

数理工学実験
テーマ6 最小二乗法

2021年12月27日 提出

工学部情報学科数理工学コース2年
1029-32-7314 岡本淳志

この実験では、最小二乗法（最小二乗誤差推定）についての手法を学ぶ。最小二乗法とは、与えられたデータに合う数理モデルあるいはそのパラメータを決定するために二乗誤差を目的関数とした最小化問題を解いて推定値を得る手法のことである。

また、実験を通して、この最適化問題を構成するデータの取得方法や用いるデータ数、得られた最適解の評価の仕方を工夫することが重要であることを確認する。

1 実験準備

1.1 最小二乗法とその評価方法

次の関係式を満たす方程式が与えられているとする。

$$z = f(\theta, x) \quad (1)$$

ここで、 $z \in \mathbb{R}^m$, $\theta \in \mathbb{R}^p$ とし、 $f: \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ とする。これは、パラメータを θ とし、 x を入力、 z をしゅつりよくするというシステム f である。一般に、観測には誤差を伴う。ここでは出力のみ観測誤差がある場合を考える。

誤差を $w \in \mathbb{R}^m$ とおき、加法的な誤差を考えると、観測値は次で与えられる。

$$y = z + w \quad (2)$$

ただし、観測誤差には偶然誤差と系統誤差の2種類あるが、ここでは偶然誤差のみを考える。また、ここでは偶然誤差は次の性質を持つものと仮定する。

1. 各試行ごとに観測誤差は同じ確率分布 P から確率的に発生する。
2. 入力 x やパラメータ θ とは独立で、観測ごとに独立に定まる。
3. i 回目の観測に対する誤差を $w_i \in \mathbb{R}^m$ で表す。このとき、次の2つが成り立つものとする。

$$\mathbb{E}_P[w_i] = 0, \mathbb{E}_P[w_i w_j^\top] = \delta_{i,j} V \quad (3)$$

ここで 0 は m 次のゼロベクトルを表し、 V は m 次の正定値対称行列である。 $\delta_{i,j}$ は Kronecker のデルタである。

ここで関数 f が既知であるが、パラメータ θ が道である場合を考える。このとき、入力と観測値のデータから θ を決定したい。 θ さえ適切に決められれば、未知のデータに対する出力の予測ができるようになり、様々な場面で恩恵を受けることができるからである。 N 組の入力及び観測値のデータ $\{(x_i, y_i)\}_{i=1}^N$ が与えられているとする。このとき、

$$\min_{\theta \in \mathbb{R}^n} \sum_{i=1}^N \|y_i - f(\theta, x_i)\|^2 \quad (4)$$

を解くことで、真のパラメータ θ^* を得ることを考える。このパラメータを求めることを、 f が θ に関して線形である場合を線形最小二乗法（または単に最小二乗法）と、非線形である場合を非線形最小二乗法と呼ぶ。一般に最小二乗法は、適切な最適化の手法を使って解けばよい。しかし、特別な場合は解を陽に書き下せたり、後からデータが追加されても逐次的に下位を得られる場合があり、データ容量を抑えられるなどメリットがある。

1.1.1 回帰問題

まずは θ が f に対して線形である場合を扱う。このとき、適切な行列値関数 $\varphi: \mathbb{R}^p \rightarrow \mathbb{R}^{m \times n}$ が存在して、

$$f(\theta, x) = \varphi(x)\theta \quad (5)$$

となる。ここで、 $\sum_{i=1}^N \varphi(x_i)^\top \varphi(x_i)$ が正則ならば、最適化問題 (4) の最適解は次で求められる。

$$\hat{\theta}_N = \left(\sum_{i=1}^N \varphi(x_i)^\top \varphi(x_i) \right)^{-1} \sum_{j=1}^N \varphi(x_j)^\top y_j \quad (6)$$

この $\hat{\theta}_N$ を、最小二乗誤差推定値あるいは最小二乗推定値という。このデータをもらったときの処理の仕方のことを、最小二乗誤差推定量という。また、このとき観測誤差 w_i , ($i = 1, \dots, N$) の共分散行列の推定値は、

$$\hat{V}_N = \frac{1}{N-n} \sum_{i=1}^N \left(y_i - \varphi(x_i) \hat{\theta}_N \right) \left(y_i - \varphi(x_i) \hat{\theta}_N \right)^\top \quad (7)$$

で与えられる。

得られたパラメータがどれほどよくデータを説明するかの指標として用いるには向いてない。観測誤差の確率分布が不明であるならば、決定変数と呼ばれる量が用いられる。

$$C = \frac{\sum_{i=1}^N \|\varphi_i \hat{\theta}_N - \bar{y}\|^2}{\sum_{i=1}^N \|y_i - \bar{y}\|^2} \quad (8)$$

ここで、 $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ とした。最小二乗推定値 $\hat{\theta}_N$ の決定変数は $[0, 1]$ の範囲に値を取り、その値が 1 に近いほどよい推定であるとみなされる。逆に 0 に近い場合は得られたパラメータ $\hat{\theta}_N$ が出力 y の推定に役に立たないことを意味する。

1.1.2 重みつき最小二乗法

データ $\{(x_i, y_i)\}_{i=1}^N$ を得たとき、それぞれのデータの信用度を考えて最小二乗法を解きたい場合もある。このとき、一般には重み行列 $Q_i \geq 0$, $i =$

1, 2, ..., N を用いた次の重みつき最小二乗法

$$\min_{\theta \in \mathbb{R}^n} \sum_{i=1}^N (y_i - \varphi_i \theta)^\top Q_i (y_i - \varphi_i \theta) \quad (9)$$

を考えればよい。この場合、 $\sum_{i=1}^N \varphi_i^\top Q_i \varphi_i$ が正則ならば、最適な推定値は

$$\hat{\theta}_N = \left(\sum_{i=1}^N \varphi_i^\top Q_i \varphi_i \right)^{-1} \sum_{j=1}^N \varphi_j^\top Q_j y_j \quad (10)$$

となる。

より一般的には、観測誤差の独立性が仮定できない場合、重み付き最小二乗法を拡張して、次で得られる推定値が BLUE（最良線形不偏推定量）であることが分かる。

$$\hat{\theta}_N = (\varphi^\top V^{-1} \varphi)^{-1} \varphi^\top V^{-1} y^\top \quad (11)$$

1.2 推定値の合成と逐次最小二乗法

1.2.1 異なるデータセットからの推定値の合成

データ $\{(x_i, y_i)\}_{i=1}^N$ と別のデータセット $\{(x'_i, y'_i)\}_{i=1}^M$ からそれぞれ求めた 2 つの推定値を用いて、推定値を更新することができる。

$$\hat{\theta}_{N+M} = (\Phi_N^{-1} + \Phi_M'^{-1})^{-1} (\Phi_N^{-1} \hat{\theta}_N + \Phi_M'^{-1} \hat{\theta}'_M) \quad (12)$$

ただし、

$$\Phi_N = \left(\sum_{i=1}^N \varphi_i^\top \varphi_i \right)^{-1}, \quad \varphi_i = \varphi(x_i) \quad (13)$$

である。

観測誤差の大きさが異なる場合、重み付きの最小二乗法を使った方が推定結果が良くなることが期待される。データ数がそれぞれ N_1, N_2 のデータセットが 2 組与えられているとき、重み行列 Q_1, Q_2 を使ったときのそれぞれのデータセットにおける推定値を $\hat{\theta}_{N_1}, \hat{\theta}_{N_2}$ とおくと、

$$\hat{\theta}_{N_1+N_2} = (\Phi_{Q_1, N_1}^{-1} + \Phi_{Q_2, N_2}^{-1})^{-1} \times (\Phi_{Q_1, N_1}^{-1} \hat{\theta}_{N_1} + \Phi_{Q_2, N_2}^{-1} \hat{\theta}_{N_2}) \quad (14)$$

となる。ここで、

$$\Phi_{Q, N} = \left(\sum_{i=1}^N \varphi_i^\top Q \varphi_i \right)^{-1} \quad (15)$$

とした。

2 問題 1

$\sum_{i=1}^N \varphi_i^\top \varphi_i$ が正則行列であるとき、式 (6) が成り立つことを示せ。

最適化問題 (4) の最適解が (6) となることを示す。

まず、

$$\begin{aligned}
 \|y_i - f(\theta, x_i)\|^2 &= \|y_i - \varphi_i(x)\theta\|^2 \\
 &= (y_i - \varphi_i(x)\theta)^\top (y_i - \varphi_i(x)\theta) \\
 &= (y_i^\top - \theta^\top \varphi_i(x)^\top)(y_i - \varphi_i(x)\theta) \\
 &= \|y_i\|^2 - y_i^\top \varphi_i(x)\theta - \theta^\top \varphi_i(x)^\top y_i + \theta^\top \varphi_i(x)^\top \varphi_i(x)\theta
 \end{aligned} \tag{16}$$

である。

ここで、

$$\min_{\theta \in \mathbb{R}^n} \sum_{i=1}^N \|y_i - f(\theta, x_i)\|^2 \tag{17}$$

を満たす $\theta (= \hat{\theta}_N)$ では、

$$\frac{\partial}{\partial \theta} \sum_{i=1}^N \|y_i - f(\theta, x_i)\|^2 = 0 \tag{18}$$

が成立する。

$$\begin{aligned}
 \frac{\partial}{\partial \theta} \|y_i\|^2 &= 0 \\
 \frac{\partial}{\partial \theta} y_i^\top \varphi_i(x)\theta &= y_i^\top \varphi_i(x) \\
 \frac{\partial}{\partial \theta} \theta^\top \varphi_i(x)^\top y_i &= \varphi_i(x)^\top y_i \\
 \frac{\partial}{\partial \theta} \theta^\top \varphi_i(x)^\top \varphi_i(x)\theta &= \left\{ \left(\varphi_i(x)^\top \varphi_i(x) \right) + \left(\varphi_i(x)^\top \varphi_i(x) \right)^\top \right\} \theta
 \end{aligned} \tag{19}$$

より、

$$\begin{aligned}
 &\frac{\partial}{\partial \theta} \sum_{i=1}^N \|y_i - f(\theta, x_i)\|^2 \\
 &= \sum_{i=1}^N \left[\left\{ \left(\varphi_i(x)^\top \varphi_i(x) \right) + \left(\varphi_i(x)^\top \varphi_i(x) \right)^\top \right\} \theta - y_i^\top \varphi_i(x) - \varphi_i(x)^\top y_i \right] \\
 &= \sum_{i=1}^N \left[\left(\varphi_i(x)^\top \varphi_i(x) \right) \theta - \varphi_i(x)^\top y_i \right] \\
 &= 0
 \end{aligned} \tag{20}$$

が成立し、 $\sum_{i=1}^N \varphi_i^\top \varphi_i$ が正則行列であれば、

$$\hat{\theta}_N = \left(\sum_{i=1}^N \varphi(x_i)^\top \varphi(x_i) \right)^{-1} \sum_{j=1}^N \varphi(x_j)^\top y_j \quad (21)$$

が成立する。

3 課題 8

各入力 $x_i \in \mathbb{R}^2$, $i = 1, 2, \dots, 10000$ に対し、次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする。

$$y_i = x_i^\top \theta + w_i \quad (22)$$

ここで観測誤差は $\mathcal{N}(0, 1)$ に従って発生している。このとき、次の問いを考える。

1. θ の最小二乗推定量を、与えられた全てのデータ $\{(x_i, y_i)\}_{i=1}^{10000}$ を使って求めよ。またそのときの推定誤差共分散行列を求めよ。
2. $N = 2, 4, 8, \dots, 2^{13}$ としたとき、 $\hat{\theta}_N$ の各要素が収束していくことを片対数グラフでプロットして確認せよ。
3. 全てのデータを用いたときの決定変数を求めよ。

3.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「8.c」である。また、問題 2 は gnuplot を用いて図の作成を行なった。

3.2 実験結果

問題 1 の結果は、

$$\hat{\theta}_N = \begin{pmatrix} 1.506551 \\ 1.997696 \end{pmatrix} \quad (23)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 1.5 \\ 2 \end{pmatrix} \quad (24)$$

であるので、高い精度で推定できたと考えられる。

また、推定誤差共分散行列 $\hat{V}_N = 0.998694$ となった。真値が $V = 1$ であるので、これも高い精度で推定できたと考えられる。

問題 2 の結果は、以下の図のようになった。

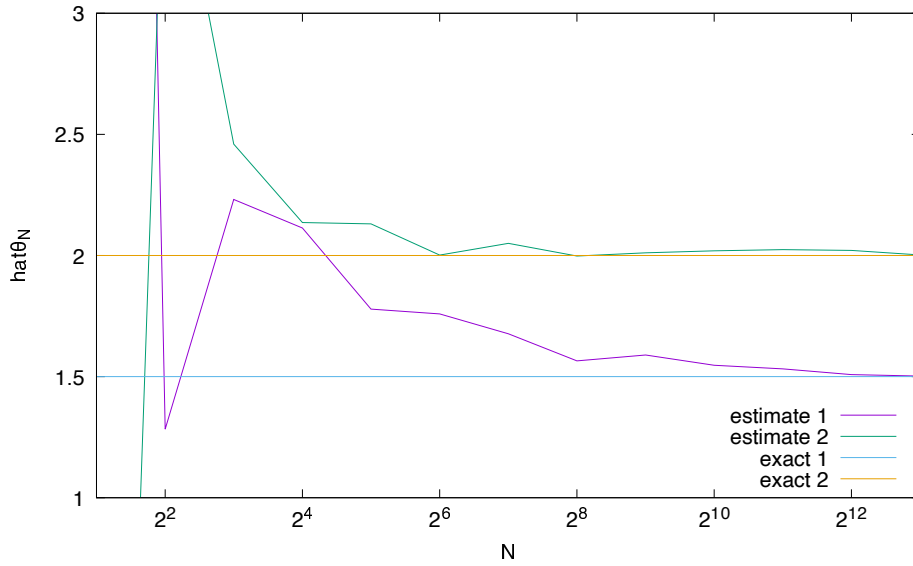


図 1: $N = 2, 4, 8, \dots, 2^{13} = 8192$ の時の $\hat{\theta}_N$, 横軸は $2 \leq N \leq 8192$, 縦軸は $1 \leq \text{hat}\theta_N (= \hat{\theta}_N) \leq 3$ とした

ここで、estimate1, estimate2 はそれぞれ数値計算で求めた $\hat{\theta}_N$ の第 1 成分、第 2 成分を、exact1, exact2 はそれぞれ真値 θ の第 1 成分、第 2 成分を示す。この図から、 N を大きくしていくと $\hat{\theta}_N$ の各要素が収束していく様子が確認できる。第 1 成分は $N = 256$ 以上で、第 2 成分は $N = 4096$ 以上でほぼ真値に近い値を取る様子が確認できた。

問題 3 の結果は、決定変数 $C = 0.862973$ となった。決定変数が 1 に近い値をとったので、これはよい推定であったといえることができる。

4 課題 9

各入力 $x_i \in \mathbb{R}$, $i = 1, 2, \dots, 10000$ に対し、次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする。

$$y_i = \varphi(x_i)\theta + w_i, \quad \varphi(x) = [1 \ x \ x^2 \ x^3] \quad (25)$$

ここで観測誤差は $\mathcal{N}(0, 9)$ に従って発生している。このとき、次の問いを考える。

1. θ の最小二乗推定量を、与えられた全てのデータ $\{(x_i, y_i)\}_{i=1}^{10000}$ を使って求めよ。またそのときの推定誤差共分散行列を求めよ。
2. $N = 4, 8, \dots, 2^{13}$ としたとき、 $\hat{\theta}_N$ の各要素が収束していくことを片対数グラフでプロットして確認せよ。

3. 全てのデータを用いたときの決定変数を求めよ。

4.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「9.c」である。(コード中には、課題 11 で用いる値の計算も含まれている。) また、問題 2 は gnuplot を用いて図の作成を行なった。

4.2 実験結果

問題 1 の結果は、

$$\hat{\theta}_N = \begin{pmatrix} -0.509029 \\ 1.975861 \\ 0.197744 \\ -0.098667 \end{pmatrix} \quad (26)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} -0.5 \\ 2 \\ 0.2 \\ -0.1 \end{pmatrix} \quad (27)$$

であるので、高い精度で推定できたと考えられる。

また、推定誤差共分散行列 $\hat{V}_N = 8.896506$ となった。真値が $V = 9$ であるので、これも高い精度で推定できたと考えられる。

問題 2 の結果は、以下の図のようになった。

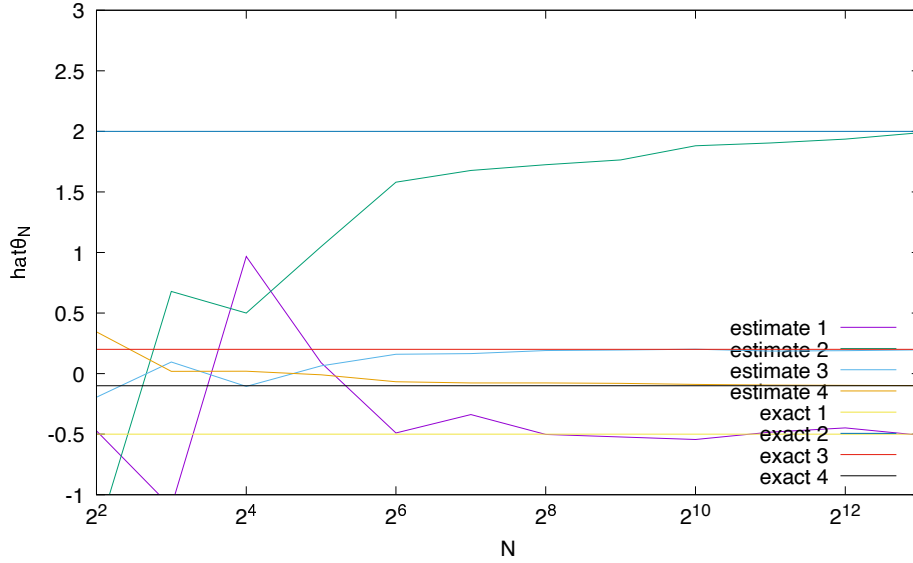


図 2: $N = 4, 8, \dots, 2^{13} = 8192$ の時の $\hat{\theta}_N$, 横軸は $4 \leq N \leq 8192$, 縦軸は $-1 \leq \hat{\theta}_N (= \hat{\theta}_N) \leq 3$ とした

ここで、estimate1,2,3,4 はそれぞれ数値計算で求めた $\hat{\theta}_N$ の第 1,2,3,4 成分を、exact1,2,3,4 はそれぞれ真値 θ の第 1,2,3,4 成分を示す。この図から、 N を大きくしていくと $\hat{\theta}_N$ の各要素が収束していく様子が確認できる。第 2,4 成分の収束は $N = 128$ 以上でほぼ真値に近い値を取る様子が確認できたが、第 1 成分の収束は $N = 256$ 以上で、第 3 成分の収束は $N = 8192$ 以上でようやく真値に近づいていく様子が確認できた。しかしこれは、グラフのスケールによるところが大きいのではないかと考えた。

問題 3 の結果は、決定変数 $C = 0.461855$ となった。この決定変数の値からは、今回の推定がよいものであったかどうかは確認できなかった。

5 課題 10

各入力 $x_i \in \mathbb{R}^2$, $i = 1, 2, \dots, 10000$ に対し、次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする。

$$y_i = x_i^\top \theta + w_i \quad (28)$$

ただし、観測誤差は標準 Cauchy 分布に従うとする。このとき、次の問いを考える。

1. 与えられた全データ $\{(x_i, y_i)\}_{i=1}^{10000}$ を用いて θ の最小二乗誤差推定量を式 (6) から求めよ。

2. $N = 2, 4, 8, \dots, 2^{13}$ としたとき、 $\hat{\theta}_N$ の各要素が収束していかないことを片対数グラフでプロットして確認せよ。

5.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「10.c」である。また、問題 2 は gnuplot を用いて図の作成を行なった。

5.2 実験結果

問題 1 の結果は、

$$\hat{\theta}_N = \begin{pmatrix} 2.373074 \\ 1.537311 \end{pmatrix} \quad (29)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 1.5 \\ 2 \end{pmatrix} \quad (30)$$

であるので、よい推定結果が得られなかった。

また、問題 2 の結果は、以下の図のようになった。

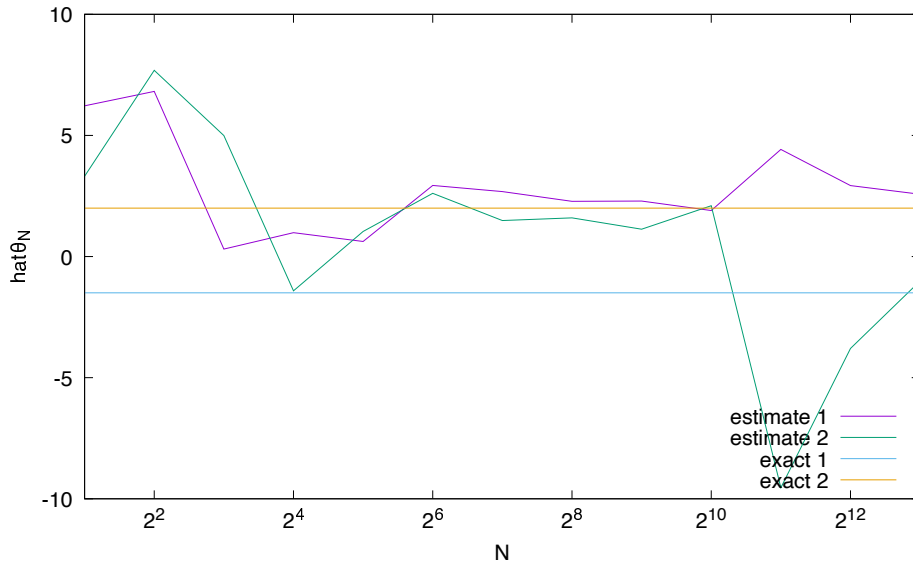


図 3: $N = 2, 4, 8, \dots, 2^{13} = 8192$ の時の $\hat{\theta}_N$, 横軸は $2 \leq N \leq 8192$, 縦軸は $-10 \leq \text{hat}\theta_N (= \hat{\theta}_N) \leq 10$ とした

ここで、estimate1, estimate2 はそれぞれ数値計算で求めた $\hat{\theta}_N$ の第 1 成分、第 2 成分を、exact1, exact2 はそれぞれ真値 θ の第 1 成分、第 2 成分を示す。

この図から、 N の値を大きくしても $\hat{\theta}_N$ の各要素が収束していかないことが確認できた。

これらのことは、観測誤差が従う標準 Cauchy 分布に対して、平均や分散が定義されないことが原因であると考えた。

ちなみに、決定変数の値は、 $C = 0.000284$ となり、0 に近い値をとった。このことから、この問題には最小二乗法を適用する意味がなく、別の推定手法が必要になることが分かった。

6 課題 11

課題 9 と同じ $\varphi(x)$, θ および観測誤差の分布 $\mathcal{N}(0, 9)$ を考える。また、 $x_i \in [0, 1]$, $i = 1, \dots, 1000$ のときに、データ $\{(x_i, y_i)\}_{i=1}^{1000}$ が与えられているものとする。このとき、 $\hat{\theta}_{1000}$ を求め、課題 9.1 の $N = 1000$ の結果と比較せよ。また、どのようにデータを取るべきか比較せよ。

6.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「11.c」(と「9.c」)である。課題 9 のデータは最初の 1000 個 ($N = 1, \dots, 1000$) のものを利用した。

6.2 実験結果

課題 11 に与えられたデータから推定した結果は、

$$\hat{\theta}_N = \begin{pmatrix} -0.577725 \\ 2.096823 \\ -0.395047 \\ 0.549125 \end{pmatrix} \quad (31)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} -0.5 \\ 2 \\ 0.2 \\ -0.1 \end{pmatrix} \quad (32)$$

であるので、ある程度の精度で推定できたと考えられる。

課題 9 に与えられたデータから推定した結果は、

$$\hat{\theta}_N = \begin{pmatrix} -0.549763 \\ 1.888781 \\ 0.204065 \\ -0.089058 \end{pmatrix} \quad (33)$$

となった。同じく与えられた真値が

$$\theta = \begin{pmatrix} -0.5 \\ 2 \\ 0.2 \\ -0.1 \end{pmatrix} \quad (34)$$

であるので、良い精度で推定できたと考えられる。

これらの結果から、今回は課題 9 に与えられたデータからの推定がより良い推定であったと言える。しかし、この 2 つの結果の差は、選んだ課題 9 に与えられたデータに、誤差が与えた影響が偶然小さかったことが大きな原因であると考えられる。ただ、課題 9 では入力 $x_i \in \mathbb{R}$ であるのに対して、課題 11 では入力が $x_i \in [0, 1]$ であり、 $x = 0.001, 0.002, \dots, 1$ と等間隔の値が一つずつ用意されていたので、これも原因の一つではないかと考えた。つまり、大きな誤差が出現すると、入力の範囲が制限されていたり、入力が規則正しく並んでいたりする方が、そうでない場合に比べてよりその影響を受けやすくなるのではないかと考えた。

このことから、データを取る場合はできれば入力の範囲を制限せず、できるだけ広い範囲で、無作為に入力を与えたほうがより良い推定結果を得られるのではないかと考えた。

7 課題 12

$x_i \in \mathbb{R}$, $i = 1, \dots, 1000$ に対し、観測値の次元が 2 次元となる場合を考える。

$$y_i = \varphi(x_i)\theta + w_i \quad (35)$$

ここで w_i は $\mathcal{N}(0, V)$ の独立同一分布に従うとし、

$$\varphi = \begin{bmatrix} 1 & x \\ 1 & x^2 \end{bmatrix}, \quad V = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \quad (36)$$

とする。このとき、式 (6) と式 (10) をそれぞれ用いて推定値を求め、それぞれの推定誤差共分散を求めよ。ただし、 $Q_i = V^{-1}$ とする。また、 $\hat{\theta}_N$ の各要素の収束の仕方をプロットせよ。

7.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「12.c」である。また、 $\hat{\theta}_N$ の各要素の収束の仕方を確認するために、データの数 $N = 100, 200, \dots, 1000$ とした。図の作成には gnuplot を用いた。

7.2 実験結果

式 (6) を用いた場合の推定値は、

$$\hat{\theta}_N = \begin{pmatrix} 1.555099 \\ -0.394144 \end{pmatrix} \quad (37)$$

となった。式 (10) を用いた場合の推定値は、

$$\hat{\theta}_N = \begin{pmatrix} 2.939084 \\ -1.986465 \end{pmatrix} \quad (38)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad (39)$$

であるので、式 (10) を用いた場合は高い精度で推定できたと考えられる。

また、式 (6) を用いた場合の推定誤差共分散行列は、

$$\hat{V}_N = \begin{bmatrix} 107.288426 & -0.018949 \\ -0.018949 & 5.963498 \end{bmatrix} \quad (40)$$

となった。式 (10) を用いた場合の推定誤差共分散行列は、

$$\hat{V}_N = \begin{bmatrix} 102.071225 & 0.081927 \\ 0.081927 & 1.035930 \end{bmatrix} \quad (41)$$

となった。真値が

$$V = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \quad (42)$$

であるので、これも式 (10) を用いた場合の方がより高い精度で推定できたと考えられる。

$\hat{\theta}_N$ の各要素の収束の仕方は、以下の図のようになった。

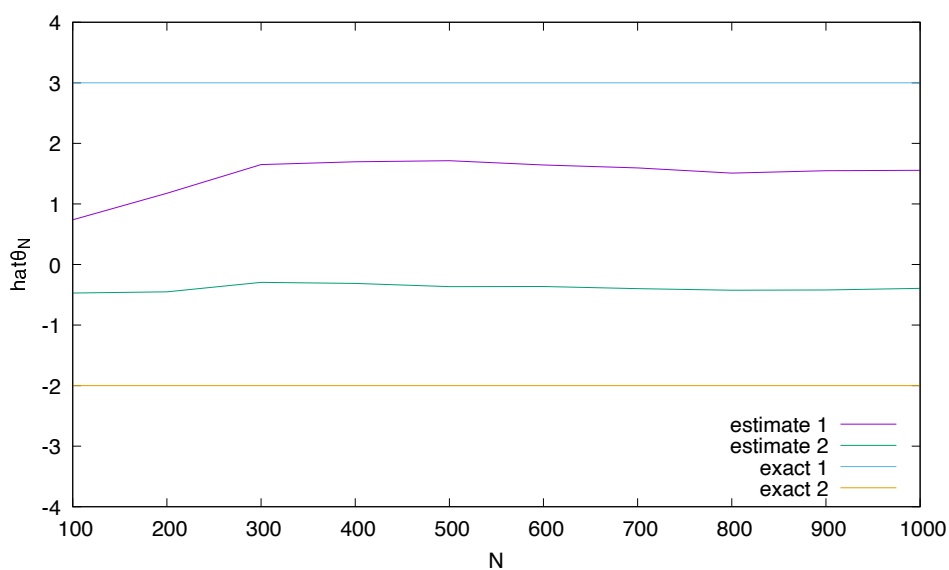


図 4: 式 (6) を用いた場合の、 $N = 100, 200, \dots, 1000$ の時の $\hat{\theta}_N$, 横軸は $100 \leq N \leq 1000$, 縦軸は $-4 \leq \hat{\theta}_N (= \hat{\theta}_N) \leq 4$ とした

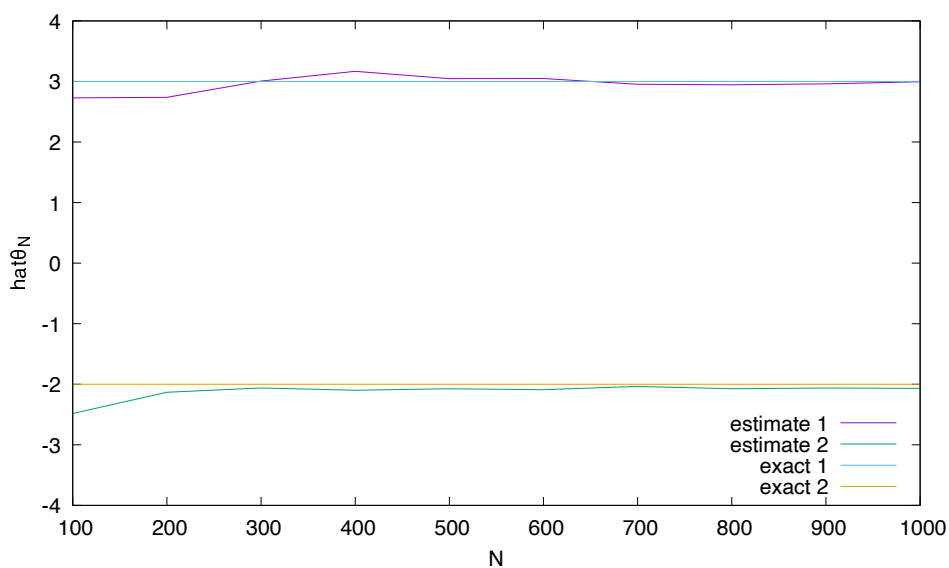


図 5: 式 (10) を用いた場合の、 $N = 100, 200, \dots, 1000$ の時の $\hat{\theta}_N$, 横軸は $100 \leq N \leq 1000$, 縦軸は $-4 \leq \hat{\theta}_N (= \hat{\theta}_N) \leq 4$ とした

これらの図からも、今回の問題では式 (10) を用いた場合の方が明らかに良い推定値を得られることが分かった。

8 課題 13

課題 12 と同様、 $x_i \in \mathbb{R}$, $i = 1, \dots, 1000$ に対し、観測値の次元が 2 次元となる場合を考える。

$$y_i = \varphi(x_i)\theta + w_i \quad (43)$$

ここで w_i は $i = 1, \dots, 500$ は $\mathcal{N}(0, V_1)$ の独立同一分布に、 $i = 501, \dots, 1000$ は $\mathcal{N}(0, V_2)$ の独立同一分布従うとし、

$$\varphi = \begin{bmatrix} 1 & x \\ 1 & x^2 \end{bmatrix}, V_1 = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}, V_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad (44)$$

とする。このとき、式 (6) と式 (10) をそれぞれ用いて推定値を求め、それぞれ収束の仕方をプロットせよ。

8.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「13.c」である。また、 $\hat{\theta}_N$ の各要素の収束の仕方を確認するために、データの数 $N = 100, 200, \dots, 1000$ とした。図の作成には gnuplot を用いた。

8.2 実験結果

式 (6) を用いた場合の推定値は、

$$\hat{\theta}_N = \begin{pmatrix} 1.897692 \\ -0.380164 \end{pmatrix} \quad (45)$$

となった。式 (10) を用いた場合の推定値は、

$$\hat{\theta}_N = \begin{pmatrix} 2.994202 \\ -2.014699 \end{pmatrix} \quad (46)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad (47)$$

であるので、式 (10) を用いた場合は高い精度で推定できたと考えられる。

また、 $\hat{\theta}_N$ の各要素の収束の仕方は、以下の図のようになった。

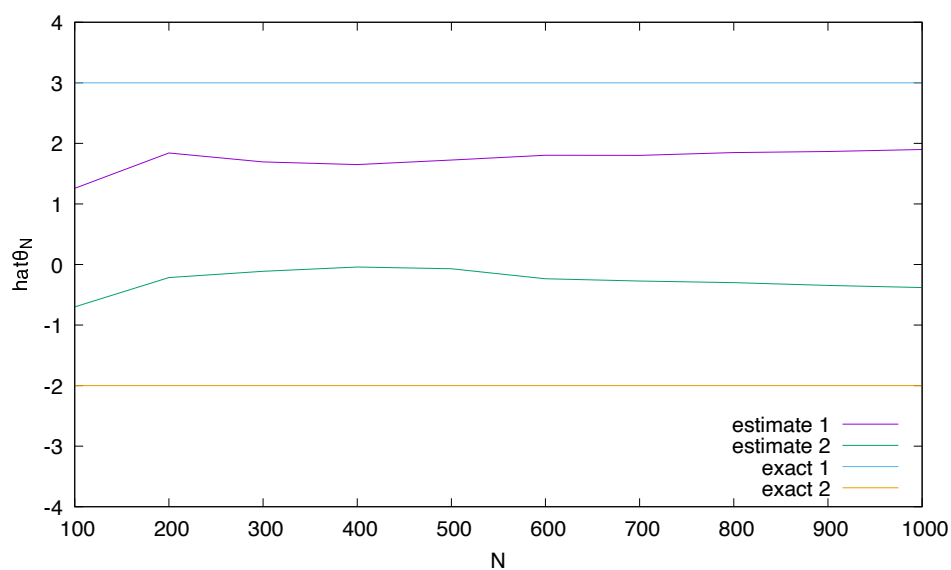


図 6: 式 (6) を用いた場合の、 $N = 100, 200, \dots, 1000$ の時の $\hat{\theta}_N$, 横軸は $100 \leq N \leq 1000$, 縦軸は $-4 \leq \hat{\theta}_N (= \hat{\theta}_N) \leq 4$ とした

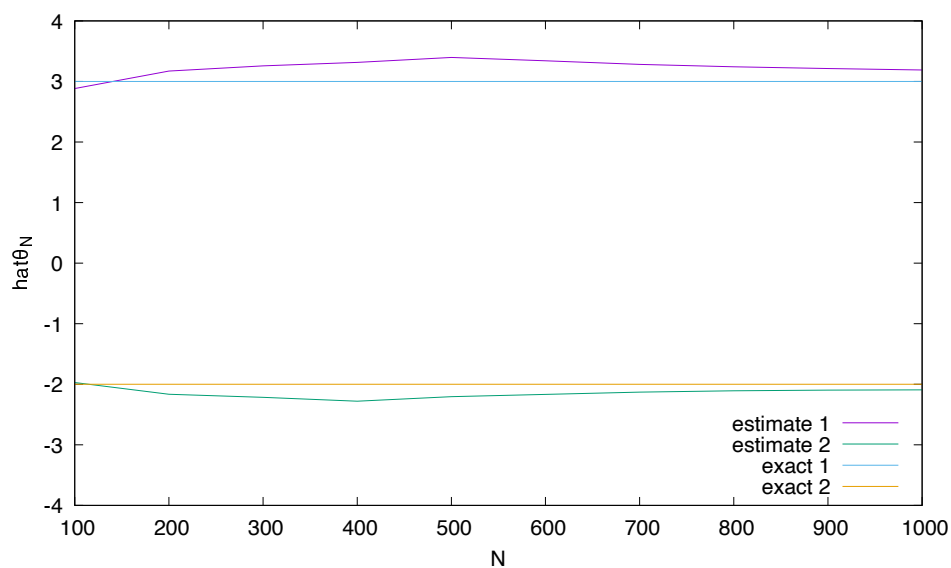


図 7: 式 (10) を用いた場合の、 $N = 100, 200, \dots, 1000$ の時の $\hat{\theta}_N$, 横軸は $100 \leq N \leq 1000$, 縦軸は $-4 \leq \hat{\theta}_N (= \hat{\theta}_N) \leq 4$ とした

これらの図からも、課題 12 と同様、今回の問題では式 (10) を用いた場合の方が明らかに良い推定値を得られることが分かった。

9 課題 14

$x_i \in \mathbb{R}$, $i = 1, \dots, 10000$ に対し、次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする。

$$y_i = \varphi(x_i)\theta + w_i \quad (48)$$

ここで

$$\varphi(x) = \begin{bmatrix} 1 & \exp(-\frac{(x_i - 1)^2}{2}) & \exp(-(x_i + 1)^2) \end{bmatrix} \quad (49)$$

であり、 w_i は平均 0、分散有限の独立同一分布から生成されているものとする。10000 組のデータの組 $\{(x_i, y_i)\}_{i=1}^{10000}$ のうち、最初の 6000 組と残りの 4000 組のデータそれぞれで $\theta \in \mathbb{R}^3$ の推定値を求め、それらから推定値を合成せよ。また、全データを使った場合の推定値と比較し、一致することを確認せよ。

9.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「14.c」である。

9.2 実験結果

まず、全データを使った場合の推定値 ($= \hat{\theta}_N$ とする) は、

$$\hat{\theta}_N = \begin{pmatrix} -0.012495 \\ 3.020430 \\ -1.984869 \end{pmatrix} \quad (50)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix} \quad (51)$$

であるので、高い精度で推定できたと考えられる。

次に、最初の 6000 組のデータを使った場合の推定値と残りの 4000 組のデータを使った場合の推定値は、それぞれ $\hat{\theta}_{N_1}$, $\hat{\theta}_{N_2}$ とすると、

$$\hat{\theta}_{N_1} = \begin{pmatrix} -0.003879 \\ 3.010715 \\ -1.989434 \end{pmatrix}, \quad \hat{\theta}_{N_2} = \begin{pmatrix} -0.025376 \\ 3.034899 \\ -1.977727 \end{pmatrix} \quad (52)$$

となった。この $\hat{\theta}_{N_1}$, $\hat{\theta}_{N_2}$ と、式 (12) を用いて推定値を合成すると、

$$\hat{\theta}_{N_1+N_2} = \begin{pmatrix} -0.012495 \\ 3.020430 \\ -1.984869 \end{pmatrix} \quad (53)$$

となり、全データを使った場合の推定値と一致することが確認できた。

10 課題 15

$x_i \in \mathbb{R}$, $i = 1, \dots, 10000$ に対し、次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする。

$$y_i = \varphi(x_i)\theta + w_i \quad (54)$$

ここで w_i は最初の 6000 組と残りの 4000 組のデータで、平均 0 の異なる独立な分布に従うものとする。

$$\varphi(x) = \begin{bmatrix} 1 & \exp(-\frac{(x_i - 1)^2}{2}) & \exp(-(x_i + 1)^2) \end{bmatrix} \quad (55)$$

である。10000 組のデータの組 $\{(x_i, y_i)\}_{i=1}^{10000}$ のうち、最初の 6000 組と残りの 4000 組のデータそれぞれで $\theta \in \mathbb{R}^3$ の推定値と偶然誤差 w_i の分散の推定値を求め、推定された分散を用いて推定値を合成せよ。また、全データを使い、式 (6) を用いて得た $\hat{\theta}_{10000}$ と、どちらが優れているか比較考察せよ。

10.1 実験方法

C 言語を使ってこの実験を行なった。用いたソースコードは添付の「15.c」である。

10.2 実験結果

まず、全データを使った場合の推定値 ($= \hat{\theta}_{10000}$ とする) は、

$$\hat{\theta}_{10000} = \begin{pmatrix} 0.043732 \\ 3.208666 \\ -2.151179 \end{pmatrix} \quad (56)$$

となった。与えられた真値が

$$\theta = \begin{pmatrix} 0.1 \\ 3.1 \\ -2.1 \end{pmatrix} \quad (57)$$

であるので、良い精度で推定できたと考えられる。

次に、最初の 6000 組のデータを使った場合の推定値と残りの 4000 組のデータを使った場合の推定値は、それぞれ $\hat{\theta}_{N_1}$, $\hat{\theta}_{N_2}$ とすると、

$$\hat{\theta}_{N_1} = \begin{pmatrix} 0.007077 \\ 3.280543 \\ -2.190900 \end{pmatrix}, \quad \hat{\theta}_{N_2} = \begin{pmatrix} 0.098483 \\ 3.100405 \\ -2.091002 \end{pmatrix} \quad (58)$$

となった。また、これらの $\hat{\theta}_{N_1}$, $\hat{\theta}_{N_2}$ を用いて、偶然誤差 w_i の分散の推定値（それぞれ、 \hat{V}_{N_1} , \hat{V}_{N_2} とする）を求めると、

$$\hat{V}_{N_1} = 96.863294, \quad \hat{V}_{N_2} = 0.010304 \quad (59)$$

となった。

これらの \hat{V}_{N_1} , \hat{V}_{N_2} と、式 (14) を用いて推定値を合成すると、

$$\hat{\theta}_{N_1+N_2} = \begin{pmatrix} 0.098516 \\ 3.098826 \\ -2.089009 \end{pmatrix} \quad (60)$$

となった。

ここで、

$$\|\hat{\theta}_{10000} - \theta\|^2 \doteq 1.76 \times 10^{-1}, \quad \|\hat{\theta}_{N_1+N_2} - \theta\|^2 \doteq 1.24 \times 10^{-3} \quad (61)$$

より、後者の方が優れた手法であると言える。 \hat{V}_{N_1} , \hat{V}_{N_2} を見て分かるように、最初の 6000 組のデータは残りの 4000 組のデータに比べてはるかに偶然誤差 w_i の分散の値が大きく、全てのデータを用いて θ の推定値を求めるとその結果に大きく影響を与えることが原因であると考えられる。よってその分散の大きさを考慮した式 (14) の方がより優れた手法となっている。

11 課題 16

11.1 実験方法

11.2 実験結果

12 課題 18

12.1 実験方法

12.2 実験結果

13 付録

14 参考文献

- 数理工学実験テキスト
- Cloud LaTeX
<https://cloudlatex.io>
- LaTeX コマンド一覧 (リスト)
<https://medemanabu.net/latex/latex-commands-list/>
- TeX で複素数 C 、実数 R 、有理数 Q 、整数 Z 、自然数 N などの白抜き文字を使う方法
<http://did2.blog64.fc2.com/blog-entry-204.html>
- LaTeX で URL を入力する方法
<https://qiita.com/hirooooooaki/items/84d45c83c6bcb1d52b79>
- 行列と行列式 - matri, pmatrix, bmatrix
<https://medemanabu.net/latex/matrix-pmatrix-bmatrix/>