数理工学実験レポート

第 3 章（最小二乗法）

学籍番号 1029366161　中塚一瑳

2025 年 10 月 28 日

概要

目次

1 はじめに 2

2 最小二乗法とその評価方法 2

2.1 原理・方法 2

2.2 課題 1（重回帰） 3

2.3 課題 2（多項式回帰） 6

2.4 課題 3（分散の観測誤差：Cauchy） 9

2.5 課題 4（入力域の制約と設計） 11

3 13

4 アルゴリズム比較 13

5 結論 13

付録 A 使用コード一覧 13

付録 B 補足導出 13

1 はじめに

本実験第 2 回では，最小二乗法の基礎とその実装手法を学ぶことを目的とする。具体的には，観測データからのパラメータ推定，重み付き・逐次最小二乗法，データ分割・推定値の合成を通して，推定精度と計算効率の違いを実験的に比較する。

2 最小二乗法とその評価方法

2.1 原理・方法

観測モデルを

$$y_i = f(\theta, x_i) + w_i \tag{1}$$

とする．加法雑音 w_i は平均 0，観測ごとに独立，同一分布であり，共分散は有限と仮定する．ここでは $f(\theta, x) = \phi(x)\theta$ としてパラメータに対して線形とし，最小二乗問題

$$\min_{\theta} \sum_{i=1}^{N} \|y_i - \phi(x_i)\theta\|^2 \tag{2}$$

を解く．解は

$$\hat{\theta}_N = \Big(\sum_{i=1}^{N} \phi_i^\top \phi_i\Big)^{-1} \sum_{i=1}^{N} \phi_i^\top y_i, \qquad \phi_i := \phi(x_i). \tag{3}$$

また雑音分散が未知の場合の推定量および推定誤差共分散は

$$\hat{\sigma}^2 = \frac{1}{N-n} \sum_{i=1}^{N} \|y_i - \phi_i \hat{\theta}_N\|^2, \tag{4}$$

$$\widehat{\mathrm{Cov}}(\hat{\theta}_N) = \hat{\sigma}^2 \Big(\sum_{i=1}^{N} \phi_i^\top \phi_i\Big)^{-1} \tag{5}$$

で与えられる．

当てはまりの評価には決定係数

$$C = \frac{\sum_{i=1}^{N} \|\phi_i \hat{\theta}_N - \bar{y}\|^2}{\sum_{i=1}^{N} \|y_i - \bar{y}\|^2}, \qquad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{6}$$

を用いる．以上は資料 3.2 の線形最小二乗および評価に対応する．

本実験では $\phi(x)$ として設計した特徴量から行列

$$X = \begin{bmatrix} \phi(x_1)^\top \\ \vdots \\ \phi(x_N)^\top \end{bmatrix} \tag{7}$$

を構成し，観測データを

$$y = [y_1; \dots; y_N] \tag{8}$$

として用いる．パラメータ推定量は

$$\hat{\theta} = (X^\top X)^{-1} X^\top y \tag{9}$$

で与えられる．

残差は

$$r = y - X\hat{\theta} \tag{10}$$

とし，雑音分散推定量は

$$\hat{\sigma}^2 = \frac{\|r\|^2}{N - p} \tag{11}$$

（p は列数）とする．

パラメータ推定の誤差共分散は

$$\hat{\sigma}^2 (X^\top X)^{-1} \tag{12}$$

で与えられる．

決定係数は

$$C = \frac{\|X\hat{\theta} - \bar{y}\mathbf{1}\|^2}{\|y - \bar{y}\mathbf{1}\|^2} \tag{13}$$

と定義する．

以下の R 関数が上記推定を実装している．

```
regression_simple <- function(x, y){
    theta_hat  <- solve(t(x) %*% x) %*% t(x) %*% y
    sigma2_hat <- as.numeric(t(y - x %*% theta_hat) %*%
                             (y - x %*% theta_hat) /
                             (nrow(x) - ncol(x)))
    err_cov_mat <- sigma2_hat * solve(t(x) %*% x)
    det_coef <- sum((x %*% theta_hat - mean(y))^2) /
                sum((y - mean(y))^2)
    list(theta_hat=theta_hat, err_cov_mat=err_cov_mat,
         sigma2_hat=sigma2_hat, det_coef=det_coef)
}
```

2.2 課題 1（重回帰）

■モデル $y_i = x_i^\top \theta + w_i$ $(i = 1, \ldots, N)$．$x_i \in \mathbb{R}^2$．w_i は独立，同分散 σ^2，平均 0．$X = [x_1^\top; \ldots; x_N^\top] \in \mathbb{R}^{N \times 2}$，$y = [y_1; \ldots, y_N] \in \mathbb{R}^N$．

■推定量　最小二乗推定量

$$\hat{\theta}_N = (X^\top X)^{-1} X^\top y.$$

残差 $r = y - X\hat{\theta}_N$ により

$$\hat{\sigma}^2 = \frac{\|r\|_2^2}{N-p}, \quad p = 2, \qquad \widehat{\mathrm{Cov}}(\hat{\theta}_N) = \hat{\sigma}^2 \, (X^\top X)^{-1}.$$

決定係数

$$R^2 = \frac{\|X\hat{\theta}_N - \bar{y}\mathbf{1}\|_2^2}{\|y - \bar{y}\mathbf{1}\|_2^2}, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

■収束確認　$N \in \{2, 4, 8, \ldots, 2^{13} = 8192\}$ で $\hat{\theta}_N$ を計算し，N を横軸とする片対数図で各成分を同一図に描く．

■結果（全データ $N = 10000$**）**

$$\hat{\theta}_N = \begin{bmatrix} 1.506551 \\ 1.997696 \end{bmatrix}, \qquad \widehat{\mathrm{Cov}}(\hat{\theta}_N) = \begin{bmatrix} 9.866491 \times 10^{-5} & -4.081657 \times 10^{-7} \\ -4.081657 \times 10^{-7} & 1.005248 \times 10^{-4} \end{bmatrix}.$$

決定係数

$$R^2 = 0.8629734.$$

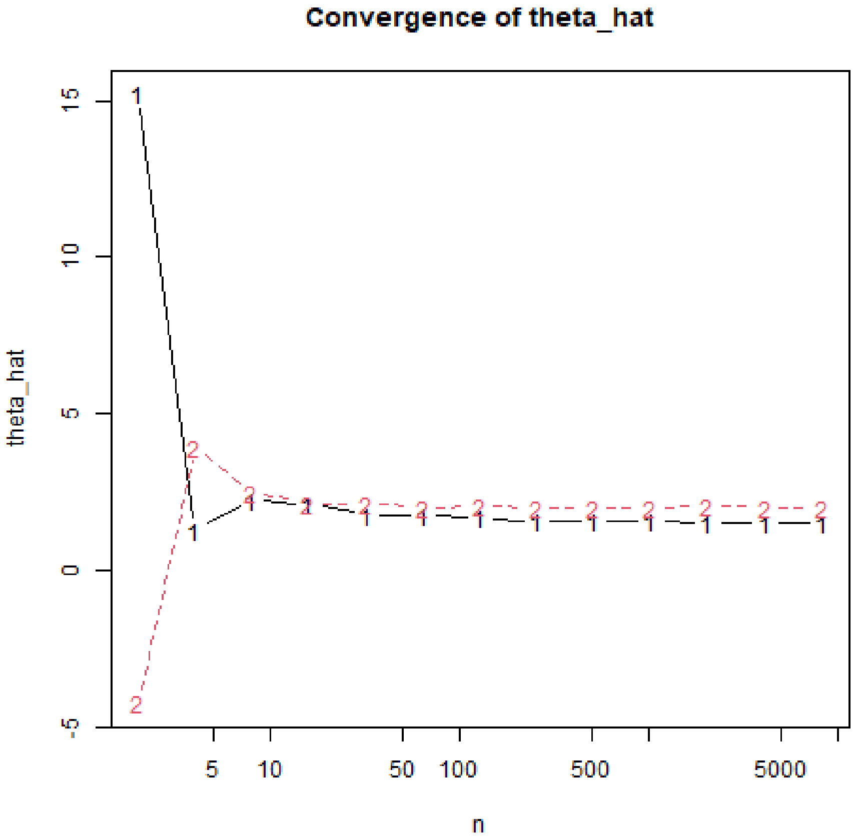

図 1: $\hat{\theta}_N$ の収束（横軸 $N = 2, 4, \ldots, 8192$ の片対数）

■R コード

```
# データ読み込み
data <- read.csv("datas/mmse_kadai1.csv", header=FALSE,
                 col.names = c("x1","x2","y"))
x <- as.matrix(data[, c("x1","x2")])
y <- as.matrix(data[, "y"])

# 最小二乗（原理・方法で用いる関数）
regression_simple <- function(x, y){
  theta_hat  <- solve(t(x) %*% x) %*% t(x) %*% y
  sigma2_hat <- as.numeric(t(y - x %*% theta_hat) %*%
                           (y - x %*% theta_hat) / (nrow(x) - ncol(x)))
  err_cov_mat <- sigma2_hat * solve(t(x) %*% x)
  det_coef <- sum((x %*% theta_hat - mean(y))^2) /
              sum((y - mean(y))^2)
  list(theta_hat=theta_hat, err_cov_mat=err_cov_mat,
       sigma2_hat=sigma2_hat, det_coef=det_coef)
}

# 部分データでの実験
exp1 <- function (x, y, n){
  x <- x[1:n, , drop=FALSE]
  y <- y[1:n, , drop=FALSE]
  result <- regression_simple(x, y)
  list(theta_hat = result$theta_hat,
       err_cov_mat = result$err_cov_mat,
       det_coef = result$det_coef)
}

# 収束図の作成
plot_exp1 <- function(x, y, out="graphs/task1.png"){
  ns <- 2^(1:13)  # 2,...,8192
  theta_hats <- matrix(NA_real_, nrow=length(ns), ncol=ncol(x))
  for(i in seq_along(ns)){
    theta_hats[i, ] <- as.vector(exp1(x, y, ns[i])$theta_hat)
  }
  dir.create(dirname(out), showWarnings=FALSE, recursive=TRUE)
  png(out, width=960, height=600, res=120)
  matplot(ns, theta_hats, type="b", log="x",
```

```
            xlab="N", ylab=expression(hat(theta)),
            main="Convergence of OLS estimates")
  legend("bottomright",
         legend=paste0("theta[", 1:ncol(x), "]"),
         lty=1:ncol(x), pch=1:ncol(x))
  dev.off()
}

# 実行例（全データ）
N_max <- nrow(x)
th    <- exp1(x, y, N_max)$theta_hat
Vhat <- exp1(x, y, N_max)$err_cov_mat
R2    <- exp1(x, y, N_max)$det_coef
print(th); print(Vhat); print(R2)

# 収束図
plot_exp1(x, y)
```

■考察

2.3　課題 2（多項式回帰）

■モデル　$y_i = \varphi(x_i)\theta + w_i, \quad \varphi(x) = [1\ x\ x^2\ x^3],\ i = 1, \ldots, N.$ 雑音 w_i は独立，同分散 σ^2，平均 0．$X = [\varphi(x_1); \ldots; \varphi(x_N)] \in \mathbb{R}^{N\times 4}$.

■推定量　推定量・共分散・決定係数は課題 1 と同じ形式（$p = 4$）で計算する．

■収束確認　$N \in \{4, 8, 16, \ldots, 8192\}$ で $\hat{\theta}_N$ を計算し，N を横軸とする片対数図で各成分を同一図に描画する．

■結果（全データ $N = 10000$）

$$\hat{\theta}_N = \begin{bmatrix} -0.50902942 \\ 1.97586067 \\ 0.19774405 \\ -0.09866691 \end{bmatrix}, \qquad \widehat{\mathrm{Cov}}(\hat{\theta}_N) = \begin{bmatrix} 2.023442\times 10^{-3} & -1.186497\times 10^{-5} & -1.348312\times 10^{-4} & 3.97116 \\ -1.186497\times 10^{-5} & 6.753015\times 10^{-4} & -1.898545\times 10^{-7} & -3.79471 \\ -1.348312\times 10^{-4} & -1.898545\times 10^{-7} & 1.603910\times 10^{-5} & 6.35182 \\ 3.971165\times 10^{-7} & -3.794719\times 10^{-5} & 6.351820\times 10^{-8} & 2.52871 \end{bmatrix}$$

標準誤差（対角の平方根）：$\mathrm{SE}(\hat{\theta}_0) \approx 0.0450$, $\mathrm{SE}(\hat{\theta}_1) \approx 0.0260$, $\mathrm{SE}(\hat{\theta}_2) \approx 0.00400$, $\mathrm{SE}(\hat{\theta}_3) \approx 0.00159$.

決定係数：

$$R^2 = 0.461855.$$

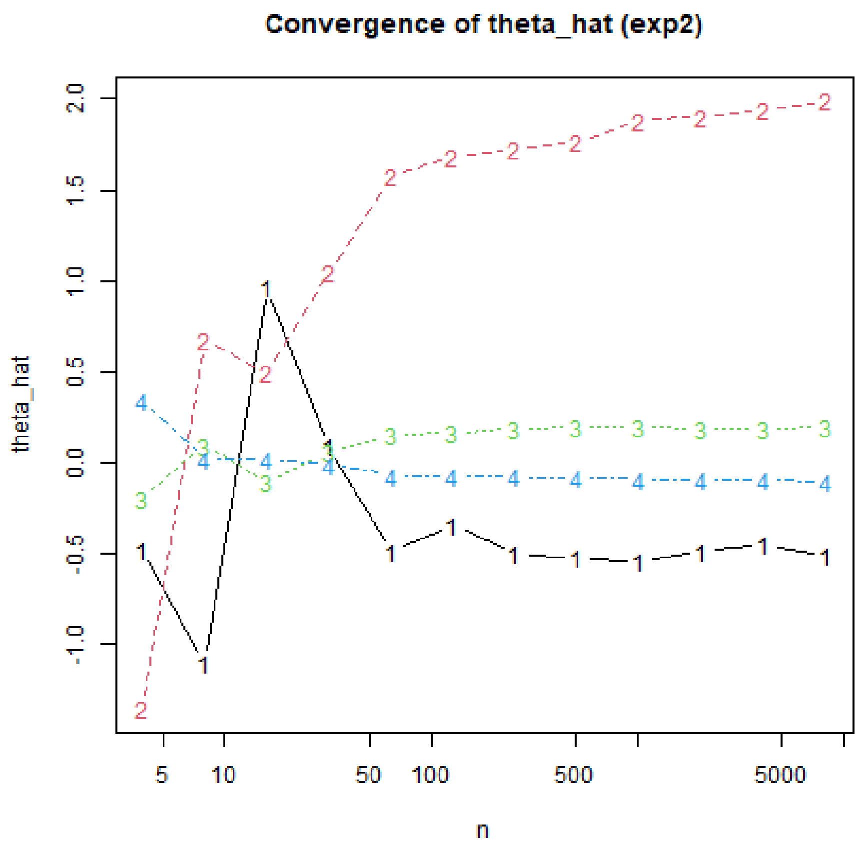

図 2: $\hat{\theta}_N$ の収束（横軸 $N = 4, 8, \ldots, 8192$ の片対数）

■R コード

```
# データ
data <- read.csv("datas/mmse_kadai2.csv", header=FALSE,
                 col.names=c("x1","y"))
x0 <- rep(1, nrow(data))
x1 <- as.matrix(data[,"x1"])
x2 <- x1^2; x3 <- x1^3
x  <- cbind(x0, x1, x2, x3)
y  <- as.matrix(data[,"y"])

# OLS 基本関数（課題 1 と同じ）
regression_simple <- function(x, y){
  theta_hat  <- solve(t(x) %*% x) %*% t(x) %*% y
  sigma2_hat <- as.numeric(t(y - x %*% theta_hat) %*%
                          (y - x %*% theta_hat) / (nrow(x) - ncol(x)))
  err_cov_mat <- sigma2_hat * solve(t(x) %*% x)
  det_coef <- sum((x %*% theta_hat - mean(y))^2) /
```

```
             sum((y - mean(y))^2)
  list(theta_hat=theta_hat, err_cov_mat=err_cov_mat,
       sigma2_hat=sigma2_hat, det_coef=det_coef)
}

# 部分データ実験
exp2 <- function(x, y, n){
  x <- x[1:n, , drop=FALSE]
  y <- y[1:n, , drop=FALSE]
  result <- regression_simple(x, y)
  list(theta_hat=result$theta_hat,
       err_cov_mat=result$err_cov_mat,
       det_coef=result$det_coef)
}

# 収束図
plot_exp2 <- function(x, y, out="graphs/task2.png"){
  ns <- 2^(2:13)  # 4,...,8192
  theta_hats <- matrix(NA_real_, nrow=length(ns), ncol=ncol(x))
  for(i in seq_along(ns)){
    theta_hats[i, ] <- as.vector(exp2(x, y, ns[i])$theta_hat)
  }
  dir.create(dirname(out), showWarnings=FALSE, recursive=TRUE)
  png(out, width=960, height=600, res=120)
  matplot(ns, theta_hats, type="b", log="x",
          xlab="N", ylab=expression(hat(theta)),
          main="Convergence of OLS estimates (poly degree 3)")
  legend("bottomright",
         legend=paste0("theta[",0:(ncol(x)-1),"]"),
         lty=1:ncol(x), pch=1:ncol(x))
  dev.off()
}

# 実行例
N_max <- nrow(x)
print(exp2(x, y, N_max)$theta_hat)
print(exp2(x, y, N_max)$err_cov_mat)
print(exp2(x, y, N_max)$det_coef)
plot_exp2(x, y)
```

■考察

2.4 課題 3（分散の観測誤差：Cauchy）

■モデルと注意 $y_i = x_i^\top \theta + w_i \ (i = 1, \dots, N)$，$x_i \in \mathbb{R}^2$．観測誤差 $w_i \sim \mathrm{Cauchy}(0,1)$ 独立同分布．Cauchy は二乗可積分でないため $\mathbb{E}[w_i^2]$ が存在せず，最小二乗法の通常仮定（有限分散と大数の法則）は満たされない。よって σ^2 や $\mathrm{Cov}(\hat{\theta})$ は理論上定義できない。以下の分散・標準誤差・R^2 は便宜的な値であり，統計的保証はない。

■推定（形式的な OLS） 推定量・共分散・決定係数は課題 1 と同じ形式（$p = 2$）で計算する．ただし観測誤差が Cauchy 分布であり理論保証がないことに注意.

■結果（全データ $N = 10000$）

$$\hat{\theta}_N = \begin{bmatrix} 2.373074 \\ 1.537311 \end{bmatrix}, \qquad \widehat{\mathrm{Cov}}(\hat{\theta}_N) = \begin{bmatrix} 9.234588215 & -0.003642121 \\ -0.003642121 & 9.264075136 \end{bmatrix}.$$

参考：対角の平方根は $\mathrm{SE}(\hat{\theta}_1) \approx 3.0388$，$\mathrm{SE}(\hat{\theta}_2) \approx 3.0437$．決定係数（参考値）は

$$R^2 = 0.0002841468.$$

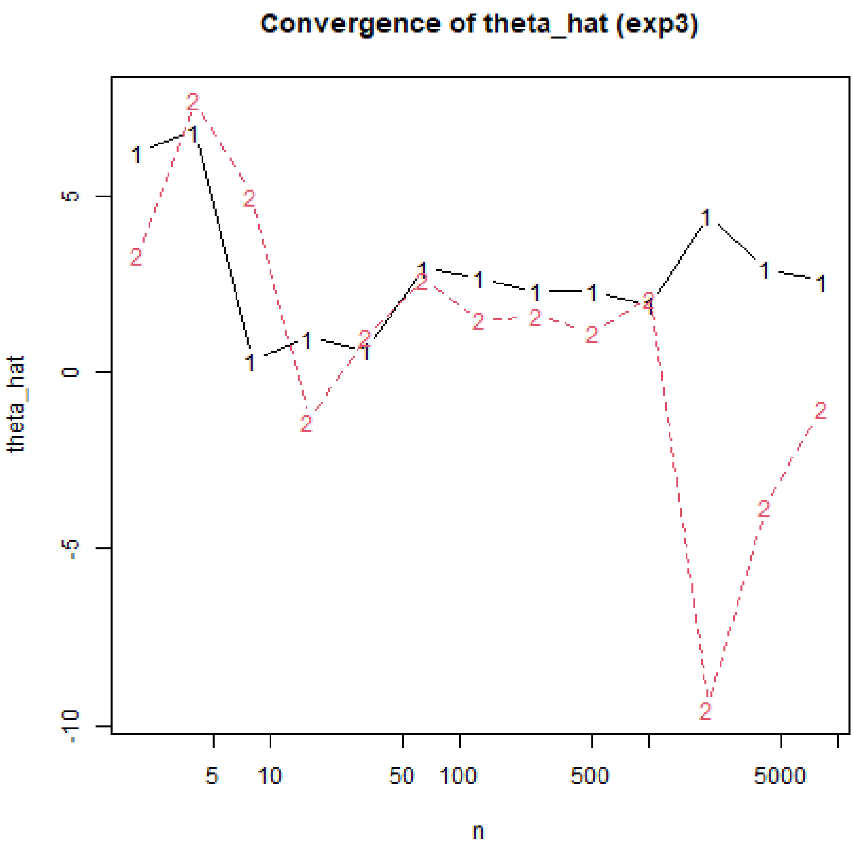

図 3: $\hat{\theta}_N$ の非収束例（横軸 $N = 2, 4, \dots, 8192$ の片対数）

■R コード

```
# データ読込
data3 <- read.csv("datas/mmse_kadai3.csv", header=FALSE,
                  col.names=c("x1","x2","y"))
x <- as.matrix(data3[, c("x1","x2")])
y <- as.matrix(data3[, "y"])

# 形式的な OLS（課題 1 と同様）
regression_simple <- function(x, y){
  theta_hat  <- solve(t(x) %*% x) %*% t(x) %*% y
  sigma2_hat <- as.numeric(t(y - x %*% theta_hat) %*%
                           (y - x %*% theta_hat) / (nrow(x) - ncol(x)))
  err_cov_mat <- sigma2_hat * solve(t(x) %*% x)
  det_coef <- sum((x %*% theta_hat - mean(y))^2) /
              sum((y - mean(y))^2)
  list(theta_hat=theta_hat, err_cov_mat=err_cov_mat,
       sigma2_hat=sigma2_hat, det_coef=det_coef)
}

# 部分データでの実験
exp3 <- function(x, y, n){
  x <- x[1:n, , drop=FALSE]
  y <- y[1:n, , drop=FALSE]
  res <- regression_simple(x, y)
  list(theta_hat=res$theta_hat,
       err_cov_mat=res$err_cov_mat,
       det_coef=res$det_coef)
}

# 収束図
plot_exp3 <- function(x, y, out="graphs/task3.png"){
  ns <- 2^(1:13)                      # 2,...,8192
  ns <- ns[ns <= nrow(x)]
  theta_hats <- matrix(NA_real_, nrow=length(ns), ncol=ncol(x))
  for(i in seq_along(ns)){
    theta_hats[i, ] <- as.vector(exp3(x, y, ns[i])$theta_hat)
  }
  dir.create(dirname(out), showWarnings=FALSE, recursive=TRUE)
  png(out, width=960, height=600, res=120)
```

```
  matplot(ns, theta_hats, type="b", log="x",
          xlab="N", ylab=expression(hat(theta)),
          main="Convergence of OLS under Cauchy noise")
  legend("topright", legend=paste0("theta[",1:ncol(x),"]"),
         lty=1:ncol(x), pch=1:ncol(x))
  dev.off()
}

# 実行例
N_max <- nrow(x)
print(exp3(x, y, N_max)$theta_hat)
print(exp3(x, y, N_max)$err_cov_mat)
print(exp3(x, y, N_max)$det_coef)
plot_exp3(x, y)
```

■考察　資料にあるように、Cauchy 誤差では外れ値の影響が支配的で，$\hat{\theta}_N$ は N を増やしても安定しにくいことが確認できた。

2.5　課題 4（入力域の制約と設計）

■設定　課題 2 と同じ $\varphi(x) = [\,1\ x\ x^2\ x^3\,]$, 真の θ, 観測誤差 $w_i \sim \mathcal{N}(0, 9)$ とする．ただし入力は $x_i \in [0, 1]$, $i = 1, \ldots, 10000$.　$X = [\varphi(x_1); \ldots; \varphi(x_N)] \in \mathbb{R}^{N\times 4}$, $y = [y_1; \ldots; y_N]$.

■推定量　推定量・共分散・決定係数は課題 1 と同じ形式（$p = 4$）で計算する．

■R コード

```
# 課題 4: x ∈ [0,1], φ (x)=[1 x x^2 x^3]
data <- read.csv("datas/mmse_kadai4.csv", header=FALSE,
                 col.names=c("x1","y"))
x0 <- rep(1, nrow(data))
x1 <- as.matrix(data[,"x1"])
x2 <- x1^2; x3 <- x1^3
x  <- cbind(x0, x1, x2, x3)
y  <- as.matrix(data[,"y"])

regression_simple <- function(x, y){
  theta_hat  <- solve(t(x) %*% x) %*% t(x) %*% y
  sigma2_hat <- as.numeric(t(y - x %*% theta_hat) %*%
                           (y - x %*% theta_hat) / (nrow(x) - ncol(x)))
  err_cov_mat <- sigma2_hat * solve(t(x) %*% x)
  det_coef <- sum((x %*% theta_hat - mean(y))^2) /
```

```
              sum((y - mean(y))^2)
  list(theta_hat=theta_hat, err_cov_mat=err_cov_mat,
       sigma2_hat=sigma2_hat, det_coef=det_coef)
}

exp4 <- function(x, y, n){
  x <- x[1:n, , drop=FALSE]
  y <- y[1:n, , drop=FALSE]
  regression_simple(x, y)
}

N_max <- nrow(x)
res4 <- exp4(x, y, N_max)
print(res4$theta_hat)
print(res4$err_cov_mat)
print(res4$det_coef)   # R^2
```

■結果（$N = 10000$）

$$\hat{\theta}_{(4)} = \begin{bmatrix} \boxed{\text{ここに数値}} \\ \boxed{\text{ここに数値}} \\ \boxed{\text{ここに数値}} \\ \boxed{\text{ここに数値}} \end{bmatrix}, \quad \widehat{\mathrm{Cov}}(\hat{\theta}_{(4)}) = \boxed{\text{ここに } 4 \times 4 \text{ 行列}}, \quad R^2 = \boxed{\text{ここに数値}}.$$

（上の枠に，直上の R 出力を転記）

■課題 2.1 との比較（$N = 10000$） 課題 2.1 の推定結果：

$$\hat{\theta}_{(2.1)} = \begin{bmatrix} -0.50902942 \\ 1.97586067 \\ 0.19774405 \\ -0.09866691 \end{bmatrix}, \quad R^2 = 0.461855.$$

評価観点：

- 分散は $\mathrm{Var}(\hat{\theta}) = \sigma^2(X^\top X)^{-1}$ に比例．$x \in [0,1]$ では列 $\{1, x, x^2, x^3\}$ が強く相関しやすく，$X^\top X$ の条件が悪化し $\hat{\theta}$ の不確かさが増えやすい．
- R^2 は母集団の y の散らばり（SST）に依存．入力域が狭いと SST が小さく，雑音分散が同じなら R^2 は低下しやすい．

■考察（どうデータを取るべきか）

- 目的は $\mathrm{Var}(\hat{\theta})$ の縮小（$= X^\top X$ を「大きく」「良条件」に）．
- 入力設計：x を区間全体で広く配置（例：一様），中心化・標準化して $[-1, 1]$ に写像，または直交基底（Legendre/Chebyshev）で回帰．

- 最適化観点：A/D 最適設計を用いて $\mathrm{tr}((X^\top X)^{-1})$ や $\det\big((X^\top X)^{-1}\big)$ を最小化する点集合を選ぶ.

3

4　アルゴリズム比較

表や図を用いて複数手法の比較を行う。計算量・精度・安定性・実行時間の観点でまとめる。

5　結論

本章での結論と今後の課題を箇条書きでまとめる。

付録 A　使用コード一覧

主要スクリプトと入手先を列挙する。

- 実験コード（R）：`https://github.com/<your-repo>`
- データ生成スクリプト：`scripts/generate_data.jl`

付録 B　補足導出

必要な導出や補助的な数学的議論を載せる。

参考文献

参考文献

[1] 数理工学実験（2025 年度配布資料）.