &= \langle (\Delta l_1 p_1 \quad \Delta l_2 p_2 \quad \dots \quad \Delta l_n p_n) \begin{pmatrix} \Delta l_1 \\ \Delta l_2 \\ \dots \\ \Delta l_n \end{pmatrix} \rangle = \langle p_1 \Delta l_1^2 + p_2 \Delta l_2^2 + \dots + p_n \Delta l_n^2 \rangle = \\ &= p_1 \langle \Delta l_1^2 \rangle + p_2 \langle \Delta l_2^2 \rangle + \dots + p_n \langle \Delta l_n^2 \rangle = p_1 \sigma_1^2 + p_2 \sigma_2^2 + \dots + p_n \sigma_n^2. \end{aligned}$$

Но $p_k \sigma_k^2 = \sigma_0^2$, $k = \overline{1, n}$, поэтому

$$\langle \Delta\mathbf{l}^T \mathbf{P} \Delta\mathbf{l} \rangle = n \sigma_0^2.$$

Ранее мы получили, что $\Delta\mathbf{x} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \Delta\mathbf{l}$, поэтому

$$\langle \Delta\mathbf{l}^T \mathbf{P} \mathbf{A} \Delta\mathbf{x} \rangle = \langle \Delta\mathbf{l}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \Delta\mathbf{l} \rangle.$$

Обозначим временно

$$\mathbf{a} = \mathbf{A}^T \mathbf{P} \Delta\mathbf{l}, \quad \text{поэтому} \quad \mathbf{a}^T = \Delta\mathbf{l}^T \mathbf{P} \mathbf{A}, \quad \text{Учитывая, что, кроме того}$$

$\Phi = \mathbf{A}^T \mathbf{P} \mathbf{A}$, будем иметь

$$\langle \Delta\mathbf{l}^T \mathbf{P} \mathbf{A} \Delta\mathbf{x} \rangle = \langle \mathbf{a}^T \Phi^{-1} \mathbf{a} \rangle.$$

Поскольку \mathbf{a} - вектор-столбец ($m \times 1$), то приведенное выше равенство - скалярная величина. Перенесём первый сомножитель на конец приведённого произведения. При этом необходимо учитывать, что

$$\mathbf{a}^T \Phi^{-1} \mathbf{a} = \text{tr} \Phi^{-1} \mathbf{a} \mathbf{a}^T,$$

где $\text{tr} \Phi^{-1}$ - след (сумма диагональных элементов) квадратной матрицы размера $(m \times m)$.

Заметим, что $\mathbf{a}^T \mathbf{b} = \text{tr} \mathbf{b} \mathbf{a}^T$. Последнее утверждение легко проверить на простом примере:

$$\mathbf{a}^T \mathbf{b} = (a_1 \quad a_2 \quad a_3) \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

С другой стороны

$$\mathbf{b} \mathbf{a}^T = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} (a_1 \quad a_2 \quad a_3) = \begin{pmatrix} a_1 b_1 & a_2 b_1 & a_3 b_1 \\ a_1 b_2 & a_2 b_2 & a_3 b_2 \\ a_1 b_3 & a_2 b_3 & a_3 b_3 \end{pmatrix}.$$

След полученной матрицы, равный сумме ее диагональных элементов, есть $a_1 b_1 + a_2 b_2 + a_3 b_3$.

Таким образом, $\mathbf{a}^T \mathbf{b} = \text{tr} \mathbf{b} \mathbf{a}^T$.

В нашем случае $\mathbf{b} = \Phi^{-1} \mathbf{a}$, отсюда следует справедливое утверждение, что

$$\langle \Delta \mathbf{l}^T \mathbf{P} \mathbf{A} \Delta \mathbf{x} \rangle = \langle \mathbf{a}^T \Phi^{-1} \mathbf{a} \rangle = \text{tr} \langle \Phi^{-1} \mathbf{a} \mathbf{a}^T \rangle = \text{tr} \langle \Phi^{-1} (\mathbf{A}^T \mathbf{P} \Delta \mathbf{l}) (\Delta \mathbf{l}^T \mathbf{P} \mathbf{A}) \rangle = \text{tr} \langle \Phi^{-1} \mathbf{A}^T \mathbf{P} \langle \Delta \mathbf{l} \Delta \mathbf{l}^T \rangle \mathbf{P} \mathbf{A} \rangle.$$

$$\text{Но } \langle \Delta \mathbf{l} \Delta \mathbf{l}^T \rangle = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} = \sigma_0^2 \begin{pmatrix} p_1^{-1} & 0 & \dots & 0 \\ 0 & p_2^{-1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n^{-1} \end{pmatrix} = \sigma_0^2 \mathbf{P}^{-1}, \text{ поэтому}$$

$$\langle \Delta \mathbf{l}^T \mathbf{P} \mathbf{A} \Delta \mathbf{x} \rangle = \text{tr} \langle \Phi^{-1} \mathbf{A}^T \mathbf{P} \mathbf{A} \rangle \sigma_0^2 = \text{tr} \mathbf{E}_m \sigma_0^2 = m \sigma_0^2.$$

Третий член равенства для $\langle L_2 \rangle$ представляет собой транспонированное выражение второго члена. А так как L_2 - скалярная величина, то оба эти члена совпадают.

Наконец, определим последнее, четвертое слагаемое

$$\begin{aligned} \langle \Delta \mathbf{x}^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \Delta \mathbf{x} \rangle &= \langle [\Phi^{-1} \mathbf{A}^T \mathbf{P} \Delta \mathbf{l}]^T \Phi [\Phi^{-1} \mathbf{A}^T \mathbf{P} \Delta \mathbf{l}] \rangle = \langle \Delta \mathbf{l}^T \mathbf{P} \mathbf{A} : \Phi^{-1} \mathbf{A}^T \mathbf{P} \Delta \mathbf{l} \rangle = \\ &= \text{tr} \Phi^{-1} \mathbf{A}^T \mathbf{P} \langle \Delta \mathbf{l} \Delta \mathbf{l}^T \rangle \mathbf{P} \mathbf{A} = \sigma_0^2 \text{tr} \Phi^{-1} \mathbf{A}^T \mathbf{P} \mathbf{A} = \sigma_0^2 \text{tr} \mathbf{E}_m = m \sigma_0^2. \end{aligned}$$

Итак, окончательно

$$\langle L_2 \rangle = n \sigma_0^2 - m \sigma_0^2 - m \sigma_0^2 + m \sigma_0^2 = (n - m) \sigma_0^2.$$

Следовательно,

$$\sigma_0^2 = \frac{\langle L_2 \rangle}{n - m} = \left\langle \frac{1}{n - m} \sum_{k=1}^n p_k v_k^2 \right\rangle.$$

Отбрасывая треугольные скобки, получим несмещенную оценку дисперсии единицы веса - квадрат средней квадратической ошибки единицы веса

$$\varepsilon_0^2 = \frac{1}{n-m} \sum_{k=1}^n p_k v_k^2,$$

где $v_k = l_k - (a_{k1}\tilde{x}_1 + a_{k2}\tilde{x}_2 \dots + a_{km}\tilde{x}_m)$, $k = \overline{1, n}$, - остаточные разности.