Table of Contents

Linear Least Squares

1. Wikipedia links

- Least squares http://en.wikipedia.org/wiki/Least_squares
    - Linear least squares http://en.wikipedia.org/wiki/Least_squares#Linear_least_squares
    - Weighted least squares http://en.wikipedia.org/wiki/Least_squares#Weighted_least_squares
- Linear least squares (mathematics) http://en.wikipedia.org/wiki/Linear_least_squares_%28mathematics%29
    - Weighted linear least squares http://en.wikipedia.org/wiki/Linear_least_squares_%28mathematics%29#Weighted_linear_least_squa
- Propagation of uncertainty http://en.wikipedia.org/wiki/Propagation_of_uncertainty

2. Uncorrelated measurements

Let $\boxed{y_i, i=1,\dots,m}$ with variances $\boxed{\sigma_i^2}$ be measurements of functions $\phi_i = \Sigma_{j=1}^n X_{ij}\beta_j$ with the known $\boxed{m \times n}$ matrix $\boxed{X}$ and unknown parameters $\boxed{\beta_j, j=1,\dots,n}$.

In linear least square method, one estimates the parameter vector β by minimizing over β the expression

$$S = \sum_{i=1}^{m} W_{ii}(y_i - \sum_{j=1}^{n} X_{ij}\beta_j)^2 \quad , \qquad (2.1)$$

where the *weight matrix* W of dimension $m \times m$ is diagonal and defined as the *inverse of the diagonal covariance matrix for* y: $\boxed{W = M_y^{-1}}$ i.e. $W_{ii} = 1/\sigma_i^2, i=1,\dots,m$.
In matrix notation (considering y and β as columns $m \times 1$ and $n \times 1$ respectively), one has

$$S = (y - X\beta)^T W (y - X\beta) = y^T W y - 2\beta^T X^T W y + \beta^T X^T W X \beta \quad . \qquad (2.2)$$

The estimate $\hat{\beta}$ is the solution of the system of equations

$$\frac{\partial S}{\partial \beta_p} = \sum_{i=1}^{m} W_{ii}(-2y_i X_{ip} + 2\sum_{j=1}^{n} X_{ij}X_{ip}\hat{\beta}_j) = 0 \quad , \quad p = 1,\dots,n \qquad (2.3)$$

or

$$X^T W X \hat{\beta} = X^T W y \quad \Leftrightarrow \quad \boxed{\hat{\beta} = (X^T W X)^{-1} X^T W y} \quad . \qquad (2.4)$$

In a general case of linear transformation $y = Ax$, the covariance matrice M_x for x is transformed into that for y via $M_y = A M_x A^T$. Hence,

$$\begin{aligned} M_{\hat{\beta}} &= (X^T W X)^{-1} X^T W M_y [(X^T W X)^{-1} X^T W]^T = \\ &= (X^T W X)^{-1} X^T W M_y W^T X (X^T W^T X)^{-1} \quad . \qquad (2.5) \end{aligned}$$

With $W = M_y^{-1}$ by the definition of W, that simplifies to

$$\boxed{M_{\hat{\beta}} = (X^T W X)^{-1}} \quad . \qquad (2.6)$$

Note, that

$$\partial^2 S/\partial\beta_p\partial\beta_q = \sum_{i=1}^{m} W_{ii}(2X_{iq}X_{ip}) = 2(X^TWX)_{qp} \quad , \qquad (2.7)$$

and

$$S(\beta_p;\beta_{q\neq p} = \hat{\beta}_q) = S_{min} + \frac{1}{2}\frac{\partial^2 S}{\partial{\beta_p}^2}\cdot(\beta_p - \hat{\beta}_p)^2 \quad . \qquad (2.8)$$

3. Correlated measurements

Let y be uncorrelated measurements as those in the previous section, and $\boxed{y = Ay'}$ (A is an invertible $m \times m$ matrix). Then $y'_1, \dots, y'_m$ are generally correlated and have the covariance matrix $\boxed{M_{y'} = A^{-1}M_y(A^{-1})^T}$.

With $M_y = AM_{y'}A^T$ and $M_y^{-1} = (A^{-1})^T M_{y'}^{-1} A^{-1}$, one gets

$$\begin{aligned}
\boxed{S} &= (y - X\beta)^T W (y - X\beta) = \\
&= (Ay' - X\beta)^T M_y^{-1} (Ay' - X\beta) = \\
&= (Ay' - X\beta)^T (A^{-1})^T M_{y'}^{-1} A^{-1} (Ay' - X\beta) = \\
&= (y' - A^{-1}X\beta)^T M_{y'}^{-1} (y' - A^{-1}X\beta) \\
&= \boxed{(y' - X'\beta)^T M_{y'}^{-1} (y' - X'\beta)} \quad , \qquad (3.1)
\end{aligned}$$

where $\boxed{X' = A^{-1}X}$.

Similarly,

$$\begin{aligned}
\boxed{\hat{\beta}} &= (X^T M_y^{-1} X)^{-1} X^T M_y^{-1} y = \\
&= (X^T [M_y^{-1}] X)^{-1} X^T [M_y^{-1}][y] = \\
&= (X'^T A^T [(A^{-1})^T M_{y'}^{-1} A^{-1}] A X')^{-1} X'^T A^T [(A^{-1})^T M_{y'}^{-1} A^{-1}][Ay'] = \\
&= \boxed{(X'^T M_{y'}^{-1} X')^{-1} X'^T M_{y'}^{-1} y'} \quad , \qquad (3.2)
\end{aligned}$$

and

$$\boxed{M_{\hat{\beta}}} = (X^T M_y^{-1} X)^{-1} = \boxed{(X'^T M_{y'}^{-1} X')^{-1}} \quad , \qquad (3.3)$$

Thus, all the formulae for the correlated measurements y' are similar to those for the uncorrelated y, with the only complication being the replacement of a diagonal weight matrix with a non-diagonal one:

$$\boxed{\text{diagonal matrix} \quad W = W_y = M_y^{-1} \qquad \longrightarrow \qquad \text{non-diagonal} \quad W_{y'} = M_{y'}^{-1}} \quad . \qquad (3.4)$$

4. An iterative solution

4.1 The procedure

Let n indices of parameters β be distributed among N groups $g_1, \dots, g_N$ with sizes $n(g_1), \dots, n(g_N)$ respectively.

$$g_k = \{i_1(g_k), \dots, i_{n(g_k)}(g_k)\}, \quad k = 1, \dots, N \quad . \qquad (4.1.1)$$

The subset of the parameters β corresponding to the k-th group of indices, can be considered as an $n(g_k) \times 1$ column

$$\beta_{(g_k)} = \{\beta_{i_1(g_k)}, \ldots, \beta_{i_{n(g_k)}(g_k)}\}^T, \quad k = 1, \ldots, N \qquad . \qquad (4.1.2)$$

Let $\bar{g_k}$ denote the set of $n - n(g_k)$ indices that are complementary to $n(g_k)$ indices in g_k.

Consider the following iterative procedure.

1. Start with an n vector $\beta^{(0)}$ as an initial approximation to $\hat{\beta}$.
2. Make N steps or, equivalently, iterations for $k = 1, \ldots, N$, thus finding n vectors $\beta^{(k)}, k = 1, \ldots, N$ by minimizing S over $\beta_{(g_k)}$ at the k-th step :

$$\beta_i^{(k)} = \begin{cases} (\hat{\beta}_{(g_k)}^{(k)})_i & \text{if } i \in g_k, \\ \beta_i^{(k-1)} & \text{if } i \notin g_k. \end{cases} \qquad , \qquad (4.1.3)$$

   where $\hat{\beta}_{(g_k)}^{(k)}$ is the $\beta_{(g_k)}$ "point" where S as a function of $n(g_k)$ parameters $\beta_{(g_k)}$, takes minimum (while the rest $n - n(g_k)$ parameters are fixed at the values obtained in the previous iteration: $(\beta_{(\bar{g}_k)})_i = (\beta^{(k-1)})_i, i \notin g_k$).
3. Repeat (4.1.3) infinitly, defining $g_k = g_{[(k-1) \bmod N]+1}$ for $k > N$ (i.e. $g_{L*N+k_0} = g_{k_0}, \ k_0 = 1, \ldots, N \ , L = 1, 2, \ldots$) .

One can expect that

$$\lim_{k \to \infty} \beta^{(k)} = \hat{\beta} \qquad . \qquad (4.1.4)$$

4.2 The step

The k-th iteration consists in finding $\beta_{(g_k)}^{(k)} = \{\beta_j^{(k)} \ , \ j \in g_k\}^T$, a set of $n(g_k)$ parameters, that minimizes

$$S = \sum_{i=1}^{m} W_{ii}(y_i - \sum_{j \in \bar{g}_k}^{n} X_{ij}\beta_j^{(k-1)} - \sum_{j \in g_k}^{n} X_{ij}\beta_j^{(k)})^2 \qquad . \qquad (4.2.1)$$

Let $\boxed{X_{g_k}}$ be an $m \times n(g_k)$ submatrix of X built from the columns of X with indices belonging to g_k (in other words, it is what remains after *removing* columns which have indices ***not*** in g_k).
Similarly, let $\boxed{X_{\bar{g}_k}}$ be an $m \times (n - n(g_k))$ submatrix consisting of the columns of X with indices belonging to $\bar{g}_k$.
Introducing

$$y' = y - X_{\bar{g}_k}\beta^{(k-1)} \qquad (4.2.2)$$

the (4.2.1) can be written as

$$S = (y' - X_{g_k}\beta_{(g_k)}^{(k)})^T W (y' - X_{g_k}\beta_{(g_k)}^{(k)}) \qquad . \qquad (4.2.3)$$

which is analogous to eq.(2.2).
Therefore, the solution is given by formulae (2.4):

$$X_{g_k}^T W X_{g_k}\beta_{(g_k)}^{(k)} = X_{g_k}^T W y' \qquad , \qquad (4.2.4)$$

or

$$\begin{aligned} \beta_{(g_k)}^{(k)} &= (X_{g_k}^T W X_{g_k})^{-1} X_{g_k}^T W y' = \\ &= (X_{g_k}^T W X_{g_k})^{-1} X_{g_k}^T W (y - X_{\bar{g}_k}\beta^{(k-1)}) \qquad . \qquad (4.2.5) \end{aligned}$$

Let us write the last expression via matrices of dimensions

$$\begin{array}{cccccc} & n\times 1, & n\times m, & 1\times n, & m\times n & \\ \text{instead of} & n(g_k)\times 1, & n(g_k)\times m, & 1\times n(g_k), & m\times n(g_k), & \text{respectively,}\end{array}$$

and

$$\begin{array}{cc} & m\times n \\ \text{instead of} & m\times n(\bar{g_k}) \end{array} \quad ,$$

such that all the arythmetics is done in rows / columns g_k (or $\bar{g_k}$) while complementary rows / columns contain zeros.

We define $n\times n$ matrices $\boxed{I_{g_k}}$ and $\boxed{I_{\bar{g_k}}}$ as follows

$$(I_{g_k})_{i,j} = \begin{cases} \delta_{i,j} & \text{if } i,j \in g_k, \\ 0 & \text{otherwise.} \end{cases} \quad , \qquad (4.2.6)$$

$$(I_{\bar{g_k}})_{i,j} = \begin{cases} \delta_{i,j} & \text{if } i,j \in \bar{g_k}, \\ 0 & \text{otherwise.} \end{cases} \quad . \qquad (4.2.7)$$

It is noteworthy that

$$I_{g_k} + I_{\bar{g_k}} = I \quad . \qquad (4.2.8)$$

We also define the $n\times n$ matrix $\boxed{(X_{g_k}^T W X_{g_k})_E^{-1}}$ (subscript E stands for *extended*) via the $n(g_k)\times n(g_k)$ matrix $(X_{g_k}^T W X_{g_k})^{-1}$:

$$[(X_{g_k}^T W X_{g_k})_E^{-1}]_{i,j} = \begin{cases} [(X_{g_k}^T W X_{g_k})^{-1}]_{i,j} & \text{if } i,j \in g_k, \\ 0 & \text{otherwise.} \end{cases} \quad . \qquad (4.2.9)$$

Then (4.2.5) can be written as

$$I_{g_k}\beta^{(k)} = (X_{g_k}^T W X_{g_k})_E^{-1} X^T W (y - X I_{\bar{g_k}}\beta^{(k-1)}) \quad . \qquad (4.2.10)$$

This is a representation of the first line of (4.1.3).
The second line of (4.1.3) can be represented with

$$I_{\bar{g_k}}\beta^{(k)} = I_{\bar{g_k}}\beta^{(k-1)} \quad . \qquad (4.2.11)$$

Summing up (4.2.10) and (4.2.11) gives

$$\begin{aligned} \boxed{\beta^{(k)}} &= (I_{g_k} + I_{\bar{g_k}})\beta^{(k)} = \\ &= (X_{g_k}^T W X_{g_k})_E^{-1} X^T W y + [I - (X_{g_k}^T W X_{g_k})_E^{-1} X_T W X] I_{\bar{g_k}}\beta^{(k-1)} = \\ &= \boxed{A_{g_k} X^T W y + B_{g_k}\beta^{(k-1)}} \quad . \qquad (4.2.12) \end{aligned}$$

where we denoted

$$\begin{aligned} \boxed{A_{g_k}} &= (X_{g_k}^T W X_{g_k})_E^{-1} \quad , \qquad (4.2.13) \\ \boxed{B_{g_k}} &= (I - A_{g_k} X^T W X) I_{\bar{g_k}} \quad . \qquad (4.2.14) \end{aligned}$$

4.3 steps

4.3.1 as a function of

Let us define matrices A_k and B_k as

$$\boxed{A_k} = \begin{cases} 0 & , \quad k = 0 \\ A_{g_k} + B_{g_k} A_{k-1} & , \quad k > 0. \end{cases} \quad , \qquad (4.3.1)$$

and

$$\boxed{B_k} = \prod_{i=k}^{1} B_{g_i} = B_{g_k} \cdot \ldots \cdot B_{g_1} \quad . \qquad (4.3.2)$$

Then, from (4.2.12),

$$\boxed{\beta^{(k)} = A_k X^T W y + B_k \beta^{(0)}} \quad . \qquad (4.3.3)$$

Indeed, by induction:

- eq.(4.3.3) holds for $k = 1$:

$$\beta^{(1)} = A_{g_1} X^T W y + B_{g_1} \beta^{(0)} = A_1 X^T W y + B_1 \beta^{(0)} \quad , \qquad (4.3.4)$$

- and assuming it is true for $\beta^{(p)}$, leads to

$$\begin{aligned} \beta^{(p+1)} &= A_{g_{p+1}} X^T W y + B_{g_{p+1}} \beta^{(p)} = \\ &= A_{g_{p+1}} X^T W y + B_{g_{p+1}} (A_p X^T W y + B_p \beta^{(0)}) = \\ &= (A_{g_{p+1}} + B_{g_{p+1}} A_p) X^T W y + B_{g_{p+1}} B_p \beta^{(0)} = \\ &= A_{p+1} X^T W y + B_{p+1} \beta^{(0)} \quad , \qquad (4.3.5) \end{aligned}$$

In particular, the eq.(4.3.3) is valid for $k = N$:

$$\boxed{\beta^{(N)} = A_N X^T W y + B_N \beta^{(0)}} \quad . \qquad (4.3.6)$$

4.3.2 The covariance matrix of

According to the general rule of covariance matrix transformation ($y = Ax \quad \Rightarrow \quad M_y = A M_x A^T$), eq.(4.3.3) gives rise to

$$\begin{aligned} M_{\beta^{(k)}} &= A_k X^T W M_y (A_k X^T W)^T = A_k X^T W M_y W^T X A_k^T = \\ &= A_k X^T W X A_k^T \quad . \qquad (4.3.7) \end{aligned}$$

Comparing this with the covariance matrix of the exact solution, eq.(2.6), yields

$$\boxed{\lim_{k \to \infty} A_k X^T W X A_k^T = M_{\hat{\beta}} = (X^T W X)^{-1}} \quad . \qquad (4.3.8)$$

4.4 steps for

With the expression (4.3.6) of the form

$$\beta^{(N)} = F(\beta^{(0)}) \quad . \qquad (4.4.1)$$

one can build the expression for the $k = 2N$ case ***in one step*** as

$$\beta^{(2N)} = F(\beta^{(N)}) = F(F(\beta^{(0)})) \quad . \qquad (4.4.2)$$

Substituting (4.3.6) into (4.4.2) gives

$$\begin{aligned} \beta^{(2N)} &= A_N X^T W y + B_N (A_N X^T W y + B_N \beta^{(0)}) = \\ &= (I + B_N) A_N X^T W y + B_N^2 \beta^{(0)} \quad . \end{aligned} \qquad (4.4.3)$$

Transforming similarly the $2N$ formula to the $4N$ one, gives

$$\beta^{(4N)} = (I + B_N^2)(I + B_N) A_N X^T W y + B_N^4 \beta^{(0)} \quad . \qquad (4.4.4)$$

Then for the $8N$ case one has

$$\beta^{(8N)} = (I + B_N^4)(I + B_N^2)(I + B_N) A_N X^T W y + B_N^8 \beta^{(0)} \quad , \qquad (4.4.4)$$

and so on...

This generalizes to

$$\boxed{\beta^{(2^p N)} = \left(\prod_{i=p-1}^{0} (I + B_N^{2^i}) \right) A_N X^T W y + B_N^{2^p} \beta^{(0)} \quad , \quad p = 1, 2, \ldots} \quad , \qquad (4.4.5)$$

This topic: Main > AVFLogA012LeastSquares
Topic revision: r14 - 2017-01-15 - AlexanderFedotov