

二维情形下的加权最小二乘拟合问题

刘牛*

2021-03-07

摘要

天体测量的数据中，因变量通常是二维的，如位置偏差、自行、章动角、极移等，且这些二维观测量的分量存在一定的相关性。教科书上常用的最小二乘估计法一般是针对一维因变量型的数据，而缺乏基于二维因变量型的观测数据进行最小二乘估计的方案，尤其是需要考虑观测量分量之间相关性的情形。在这里我尝试推导过二维情形下的最小二乘估计的目标函数，并给出了相应的正规方程及其矩阵形式。

1 引言

根据《天文数据处理》中的描述 [1]，参数估计问题可以转换为下述的数学问题：若两个观测量 x 和 y 存在某种关于 m 个待定参数 c_k ($k = 1, 2, \dots, m$) 的形式已知的函数关系，对于 N 对观测值 (x_i, y_i) ($i = 1, 2, \dots, N$)，确定参数 c_k 的某一估计值 \hat{c}_k ，使某个目标函数

$$d = d(x_1, x_2, \dots, x_N; y_1, y_2, \dots, y_N; c_1, c_2, \dots, c_m) \quad (1)$$

因此，曲线拟合就是对目标函数进行最优化计算，寻求是目标函数 d 取极值的一组参数值。显然，这样的表述是针对观测量 y 为一维的情形给出的。当观测量 \mathbf{y} (为了区分我使用了黑体) 为二维变量时，我们可以将上面的描述外推，得到当因变量为二维变量时的最小二乘参数估计形式。

2 最小二乘估计的目标函数

2.1 回顾一维情形

首先考虑双变量型数据，数据形式可以表述为

$$(x_i, \sigma_{x,i}, y_i, \sigma_{y,i}), i = 1, 2, \dots, n \quad (2)$$

其中， x_i 为自变量， y_i 为因变量，也就是观测量， $\sigma_{x,i}$ 和 $\sigma_{y,i}$ 分别为相应的测量误差。一般假设观测量彼此之间相互独立，即对于任意的 $i \neq j$ ， y_i 与 y_j 相互独立。一般假设自变量的测量误差为零，即

$$\sigma_{x,i} = 0 \quad (3)$$

*电子邮箱: niu.liu@nju.edu.cn

在天文观测资料处理过程中，对观测误差的估计通常与观测值本身之间相互独立，即 y_i 与 $\sigma_{y,i}$ 相互独立。为了讨论方便，假设观测量的理论模型为线性模型，具体形式为

$$F(x_i, \mathbf{c}) = \sum_{j=1}^m c_j F_j(x_i) \quad (4)$$

其中模型中有 m 个未知参数需要估计，待估参数向量可记为

$$\mathbf{c} = (c_1, c_2, \dots, c_m)^T \quad (5)$$

则第 i 次的观测值可写为

$$y_i = F(x_i, \mathbf{c}) + \epsilon_i \quad (6)$$

其中 ϵ_i 为第 i 次观测值与理论值之间的偏差，即残差。归一化残差（即加权后的残差）为

$$X_i = \frac{\epsilon_i}{\sigma_{y,i}} \quad (7)$$

最小二乘意义下的参数估计就是选取各观测点的归一化残差平方和作为目标函数

$$X^2 = \sum_{i=1}^N X_i^2 = \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_{y,i}^2} = \sum_{i=1}^N \frac{\left(y_i - \sum_{j=1}^m c_j F_j(x_i)\right)^2}{\sigma_{y,i}^2} \quad (8)$$

X^2 也就是常说的 χ^2 。选取一组合适的参数值 \hat{c}_k ，使得目标函数 X^2 取极小值，此时， X^2 相对于未知参数 c_k 的偏导为零，即

$$0 = \frac{\partial X^2}{\partial c_j} = \sum_{i=1}^N \frac{\partial X_i^2}{\partial c_j}, j = 1, 2, \dots, m \quad (9)$$

2.2 延伸至二维情形

现在我们考虑当观测量 \mathbf{y}_i 为二维变量时的情形，即

$$\mathbf{y}_i = \left(y_i^{(1)}, y_i^{(2)}\right)^T \quad (10)$$

数据形式变为

$$(x_i, \sigma_{x,i}, y_i^{(1)}, y_i^{(2)}, \sigma_{y,i}^{(1)}, \sigma_{y,i}^{(2)}, \rho_i), i = 1, 2, \dots, n \quad (11)$$

其中， ρ_i 表示第 i 次的观测值向量 \mathbf{y}_i 两个分量 $y_i^{(1)}$ 和 $y_i^{(2)}$ 之间的（Pearson）相关性系数。这里，我们再次假设任意两次观测的观测量分量之间相互独立，即对于任意的 $i \neq j$ ， $y_i^{(1)}$ 和 $y_i^{(2)}$ 、 $y_i^{(1)}$ 和 $y_j^{(2)}$ 、 $y_j^{(1)}$ 和 $y_i^{(2)}$ 、 $y_j^{(1)}$ 和 $y_j^{(2)}$ 均相互独立。

第 i 次的观测量 \mathbf{y}_i 的两个分量对应的理论模型为

$$F^{(1)}(x_i, \mathbf{c}) = \sum_{j=1}^m c_j F_j^{(1)}(x_i) \quad (12)$$

$$F^{(2)}(x_i, \mathbf{c}) = \sum_{j=1}^m c_j F_j^{(2)}(x_i) \quad (13)$$

则第 i 次观测值的两个分量可以表示为

$$y_i^{(1)} = F^{(1)}(x_i, \mathbf{c}) + \epsilon_i^{(1)} \quad (14)$$

$$y_i^{(2)} = F^{(2)}(x_i, \mathbf{c}) + \epsilon_i^{(2)} \quad (15)$$

归一化的残差分别为

$$X_i^{(1)} = \frac{\epsilon_i^{(1)}}{\sigma_{y,i}^{(1)}}, \quad X_i^{(2)} = \frac{\epsilon_i^{(2)}}{\sigma_{y,i}^{(2)}} \quad (16)$$

借用 Mignard 等人 [2] 提出的 X 统计量, 归一化的二维残差 X_i 可由下式导出

$$X_i^2 = \begin{bmatrix} X_i^{(1)} & X_i^{(2)} \end{bmatrix} \begin{bmatrix} 1 & \rho_i \\ \rho_i & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_i^{(1)} \\ X_i^{(2)} \end{bmatrix} \quad (17)$$

$$= \frac{1}{1 - \rho_i^2} \begin{bmatrix} X_i^{(1)} & X_i^{(2)} \end{bmatrix} \begin{bmatrix} 1 & -\rho_i \\ -\rho_i & 1 \end{bmatrix} \begin{bmatrix} X_i^{(1)} \\ X_i^{(2)} \end{bmatrix} \quad (18)$$

$$= \frac{1}{1 - \rho_i^2} \left[\left(X_i^{(1)} \right)^2 + \left(X_i^{(2)} \right)^2 - 2\rho_i X_i^{(1)} X_i^{(2)} \right] \quad (19)$$

目标函数即归一化的二维残差平方和为

$$X^2 = \sum_{i=1}^N X_i^2 = \sum_{i=1}^N \frac{1}{1 - \rho_i^2} \left[\left(X_i^{(1)} \right)^2 + \left(X_i^{(2)} \right)^2 - 2\rho_i X_i^{(1)} X_i^{(2)} \right] \quad (20)$$

当观测量 \mathbf{y}_i 的两个分量之间不相关时, 即

$$\rho_i = 0, \quad i = 1, 2, \dots, N \quad (21)$$

此时, 目标函数变为

$$X^2 = \sum_{i=1}^N X_i^2 = \sum_{i=1}^N \left(X_i^{(1)} \right)^2 + \sum_{i=1}^N \left(X_i^{(2)} \right)^2 = \left(X^{(1)} \right)^2 + \left(X^{(2)} \right)^2 \quad (22)$$

即将两个分量视为相互独立的一维观测量时得到的目标函数之和。

2.3 由极大似然估计间接导出

在观测量 \mathbf{y} 为一维情形下, 若观测残差 ϵ_i 只包含随机噪声, 即观测残差服从高斯分布

$$\epsilon_i \sim N(0, \sigma_{y,i}) \quad (23)$$

此时, 极大似然估计等价于最小二乘估计。这一结论也可以推广到二维甚至更高维情形。现在我先简单给出一维情形下的不严格证明, 之后再从二维情形下的极大似然估计出发给出二维情形下的最小二乘估计的目标函数。

假设在方程 (6) 对应的模型下, 在待估参数 \mathbf{c} 处得到一对观测值 (x_i, y_i) 的概率密度为 $f(x_i, y_i; \mathbf{c})$ 。现在进行了 N 次观测之后, 得到一组如方程 (2) 所示的观测值这一事件 p 发生的概率密度为 [3]

$$L(\mathbf{c}) = \prod_{i=1}^N f(x_i, y_i; \mathbf{c}) \quad (24)$$

$L(\mathbf{c})$ 即似然函数。由于事件 p 是真实发生的, 因此在待估参数 \mathbf{c} 取得最佳估计 $\hat{\mathbf{c}}$ 时, $L(\mathbf{c})$ 取得极大值。但因为 $L(\mathbf{c})$ 是连乘形式, 不利于计算, 常选取其对数形式, 即对数似然函数 $l(\mathbf{c})$ 来进行参数估计, 其表达式为

$$l(\mathbf{c}) = \ln L(\mathbf{c}) = \sum_{i=1}^N \ln f(x_i, y_i; \mathbf{c}) \quad (25)$$

这就是极大似然估计法的目标函数。不同于最小二乘估计, 这里取目标函数的极大值。在方程 (23) 的假设下, 有

$$f(x_i, y_i; \mathbf{c}) = f(\epsilon_i; \mathbf{c}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\epsilon_i^2}{\sigma_{y,i}^2}\right) \quad (26)$$

由此可以计算得

$$l(\mathbf{c}) = \sum_{i=1}^N \left[-\ln(\sqrt{2\pi}\sigma_i) - \frac{\epsilon_i^2}{\sigma_{y,i}^2} \right] = -\sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_{y,i}^2} + \text{const} \quad (27)$$

上式等号右边最后一项为与任一待估参数 c_k 均无关的常数项。可见, 对数似然函数 $l(\mathbf{c})$ 取极大值等价于方程 (8) 中的 X^2 取极小值, 因此, 此时的极大似然估计等价于最小二乘估计。我不加证明地将这一结论外推到二维乃至高维情形, 并利用这一结论反推出二维情形下最小二乘估计法的目标函数。

在二维情形下, 若观测残差 ϵ_i 服从二维高斯分布, 则有

$$\begin{aligned} & f(\epsilon_i^{(1)}, \epsilon_i^{(2)}; \mathbf{c}) \\ &= \frac{1}{2\pi\sigma_{y,i}^{(1)}\sigma_{y,i}^{(2)}\sqrt{1-\rho_i^2}} \exp\left\{-\frac{1}{2(1-\rho_i^2)}\left[\left(\frac{\epsilon_i^{(1)}}{\sigma_{y,i}^{(1)}}\right)^2 + \left(\frac{\epsilon_i^{(2)}}{\sigma_{y,i}^{(2)}}\right)^2 - 2\rho_i\frac{\epsilon_i^{(1)}}{\sigma_{y,i}^{(1)}}\frac{\epsilon_i^{(2)}}{\sigma_{y,i}^{(2)}}\right]\right\} \\ &= \frac{1}{2\pi\sigma_{y,i}^{(1)}\sigma_{y,i}^{(2)}\sqrt{1-\rho_i^2}} \exp\left\{-\frac{1}{2(1-\rho_i^2)}\left[\left(X_i^{(1)}\right)^2 + \left(X_i^{(2)}\right)^2 - 2\rho_i X_i^{(1)} X_i^{(2)}\right]\right\} \end{aligned} \quad (28)$$

由此可知对数似然函数 $l(\mathbf{c})$ 的表达式为

$$\begin{aligned} l(\mathbf{c}) &= \sum_{i=1}^N \left\{ -\frac{1}{2(1-\rho_i^2)} \left[\left(X_i^{(1)}\right)^2 + \left(X_i^{(2)}\right)^2 - 2\rho_i X_i^{(1)} X_i^{(2)} \right] - \ln\left(2\pi\sigma_{y,i}^{(1)}\sigma_{y,i}^{(2)}\sqrt{1-\rho_i^2}\right) \right\} \\ &= -\sum_{i=1}^N \frac{1}{2(1-\rho_i^2)} \left[\left(X_i^{(1)}\right)^2 + \left(X_i^{(2)}\right)^2 - 2\rho_i X_i^{(1)} X_i^{(2)} \right] + \text{const} \end{aligned} \quad (29)$$

可见, 上式的对数似然函数 $l(\mathbf{c})$ 取极大值等价于方程 (20) 中的 X^2 取极小值。这也反过来说明了方程 (20) 的正确性。

3 二维情形下最小二乘估计法的实现

前面分别给出了最小二乘估计法在一维和二维情形下的目标函数, 这里我将导出相应的正规方程, 从而给出最小二乘估计法的具体实现。我依然先推导一维情形的正规方程。利用方程 (8) 将方程 (9) 展开, 对任意的 $j = 1, 2, \dots, m$ 均有

$$0 = \sum_{i=1}^N \frac{\partial X_i^2}{\partial c_j} = \sum_{i=1}^N 2X_i \frac{\partial X_i}{\partial c_j} = \sum_{i=1}^N -2 \left(y_i - \sum_k^m c_k F_k(x_i) \right) \frac{F_j(x_i)}{\sigma_{y,i}^2} \quad (30)$$

进一步展开, 有

$$\sum_{i=1}^N \sum_k^m \frac{c_k F_j(x_i) F_k(x_i)}{\sigma_{y,i}^2} = \sum_{i=1}^N \frac{F_j(x_i) y_i}{\sigma_{y,i}^2}, j = 1, 2, \dots, m \quad (31)$$

考虑一个模型理论值向量

$$\mathbf{F} = (F(x_1, \mathbf{c}), F(x_2, \mathbf{c}), \dots, F(x_N, \mathbf{c}))^T \quad (32)$$

向量 \mathbf{F} 相对于 \mathbf{c} 的偏导构成了雅可比矩阵

$$\mathcal{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{c}}, \quad \mathcal{J}_{ij} = \frac{\partial F(x_i, \mathbf{c})}{\partial c_j} = F_j(x_i, \mathbf{c}) \quad (33)$$

向量 \mathbf{F} 相对于任一待估参数 c_j 的偏导构成了向量

$$\frac{\partial \mathbf{F}}{\partial c_j} = (F_j(x_1, \mathbf{c}), F_j(x_2, \mathbf{c}), \dots, F_j(x_N, \mathbf{c}))^T \quad (34)$$

可见, $\frac{\partial \mathbf{F}}{\partial c_j}$ 是雅可比矩阵 \mathcal{J} 的第 j 列, 而 $\left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T$ 是雅可比矩阵的转置矩阵 \mathcal{J}^T 的第 j 行. 引入一个加权矩阵 \mathcal{M} , 它可以看作是观测向量 \mathbf{y} 的协方差矩阵 \mathcal{V} 的求逆, 其表达式分别为

$$\mathcal{V} = \begin{pmatrix} \sigma_{y,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{y,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{y,N}^2 \end{pmatrix} \quad (35)$$

$$\mathcal{M} = \begin{pmatrix} \sigma_{y,1}^{-2} & 0 & \dots & 0 \\ 0 & \sigma_{y,2}^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{y,N}^{-2} \end{pmatrix} = \mathcal{V}^{-1} \quad (36)$$

一组观测值构成一个观测值向量, 即

$$\mathbf{y} = (y_1, y_2, \dots, y_N)^T \quad (37)$$

那么, 方程 (31) 中对 i 求和可以写为

$$\sum_{i=1}^N \frac{F_j(x_i) F_k(x_i)}{\sigma_{y,i}^2} c_k = \left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \mathcal{M} \frac{\partial \mathbf{F}}{\partial c_k} c_k \quad (38)$$

$$\sum_{i=1}^N \frac{F_j(x_i) y_i}{\sigma_{y,i}^2} = \left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \mathcal{M} \mathbf{y} \quad (39)$$

类似地, 方程 (38) 再对 k 求和也可以写成矩阵乘积的形式, 为

$$\left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \sum_{k=1}^m \frac{\partial \mathbf{F}}{\partial c_k} c_k = \left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \mathcal{J} \mathbf{c} \quad (40)$$

方程 (31) 可以写成矩阵相乘的形式, 有

$$\left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \mathcal{M} \mathcal{J} \mathbf{c} = \left(\frac{\partial \mathbf{F}}{\partial c_j}\right)^T \mathcal{M} \mathbf{y}, j = 1, 2, \dots, m \quad (41)$$

将所有的 $j = 1, 2, \dots, m$ 对应的上述方程叠加在一起, 则有

$$\mathcal{J}^T \mathcal{M} \mathcal{J} \mathbf{c} = \mathcal{J}^T \mathcal{M} \mathbf{y} \quad (42)$$

从而, 可以求出未知参数向量的最小二乘估计

$$\hat{\mathbf{c}} = (\mathcal{J}^T \mathcal{M} \mathcal{J})^{-1} \mathcal{J}^T \mathcal{M} \mathbf{y} \quad (43)$$

当观测量 \mathbf{y} 为二维时, 情况要复杂得多。类似于方程 (30), 有

$$\begin{aligned} & \frac{\partial X_i^2}{\partial c_j} \\ &= \frac{1}{1 - \rho_i^2} \left(2X_i^{(1)} \frac{\partial X_i^{(1)}}{\partial c_j} + 2X_i^{(2)} \frac{\partial X_i^{(2)}}{\partial c_j} - 2\rho_i \frac{\partial X_i^{(1)}}{\partial c_j} X_i^{(2)} - 2\rho_i X_i^{(1)} \frac{\partial X_i^{(2)}}{\partial c_j} \right) \\ &= \frac{2}{1 - \rho_i^2} \left[- \left(y_i^{(1)} - \sum_k c_k F_k^{(1)}(x_i) \right) \frac{F_j^{(1)}(x_i)}{(\sigma_{y,i}^{(1)})^2} - \left(y_i^{(2)} - \sum_k c_k F_k^{(2)}(x_i) \right) \frac{F_j^{(2)}(x_i)}{(\sigma_{y,i}^{(2)})^2} \right. \\ & \quad \left. + \rho_i \left(y_i^{(2)} - \sum_k c_k F_k^{(2)}(x_i) \right) \frac{F_j^{(1)}(x_i)}{\sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)}} + \rho_i \left(y_i^{(1)} - \sum_k c_k F_k^{(1)}(x_i) \right) \frac{F_j^{(2)}(x_i)}{\sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)}} \right], \\ & j = 1, 2, \dots, m \end{aligned} \quad (44)$$

将上式对 i 求和, 并令其等于 0, 有

$$\begin{aligned} & \sum_{i=1}^N \sum_k^m \frac{c_k}{1 - \rho_i^2} \left[\frac{F_j^{(1)}(x_i) F_k^{(1)}(x_i)}{(\sigma_{y,i}^{(1)})^2} + \frac{F_j^{(2)}(x_i) F_k^{(2)}(x_i)}{(\sigma_{y,i}^{(2)})^2} - \frac{\rho_i (F_j^{(1)}(x_i) F_k^{(2)}(x_i) + F_j^{(2)}(x_i) F_k^{(1)}(x_i))}{\sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)}} \right] \\ &= \sum_{i=1}^N \sum_k^m \frac{1}{1 - \rho_i^2} \left[\frac{F_j^{(1)}(x_i) y_i^{(1)}}{(\sigma_{y,i}^{(1)})^2} + \frac{F_j^{(2)}(x_i) y_i^{(2)}}{(\sigma_{y,i}^{(2)})^2} - \frac{\rho_i (F_j^{(1)}(x_i) y_i^{(2)} + F_j^{(2)}(x_i) y_i^{(1)})}{\sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)}} \right], \\ & j = 1, 2, \dots, m \end{aligned} \quad (45)$$

方程 (45) 等式左边求和子项可以写为

$$\begin{aligned} & \begin{pmatrix} F_j^{(1)}(x_i) & F_j^{(2)}(x_i) \end{pmatrix} \begin{pmatrix} (\sigma_{y,i}^{(1)})^{-1} & 0 \\ 0 & (\sigma_{y,i}^{(2)})^{-1} \end{pmatrix} \frac{1}{1 - \rho_i^2} \begin{pmatrix} 1 & -\rho_i \\ -\rho_i & 1 \end{pmatrix} \\ & \times \begin{pmatrix} (\sigma_{y,i}^{(1)})^{-1} & 0 \\ 0 & (\sigma_{y,i}^{(2)})^{-1} \end{pmatrix} \begin{pmatrix} F_k^{(1)}(x_i) \\ F_k^{(2)}(x_i) \end{pmatrix} \\ &= \begin{pmatrix} F_j^{(1)}(x_i) & F_j^{(2)}(x_i) \end{pmatrix} \begin{pmatrix} (\sigma_{y,i}^{(1)})^2 & \rho_i \sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)} \\ \rho_i \sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)} & (\sigma_{y,i}^{(2)})^2 \end{pmatrix}^{-1} \begin{pmatrix} F_k^{(1)}(x_i) \\ F_k^{(2)}(x_i) \end{pmatrix} \end{aligned} \quad (46)$$

上式第二项即为第 i 次观测量 \mathbf{y}_i 的协方差矩阵。同样地, 方程 (45) 等式右边求和子项可以写为

$$\begin{pmatrix} F_j^{(1)}(x_i) & F_j^{(2)}(x_i) \end{pmatrix} \begin{pmatrix} (\sigma_{y,i}^{(1)})^2 & \rho_i \sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)} \\ \rho_i \sigma_{y,i}^{(1)} \sigma_{y,i}^{(2)} & (\sigma_{y,i}^{(2)})^2 \end{pmatrix}^{-1} \begin{pmatrix} y_i^{(1)} \\ y_i^{(2)} \end{pmatrix} \quad (47)$$

我们可以构建合适的矩阵 \mathcal{J} 、 \mathcal{M} 和 \mathbf{y} ，来将方程 (45) 改写成如方程 (42) 的矩阵形式。至少有两种方案可以实现上述目的。第一种方案的矩阵 \mathcal{J} 、 \mathcal{V} 和 \mathbf{y} 如下所示：

$$\mathcal{J} = \begin{pmatrix} \frac{\partial \mathbf{F}^{(1)}}{\partial \mathbf{c}} & \frac{\partial \mathbf{F}^{(2)}}{\partial \mathbf{c}} \end{pmatrix}, \quad (48)$$

$$\mathcal{J}_{ij} = \frac{\partial F^{(1)}(x_i, \mathbf{c})}{\partial c_j} = F_j^{(1)}(x_i, \mathbf{c}) \quad (49)$$

$$\mathcal{J}_{(i+N),j} = \frac{\partial F^{(2)}(x_i, \mathbf{c})}{\partial c_j} = F_j^{(2)}(x_i, \mathbf{c}) \quad (50)$$

$$\mathcal{V} = \begin{pmatrix} (\sigma_{y,1}^{(1)})^2 & \cdots & O & \rho_1 \sigma_{y,1}^{(1)} \sigma_{y,1}^{(2)} & \cdots & O \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ O & \cdots & (\sigma_{y,N}^{(1)})^2 & O & \cdots & \rho_N \sigma_{y,N}^{(1)} \sigma_{y,N}^{(2)} \\ \rho_1 \sigma_{y,1}^{(1)} \sigma_{y,1}^{(2)} & \cdots & O & (\sigma_{y,1}^{(2)})^2 & \cdots & O \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ O & \cdots & \rho_N \sigma_{y,N}^{(1)} \sigma_{y,N}^{(2)} & O & \cdots & (\sigma_{y,N}^{(2)})^2 \end{pmatrix} \quad (51)$$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}^T = (y_1^{(1)}, y_2^{(1)}, \dots, y_N^{(1)}, y_1^{(2)}, y_2^{(2)}, \dots, y_N^{(2)})^T \quad (52)$$

另一种方案则是这样的：

$$\mathcal{J}_{(2i-1),j} = \frac{\partial F^{(1)}(x_i, \mathbf{c})}{\partial c_j} = F_j^{(1)}(x_i, \mathbf{c}) \quad (53)$$

$$\mathcal{J}_{(2i),j} = \frac{\partial F^{(2)}(x_i, \mathbf{c})}{\partial c_j} = F_j^{(2)}(x_i, \mathbf{c}) \quad (54)$$

$$\mathcal{V} = \begin{pmatrix} (\sigma_{y,1}^{(1)})^2 & \rho_1 \sigma_{y,1}^{(1)} \sigma_{y,1}^{(2)} & \cdots & O & O \\ \rho_1 \sigma_{y,1}^{(1)} \sigma_{y,1}^{(2)} & (\sigma_{y,1}^{(2)})^2 & \cdots & O & O \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & (\sigma_{y,N}^{(1)})^2 & \rho_N \sigma_{y,N}^{(1)} \sigma_{y,N}^{(2)} \\ O & O & \cdots & \rho_N \sigma_{y,N}^{(1)} \sigma_{y,N}^{(2)} & (\sigma_{y,N}^{(2)})^2 \end{pmatrix} \quad (55)$$

$$\mathbf{y} = (y_1^{(1)}, y_1^{(2)}, y_2^{(1)}, y_2^{(2)}, \dots, y_N^{(1)}, y_N^{(2)})^T \quad (56)$$

\mathcal{M} 可由方程 (51) 或 (55)，结合方程 (36) 得出。利用方程 (42) 和 (43)，就可以得到未知参数的估计值。方程 (52) 和 (56) 实际上是将二维的观测值向量转换成了一维观测值向量。

当观测量 \mathbf{y} 的两个分量不相关时，目标函数可以简化为

$$X^2 = \sum_{i=1}^N X_i^2 = \sum_{i=1}^N (X_i^{(1)})^2 + \sum_{i=1}^N (X_i^{(2)})^2 = (X^{(1)})^2 + (X^{(2)})^2 \quad (57)$$

方程 (51) 和 (55) 给出的协方差矩阵 \mathcal{V} 和相应的加权矩阵 \mathcal{M} 均变为对角矩阵。

4 总结与讨论

借由上述推导，我们得到了在观测因变量为二维情况下的最小二乘估计的具体实现。对于更高维度的情形，可以进行类似的分析。对我个人而言，最容易理解的就是方程 (48) 到 (56) 的矩阵形式。对于更加复杂的情况，如不同次的观测值 y_i 之间存在某种相关性且相关性系数（或者协方差）已知，只需将协方差矩阵 \mathbf{V} 的对应元素作出相应调整（从零调整为协方差）即可。观测量分量之间的相关性究竟如何影响到参数的最小二乘估计结果？对于这一问题，目前我还没有明确的结果和结论，后续我会利用一些典型的研究问题进行分析比较。

参考文献

- [1] 丁月蓉. 天文数据处理方法. 南京大学出版社, 1998.
- [2] F. Mignard, S. Klioner, L. Lindegren, U. Bastian, A. Bombrun, J. Hernández, D. Hobbs, U. Lammers, D. Michalik, M. Ramos-Lerate, M. Biermann, A. Butkevich, G. Comoretto, E. Joliet, B. Holl, A. Hutton, P. Parsons, H. Steidelmüller, A. Andrei, G. Bourda, and P. Charlot. Gaia Data Release 1. Reference frame and optical properties of ICRF sources. A&A, 595:A5, November 2016.
- [3] Eric D Feigelson and G Jogesh Babu. Modern statistical methods for astronomy: with R applications. Cambridge University Press, 2012.