

例 1 データ

$$(A, B, Y) = (1, 3, 3.2), (-6, 2, 0.8)$$

に対して, (1) は

$$\begin{cases} \alpha + 3\beta = 3.2 \\ -6\alpha + 2\beta = 0.8 \end{cases} \quad (2)$$

となり, この連立方程式を解いて

$$\alpha = 0.2, \quad \beta = 1$$

を得る.

例 2 データ

$$(A, B, Y) = (1, 3, 3), (-6, 2, 2)$$

に対して, (1) は

$$\begin{cases} \alpha + 3\beta = 3 \\ -6\alpha + 2\beta = 2 \end{cases} \quad (3)$$

となり, この連立方程式を解いて

$$\alpha = 0, \quad \beta = 1$$

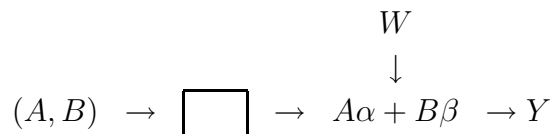
を得る. この機械では入力 A は出力に影響しない (A の入力端子の具合が悪いのかもしれない).

2 ノイズで汚染されたデータ

入力 A, B に対して機械は $A\alpha + B\beta$ を出力するが、今度はそれにノイズ W が加わった値を観測するものとする:

$$Y = A\alpha + B\beta + W.$$

ただし、我々が観測できるのは (A, B, Y) のみであって、 W の値は知ることはできない。



次の表はある α, β の値に対して、いろいろな入力 A, B と対応する観測 Y を与えている (入力ごとにノイズの値は異なる)。

A	B	Y
2.62	3.99	10.29
3.22	4.75	15.25
2.37	3.24	7.67
3.32	4.54	11.2
3.36	4.13	8.71
3.05	4.71	12.98
3.17	5.93	20.24
2.82	4.72	17.16
3.04	3.64	10.91
2.76	3.14	4.29

A	B	Y
2.98	5.52	18.79
2.79	5.17	17.29
3.12	3.37	5.18
3.14	4.22	11.87
2.75	3.51	7.21
3.48	4.04	11.3
2.79	2.25	1.59
2.31	4.68	15.78
3.04	5.67	16.77
2.61	4.85	15.52

表のデータに左上から順に番号を付け、 i 番目のデータを (A_i, B_i, Y_i) で表し、対応するノイズを W_i で表す。つまり、

$$A_i \alpha + B_i \beta + W_i = Y_i \quad (i = 1, 2, \dots, 20) . \quad (4)$$

我々は W_i が平均的に 0 であることを想定しているが、個々の W_i が 0 であるわけではない。したがって、 W_i があるために、(1) のような連立方程式を書くことができない ($A_i \alpha + B_i \beta = Y_i - W_i$ としても我々は W_i の値を知らない)。

3 パラメータ α, β の推定

データ (A_i, B_i, Y_i) ($i = 1, 2, \dots, n$) を用いて、値が未知のパラメータ α, β の推定をしたい。

最小二乗法。

$$\sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2$$

を最小にする α, β の値をそれぞれのパラメータの推定値とする。

Y_i と $A_i \alpha + B_i \beta$ の相違

$$|Y_i - (A_i \alpha + B_i \beta)| \quad (5)$$

が小さいとき、出力 Y_i が入力 A_i, B_i でよく説明できたといえるが、この説明誤差は i 毎に異なるので、誤差の総体

$$\sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2$$

を最小にするようにパラメータ α, β を調整するのである。²

² 2 乗するのは、一つには扱いが容易であるからである。たとえば (5) をそのまま加えるこ

R, S, T を

$$R = \sum_{i=1}^n A_i^2, \quad S = \sum_{i=1}^n A_i B_i, \quad T = \sum_{i=1}^n B_i^2$$

とおく．さらに，

$$C = \sum_{i=1}^n A_i Y_i, \quad D = \sum_{i=1}^n B_i Y_i$$

とおく．

$\hat{\alpha}, \hat{\beta}$ が次の連立一次方程式の解であるとする：

$$\begin{cases} R\hat{\alpha} + S\hat{\beta} = C \\ S\hat{\alpha} + T\hat{\beta} = D \end{cases} \quad (6)$$

このとき，

$$\begin{aligned} \sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2 &= \sum_{i=1}^n \left\{ Y_i - (A_i \hat{\alpha} + B_i \hat{\beta}) \right\}^2 \\ &\quad + \sum_{i=1}^n \left\{ A_i (\hat{\alpha} - \alpha) + B_i (\hat{\beta} - \beta) \right\}^2 \quad (7) \end{aligned}$$

が成り立つ．証明はセクション 4 を見よ．(7) から

$$\alpha = \hat{\alpha}, \quad \beta = \hat{\beta}$$

とするとき

$$\sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2$$

が最小となることがわかる．すなわち， $\hat{\alpha}, \hat{\beta}$ が α, β の最小二乗推定値である．
まとめると，

最小二乗推定値 $\hat{\alpha}, \hat{\beta}$ の求め方：連立一次方程式 (6) を解く．

とも可能であるが，推定値は最小二乗法によるものとは違ったものになる．状況によってはその方が良い推定を与えることもある．

例 3 セクション 2 で与えられたデータセットに対して最小二乗推定を実行しよう。まず、

$$R = 174.43, \quad S = 253.9176, \quad T = 387.0279, \\ C = 707.0342, \quad D = 1119.9059$$

となることが計算からわかる（懸命に手計算するか計算機を使う）。したがって、連立一次方程式 (6) は

$$\begin{cases} 174.43 \hat{\alpha} + 253.9176 \hat{\beta} = 707.0342 \\ 253.9176 \hat{\alpha} + 387.0279 \hat{\beta} = 1119.9059 \end{cases}$$

となる。これを解いて、

$$\hat{\alpha} = -3.53258681067067, \quad \hat{\beta} = 5.21123119226586$$

をパラメータの推定値として得る。³

セクション 2 のデータは、モデル (4) に基づいて計算機で人工的に作ったものである。それに用いたパラメータの真の値は

$$\alpha = -3.6, \quad \beta = 5.4$$

である。上記の推定値はこの真の値をかなり良く当てているといえよう。

ちなみに、同じやり方でデータを 1000 個発生させて、それに基づいてパラメータ推定を行い、

$$\hat{\alpha} = -3.53642419921925, \quad \hat{\beta} = 5.35083152858422 \quad (1000 \text{ 個})$$

なる結果を得た。推定の精度はすこし改善がみられる。

データ数が少ないときは、良い推定精度は期待できない。たとえばはじめの 10 個のデータだけを使って推定すると、

$$\hat{\alpha} = -4.51509026059096, \quad \hat{\beta} = 5.91950723733901 \quad (10 \text{ 個})$$

となり、推定誤差は大きい。

³実際のデータ処理においては観測値の記述の誤差がもともとあるから、推定値は適当な桁の数とすべきであろう。

4 式(7)の証明

まず,

$$\sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2 = \sum_{i=1}^n \left\{ [Y_i - (A_i \hat{\alpha} + B_i \hat{\beta})] + [A_i (\hat{\alpha} - \alpha) + B_i (\hat{\beta} - \beta)] \right\}^2 \quad (8)$$

として, 右辺を展開することを考える. クロスタームは

$$\begin{aligned} & \sum_{i=1}^n [Y_i - (A_i \hat{\alpha} + B_i \hat{\beta})][A_i (\hat{\alpha} - \alpha) + B_i (\hat{\beta} - \beta)] \\ = & \sum_{i=1}^n [Y_i - (A_i \hat{\alpha} + B_i \hat{\beta})]A_i (\hat{\alpha} - \alpha) + \sum_{i=1}^n [Y_i - (A_i \hat{\alpha} + B_i \hat{\beta})]B_i (\hat{\beta} - \beta) \\ = & (\hat{\alpha} - \alpha) \left[\sum_{i=1}^n A_i Y_i - \sum_{i=1}^n A_i^2 \hat{\alpha} - \sum_{i=1}^n A_i B_i \hat{\beta} \right] + (\hat{\beta} - \beta) \left[\sum_{i=1}^n B_i Y_i - \sum_{i=1}^n A_i B_i \hat{\alpha} - \sum_{i=1}^n B_i^2 \hat{\beta} \right] \\ = & (\hat{\alpha} - \alpha) [C - R \hat{\alpha} - S \hat{\beta}] + (\hat{\beta} - \beta) [D - S \hat{\alpha} - T \hat{\beta}] \\ = & 0. \end{aligned}$$

ここで, (6)を用いた. したがって,

$$\sum_{i=1}^n \{Y_i - (A_i \alpha + B_i \beta)\}^2 = \sum_{i=1}^n [Y_i - (A_i \hat{\alpha} + B_i \hat{\beta})]^2 + \sum_{i=1}^n [A_i (\hat{\alpha} - \alpha) + B_i (\hat{\beta} - \beta)]^2$$

となる. これが(7)であった. \square

5 DO-IT-YOURSELF!

A	B	Y	A^2	AB	B^2	AY	BY
1.62	.13	.76					
2.22	.89	1.3					
1.37	-.62	1.05					
2.32	.68	1.05					
2.36	.27	1.26					
2.05	.85	.76					
2.17	2.07	.42					
1.82	.86	1.14					
2.04	-.22	1.9					
1.76	-.72	1.13					
1.98	1.66	.5					
1.79	1.31	.5					
2.12	-.49	1.43					
2.14	.36	1.38					
1.75	-.35	1.09					
2.48	.18	1.98					
1.79	-1.61	1.97					
1.31	.82	.31					
2.04	1.81	.04					
1.61	.99	.41					

問題 1 上の表のデータの, はじめの $n = 2, 5, 10, 20$ 個を用いて, それぞれの場合にパラメータの推定値 $\hat{\alpha}, \hat{\beta}$ を求めよう. n の増加とともに $\hat{\alpha}, \hat{\beta}$ が変化することを観察しよう.

..... MEMO

..... MEMO

6 問題1の答え

実はこのデータは計算機で人工的に発生させたものなので、パラメータの真の値がはじめからわかっている：真値は $\alpha = 0.7$, $\beta = -0.5$ である。以下、データ数 n のときの推定値である。

$$n = 2, \quad \hat{\alpha} = .439993062781824, \quad \hat{\beta} = .363163371488033$$

$$n = 5, \quad \hat{\alpha} = .577546802926555, \quad \hat{\beta} = -.226951312914061,$$

$$n = 10, \quad \hat{\alpha} = .631057769015935, \quad \hat{\beta} = -.392576192424761,$$

$$n = 20, \quad \hat{\alpha} = .646200209711328, \quad \hat{\beta} = -.49090207836115,$$

ちなみに、もっとデータを集めて推定すると次のようになった。

$$n = 100, \quad \hat{\alpha} = .669072660308115, \quad \hat{\beta} = -.513454554131925,$$

$$n = 1000, \quad \hat{\alpha} = .697981340950558, \quad \hat{\beta} = -.506901806862601,$$

データ数が多くなると推定値はパラメータの真の値に非常に近いことがわかる。

..... MEMO