

卒業論文A

マルコフパラメータ推定における 物理制約下での入力設計

片岡 慎治

大阪大学工学部 応用自然科学科
応用物理学科目

指導教員
和田 孝之 准教授



令和 7年 2月 6日

目次

第1章	序論	1
1.1	背景	1
1.2	目的	2
1.3	構成	3
第2章	入力に制約がない場合のマルコフパラメータ推定	4
2.1	問題設定	4
2.2	マルコフパラメータについて	4
2.3	マルコフパラメータの最小二乗推定の手順	5
2.4	先行研究における推定精度	6
第3章	入力に制約がある場合のマルコフパラメータ推定	9
3.1	入力の設計方法	9
3.2	入力を生成する確率分布の検討	11
3.3	推定誤差の上界	15
第4章	結論	21
	謝辞	23
	参考文献	24
付録A	定理1の証明に用いる補題	26
付録B	Sub-Gaussian性をもつ確率変数の性質	28

第1章

序論

1.1 背景

システム同定とは，入出力のデータからシステムのダイナミクスを推定することであり，制御システムの効率的かつ高精度な設計，システムの動作予測，故障診断などにおいて重要な役割を果たす[1, 2]．特に線形システムは，その扱いやすさから製造業におけるロボットの制御や自動車産業における安定性制御，経済学では市場予測や価格変動モデルなど幅広い分野で近似モデルとして採用されている．これまでシステム同定の研究では，収集された入出力データのサンプル数を無限に仮定した漸近的な解析が主流であり，入出力のデータ数が無限に増加する場合における性能保証が中心的なテーマだった．しかし，現実においては利用可能なデータが有限である場合が一般的であり，漸近的な結果に基づく設計では実用上の精度が保証されないことがある．また，故障確率や過渡現象などの概念を解析するのに適しているといった理由で，非漸近的な解析が注目されるようになり，有限のデータにおける特性を解析する研究が行われてきた．

近年の研究動向として，高次元統計[3, 4]や機械学習理論[5]の発展に伴い，これらのツールを活用することで非漸近的な解析における理論的基盤がさらに強化された．特に，線形システムにおける有限の入出力データを用いた解析に関する理論が整備され，システムパラメータを所定の精度レベルで同定するために必要なデータ数の下限を導出するための情報理論的議論もされている．これにより，システムの学習効率を定性的に評価できるようになった[6]．

その中でも、線形時不変システムの同定は、構造がシンプルでありながら広範な応用が可能であるため、従来から盛んに研究されてきた。特に入出力データから得られるシステム行列の最小二乗推定値について、誤差境界やサンプル複雑性が議論されてきた。線形システムにおいて最も基本的なケースである、入力を印加しない完全観測システムについて、システム行列の最小二乗推定値に関する有限時間での推定誤差の上界や任意の精度を保証するために必要なデータ数の下界を導出している[7, 8]。また、入力を印加するシステムについてシステム行列の誤差を制御設計に適用し最適二次レギュレータを用いて安定化制御を可能にする推定誤差とサンプル複雑性について議論されている[9]。

入出力データからシステム同定をする手法の一つとして、パルス伝達関数の係数であるマルコフパラメータを推定する手法がある。システムが安定な場合についてひとつの軌道の入出力データから、安定性を要求しない場合について複数の独立した軌道の入出力データから、それぞれマルコフパラメータ行列の最小二乗推定値の推定誤差の上界を導出している[10, 11]。

1.2 目的

システムの入出力関係を示すマルコフパラメータを推定することは、直接的にシステムの入出力の関係を表すことができる点から、重要である。従来の手法では、入力が正規分布に従って生成されるというケースが一般的であり、入力信号に制約を設けず、理想的に設計可能であることを前提としている。一方、現実の物理的なシステムでは、理想的な乱数に従う入力信号を生成することが困難な場合や、アクチュエータの物理的制約によって入力大きさに上限が課される場合、電池駆動のシステムにおける電流の制限など、実際の応用環境では様々な制約が課されることが多い。そのため入力に制約がある場合でも適用可能な理論的保証が求められる。このため本研究では、物理的な制約を考慮し、入力大きさに上限がある条件下において有限の入出力データから高精度な推定値を得る入力設計手法を確立し、理論的な保証を与えることを目的とする。

1.3 構成

本論文の構成は以下の通りである．まず，第2章では，マルコフパラメータおよび対象とするシステムの設定など，本論文で用いる数学的概念の定義を行い，マルコフパラメータの最小二乗推定の手順について述べる．また，先行研究の入力が正規分布に従い生成される場合のマルコフパラメータの推定精度について紹介する．第3章では入力の大きさに上限がある場合にも適用可能な，本論文で提案する入力設計アルゴリズムを紹介し，数値実験から最良の入力を得ることのできる確率分布について検討している．また，入力が決定した場合のマルコフパラメータ行列の推定誤差のスペクトルノルムについての上界を導出している．最後に第4章で結論を述べる．

第2章

入力に制約がない場合のマルコフパラメータ推定

2.1 問題設定

以下の離散時間線形時不変システム

$$x_{t+1} = Ax_t + Bu_t + w_t \quad (2.1)$$

$$y_t = Cx_t + Du_t + z_t \quad (2.2)$$

を考える．ただし， $x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^p, y_t \in \mathbb{R}^m$ はそれぞれ時刻 t におけるシステムの状態，入力，出力を表す．また， $w_t \in \mathbb{R}^n, z_t \in \mathbb{R}^m$ はプロセスノイズおよび観測ノイズを表し， $w_t \sim \mathcal{N}(0, \sigma_w^2 I), z_t \sim \mathcal{N}(0, \sigma_z^2 I)$ である．初期状態 $x_1 = 0$ であり， $\rho(A) \leq \eta < 1$ を仮定する．ここで， $\rho(\cdot)$ は行列の最大固有値の絶対値を表す．

本章では，未知の行列 A, B, C, D を推定することを目的とする．

2.2 マルコフパラメータについて

システム(2.1), (2.2)において，入出力関係は

$$y_t = Du_t + \sum_{i=1}^{t-1} CA^{i-1}Bu_{t-i} + \sum_{i=1}^{t-1} CA^{i-1}w_{t-i}$$

と表される．このときの入力の係数行列である D, CB, CAB, CA^2B, \dots をシステムのマルコフパラメータと呼ぶ．

また，システムの最初の T 個のマルコフパラメータを並べた行列を

$$G = \begin{bmatrix} D & CB & CAB & \dots & CA^{T-2}B \end{bmatrix} \in \mathbb{R}^{m \times pT} \quad (2.3)$$

と定義し，この G の推定を目的とする．次の節では，この G の最小二乗推定の手順について説明する．

2.3 マルコフパラメータの最小二乗推定の手順

まず，推定手順を説明するために，データ収集の方法について説明する．今回，システムの入出力データを時間長 \bar{N} まで取得する．このとき得られるデータセットは

$$\{(y_t, u_t) : t = 1, 2, \dots, \bar{N}\}$$

である．この与えられたデータセットから長さ T のセットを N 個生成する． $\bar{N} = N + T - 1$ である．ある時刻 t において入力，ノイズを時間長 T 分並べたものを

$$\bar{u}_t = \begin{bmatrix} u_t^* & u_{t-1}^* & \dots & u_{t-T+1}^* \end{bmatrix}^* \in \mathbb{R}^{pT} \quad (2.4)$$

$$\bar{w}_t = \begin{bmatrix} w_t^* & w_{t-1}^* & \dots & w_{t-T+1}^* \end{bmatrix}^* \in \mathbb{R}^{nT} \quad (2.5)$$

と表記する． y_t を $t - T + 1$ まで再帰的に展開すると，出力 y_t と \bar{u}_t ， G の関係は

$$\begin{aligned} y_t &= Cx_t + Du_t + z_t \\ &= C(Ax_{t-1} + Bu_{t-1} + w_{t-1}) + Du_t + z_t \\ &= G\bar{u}_t + F\bar{w}_t + z_t + e_t \end{aligned} \quad (2.6)$$

となる．ここで， $F = \begin{bmatrix} 0 & C & CA & \dots & CA^{T-2} \end{bmatrix} \in \mathbb{R}^{m \times nT}$ であり， $e_t = CA^{T-1}x_{t-T+1}$ は，時刻 $t - T + 1$ におけるシステムの状態の影響による誤差に相当する．この関係を用いて， N 個の $(\bar{u}_t, y_t)_{t=T}^{\bar{N}}$ を入出力データとして扱う．

上の(2.6)式右辺の2項目以降をノイズとして最小二乗法を適用すると G の最小二乗推定値は，

$$\hat{G} = \arg \min_{X \in \mathbb{R}^{m \times pT}} \sum_{t=T}^{\bar{N}} \|y_t - X\bar{u}_t\|^2 \quad (2.7)$$

と表せる．ただし， $\|\cdot\|$ はスペクトルノルムを表す．また，

$$Y = \begin{bmatrix} y_T & y_{T+1} & \dots & y_{\bar{N}} \end{bmatrix}^* \in \mathbb{R}^{N \times m} \quad (2.8)$$

$$U = \begin{bmatrix} \bar{u}_T & \bar{u}_{T+1} & \dots & \bar{u}_{\bar{N}} \end{bmatrix}^* \in \mathbb{R}^{N \times pT} \quad (2.9)$$

と定義すると，(2.7)式より，

$$\hat{G} = \min_{X \in \mathbb{R}^{m \times pT}} \|Y - UX^*\|_F^2 \quad (2.10)$$

となる．なお， $\|\cdot\|_F$ はフロベニウスノルムを表す．最小二乗解である \hat{G} は，

$$\hat{G} = (U^\dagger Y)^* \quad (2.11)$$

と与えられる．ここで， $U^\dagger = (U^*U)^{-1}U^*$ は左疑似逆行列である．また，

$$W = \begin{bmatrix} \bar{w}_T & \bar{w}_{T+1} & \dots & \bar{w}_{\bar{N}} \end{bmatrix}^* \in \mathbb{R}^{N \times pT} \quad (2.12)$$

$$E = \begin{bmatrix} e_T & e_{T+1} & \dots & e_{\bar{N}} \end{bmatrix}^* \in \mathbb{R}^{N \times m} \quad (2.13)$$

$$Z = \begin{bmatrix} z_T & z_{T+1} & \dots & z_{\bar{N}} \end{bmatrix}^* \in \mathbb{R}^{N \times m} \quad (2.14)$$

と定義すると，システムの入出力関係を以下のように記述することができる．

$$Y = UG^* + E + Z + WF^* \quad (2.15)$$

G について解くと $G^* = U^\dagger(Y - E - Z - WF^*)$ となるため，(2.11)式より最小二乗推定値と真値の誤差は，

$$(\hat{G} - G)^* = (U^*U)^{-1}U^*(E + Z + WF^*) \quad (2.16)$$

となる．

2.4 先行研究における推定精度

先行研究では入力を $\mathcal{N}(0, \sigma_u^2 I_p)$ に従い生成した場合について G の最小二乗推定を行っていた．この設定における最小二乗推定値 \hat{G} と G の推定誤差のスペクトルノルムの上界を与えた定理を紹介する．

定理 2.1. [10] $c, C > 0$ は十分大きな定数である． $\rho(A) < 1$ であり， $\frac{N}{\log^2(Np)} \geq cpT \log^2(pT)$ であるとする． $q = p + n$ ， $N_w = cTq \log^2 Tq \log^2 Nq$ とする．このとき，高い確率でマルコフパラメータ行列 G の最小二乗推定値 \hat{G} について，

$$\|\hat{G} - G\| \leq \frac{R_w \|F\| + R_z + R_e}{\sigma_u \sqrt{N}}$$

が成り立つ．ここで，

$$\begin{aligned} R_z &= 8\sigma_z \sqrt{Tp + m} \\ R_w &= \sigma_w (\sqrt{N_w} + N_w / \sqrt{N}) \\ R_e &= C\sigma_e \sqrt{\left(1 + \frac{mT}{N(1 - \rho(A)^{T/2})}\right) (Tp + m)} \end{aligned}$$

である．また，

$$\begin{aligned} \Phi(A) &= \sup_{\tau \geq 0} \frac{\|A^\tau\|}{\rho(A)^{\tau/2}} \\ \Gamma_\infty &= \sum_{i=0}^{\infty} \sigma_w^2 A^i (A^*)^i + \sigma_u A^i B B^* (A^*)^i \end{aligned}$$

とすると，

$$\sigma_e = \Phi(A) \|CA^{T-1}\| \sqrt{\frac{T \|\Gamma_\infty\|}{1 - \rho(A)^T}}$$

である．

この定理では，十分に大きな N において，サンプル数 N に対して $\mathcal{O}(N^{-1/2})$ で推定誤差の上界が減少することが示されている．さらに，最小二乗推定した \hat{G} を用いて出力の予測誤差を示している．

補題 2.1. [10] G の最小二乗推定値 \hat{G} を用いて，観測出力 y_t の推定値を

$$\hat{y}_t = \hat{G} \bar{u}_t$$

とする．このとき，

$$\mathbb{E}[\|y_t - \hat{y}_t\|^2] \leq \sigma_w^2 \|F\|_F^2 + \sigma_u^2 \|G - \hat{G}\|_F^2 + m\sigma_z^2 + \|CA^{T-1}\|^2 \text{tr}(\Gamma_\infty)$$

である．

また、先行研究[10]において提案されている、行列 G から元のシステムの行列 A, B, C, D を実現するアルゴリズムを紹介する。

Algorithm 1 Ho-Kalman アルゴリズムによる状態空間実現

```

1: procedure HO-KALMAN の最小実現
2:   Inputs: 長さ  $T$ , マルコフパラメータの最小二乗推定値  $\hat{G}$ , システムの次数
       $n$ , ハンケル行列のサイズ  $(T_1, T_2 + 1)$ , ただし  $T_1 + T_2 + 1 = T$ 
3:   Outputs: 状態空間実現  $\hat{A}, \hat{B}, \hat{C}$ 
4:   ハンケル行列  $\hat{H} \in \mathbb{R}^{mT_1 \times pT_2}$  を  $\hat{G}$  から生成
5:    $\hat{H}_1 \leftarrow \hat{H}$  の最初の  $pT_2$  列
6:    $\hat{L} \leftarrow \hat{H}_1$  のランク  $n$  の近似行列
7:    $U, \Sigma, V \leftarrow \text{SVD}(\hat{L})$ 
8:    $\hat{O} \leftarrow U_{:,1:n} \Sigma^{1/2}$ 
9:    $\hat{Q} \leftarrow \Sigma^{1/2} V_{:,1:n}^*$ 
10:   $\hat{C} \leftarrow \hat{O}$  の最初の  $m$  行
11:   $\hat{B} \leftarrow \hat{Q}$  の最初の  $p$  列
12:   $\hat{H}^+ \leftarrow \hat{H}$  の最後の  $pT_2$  列
13:   $\hat{A} \leftarrow \hat{O}^\dagger \hat{H}^+ \hat{Q}^\dagger$ 
14:  return  $\hat{A} \in \mathbb{R}^{n \times n}, \hat{B} \in \mathbb{R}^{n \times p}, \hat{C} \in \mathbb{R}^{m \times n}$ 
15: end procedure

```

ブロック行列 $X = [X_1, X_2, \dots, X_T] \in \mathbb{R}^{m \times Tp}$ と、整数 T_1, T_2 (ただし $T_1 + T_2 \leq T$) が与えられたとき、ここで用いるハンケル行列は、 $m \times p$ サイズのブロックを持つ $T_1 \times T_2$ のブロック行列として定義され、 (i, j) 番目のブロックは、

$$H[i, j] = X_{i+j}$$

である。

また、ここから得られる行列の組 $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ の推定精度は、アルゴリズムの入力である \hat{G} の推定精度に依存しているため G をより厳密に推定することが目標となる。 [10]

第3章

入力に制約がある場合のマルコフ パラメータ推定

3.1 入力の設計方法

離散時間線形時不変システム

$$x_{t+1} = Ax_t + Bu_t + w_t \quad (3.1)$$

$$y_t = Cx_t + Du_t + z_t \quad (3.2)$$

を考える. ただし, $x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^p, y_t \in \mathbb{R}^m$ はそれぞれ時刻 t におけるシステムの状態, 入力, 出力を表す. また, $w_t \in \mathbb{R}^n, z_t \in \mathbb{R}^m$ はプロセスノイズおよび観測ノイズを表し, $w_t \sim \mathcal{N}(0, \sigma_w^2 I), z_t \sim \mathcal{N}(0, \sigma_z^2 I)$ である. 初期状態 $x_1 = 0$ であり, $\rho(A) \leq \eta < 1$ と仮定する. ここで, $\rho(\cdot)$ は行列の最大固有値の絶対値を表す. また, 入力に関する制約条件

$$|u_{tk}| \leq a, \quad (\forall t = 1, 2, \dots, \quad k = 1, 2, \dots, p)$$

を加える. ここで, a は既知の正の実数である.

2.3節の手順に従い, G の最小二乗推定を行う. G の最小二乗推定値の推定誤差のスペクトルノルムの上界は(2.16)式より,

$$\begin{aligned} \|\hat{G} - G\| &= \|(U^*U)^{-1}U^*(W \cdot F + E + Z)\| \\ &\leq \|(U^*U)^{-1}U^*\|(\|W\| \cdot \|F\| + \|E\| + \|Z\|) \end{aligned} \quad (3.3)$$

となる．(3.3)式の右辺を最小化するための入力設計を考える．入力の項について，

$$\begin{aligned}\|(U^*U)^{-1}U^*\|^2 &= \lambda_{\max}\{((U^*U)^{-1}U^*)((U^*U)^{-1}U^*)^*\} \\ &= \lambda_{\max}((U^*U)^{-1}) \\ &= \frac{1}{\lambda_{\min}(U^*U)} \\ &= \frac{1}{\sigma_{\min}^2(U)}\end{aligned}$$

が成り立つ．ここで $\lambda_{\min}(\cdot)$, $\lambda_{\max}(\cdot)$ は，それぞれ行列の最小固有値，最大固有値を表し， $\sigma_{\min}(\cdot)$ は行列の最小特異値を表す．したがって推定誤差は，

$$\|\hat{G} - G\| \leq \left\| \frac{\|W\| \cdot \|F\| + \|E\| + \|Z\|}{\sigma_{\min}(U)} \right\| \quad (3.4)$$

である．

したがって $\sigma_{\min}(U)$ を大きくすると，推定誤差のノルムがより小さくすることができる．入力信号の大きさに上界を与えて以下の問題を考える．

問題1：

$$\begin{array}{lll} \text{maximize} & & \sigma_{\min}(U) \\ \text{subject to} & & |u_{tk}| \leq a \quad \begin{array}{l} t = 1, 2, \dots, \bar{N}, \\ k = 1, 2, \dots, p \end{array} \end{array}$$

この入力を決定する問題は凸問題ではないため，解くことは容易ではない．そこで，この問題の最適値に近い値をとる入力を設計するために，以下の補題を用いる．

補題 3.1. [12] 与えた $\delta \in (0, 1)$, $\epsilon \in (0, 1)$ について，

$$M \geq \frac{\log \frac{1}{\delta}}{\log \frac{1}{1-\epsilon}}$$

を満たすサンプル数 M を考える．入力のある確率分布に従って M 回生成し，それらを(2.9)式の U に従いならべ，

$$\{U^{(i)}, 1 \leq i \leq M\}$$

とする．このとき U の最小特異値

$$\{\sigma_{\min}(U^{(i)}), 1 \leq i \leq M\}$$

の最大値

$$\hat{\gamma}_M = \max_{i=1, \dots, M} \sigma_{\min}(U^{(i)})$$

は、 $1 - \delta$ 以上の確率で

$$\mathbb{P}\{\sigma_{\min}(U) \leq \hat{\gamma}_M\} \geq 1 - \epsilon \quad (3.5)$$

を満たす。

3.2 入力を生成する確率分布の検討

補題3.1に基づき、どの確率分布を選択してアルゴリズムを適用すれば問題1の最適値により近い入力を得られるかを数値実験により検証する。

$m = 1, p = 1, n = 2$ として、システムの行列を

$$A = \begin{bmatrix} 0.6 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, C = [1 \quad 1], D = 1$$

と設定する。行列 A の固有値は、

$$\lambda_1 = \frac{-\sqrt{3} + 5}{10}, \quad \lambda_2 = \frac{\sqrt{3} + 5}{10}$$

であり、 $\lambda_1, \lambda_2 < 1$ より、安定なシステムである。プロセスノイズ、観測ノイズについて、 $\sigma_w = \sigma_z = 0.5$ とする。推定するマルコフパラメータの長さについては $T = 20$ として、学習回数を $N = 1000$ までとり、50刻みで推定誤差についてプロットした。各分布に対して、補題3.1を用いて設計した入力について G の最小二乗推定値を計算し、真値との推定誤差を10回出力し、その平均についてプロットしている。

図3.1は、補題3.1を用いる際に、正規分布に従う入力を選択した場合における分散の違いが推定精度に与える影響を比較したものである。

この図から、推定精度は分散の大きさ依存していることが分かる。入力の制約は、

$$|u_t| \leq 1, \quad t = 1, 2, \dots, \bar{N}$$

とした。そのため正規分布 $\mathcal{N}(0, \sigma_u^2)$ に従う確率変数 X について、 $X > 1$ のとき $X = 1$ 、 $X < -1$ のとき $X = -1$ としている。この分布を丸めた正規分布と呼ぶ。

図3.2では、 $\mathcal{N}(0, 1)$ を丸めた正規分布(clipped normal distribution)と切断正規分布(truncated normal distribution)を比較している。切断正規分布とは、通常の正規分

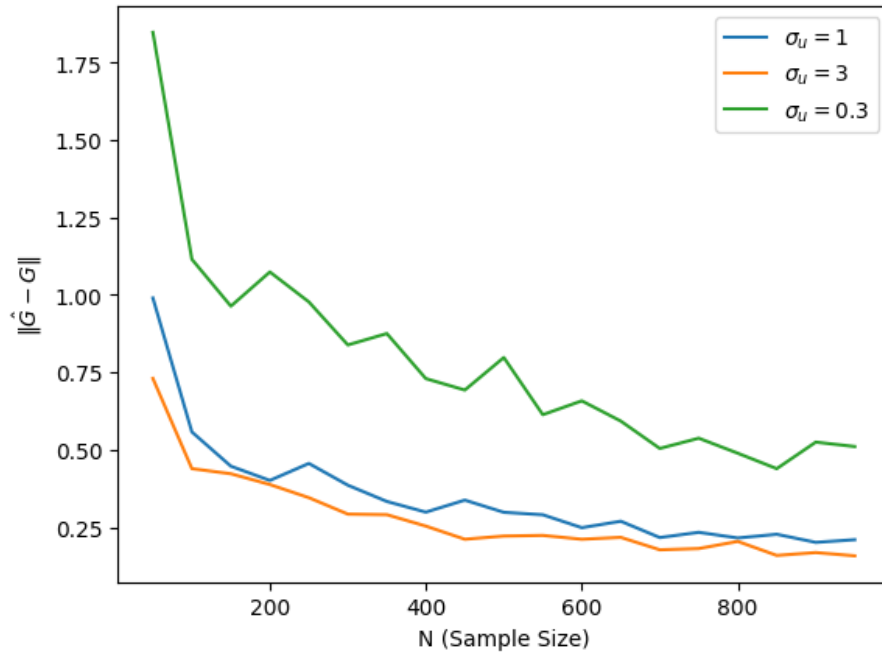


図3.1 正規分布の分散を変えたときの比較

布を $-a \leq x \leq b$ の範囲に制限した分布である。確率密度関数は

$$f(x|\mu, \sigma, a, b) = \begin{cases} 0, & \text{if } x \leq a, \\ \frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}, & \text{if } a < x < b, \\ 0, & \text{if } b \leq x. \end{cases}$$

と表される。ここで $\phi(z)$, $\Phi(z)$ はそれぞれ標準正規分布 $\mathcal{N}(0, 1)$ の確率密度関数、累積分布関数を表しており、

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \Phi(z) = \int_{-\infty}^z \phi(t) dt$$

である。図3.2において、切断正規分布は $\mathcal{N}(0, 1)$ を $[-1, 1]$ で制限している。丸めた正規分布と切断正規分布を比較すると、丸めた正規分布の方が推定精度が高いことがわかる。

図3.3では、 $\mathcal{N}(0, 1)$ を丸めた正規分布と一様分布 $\mathcal{U}[-1, 1]$, $\{-1, +1\}$ を等確率でとるベルヌーイ試行の3つの分布について比較している。

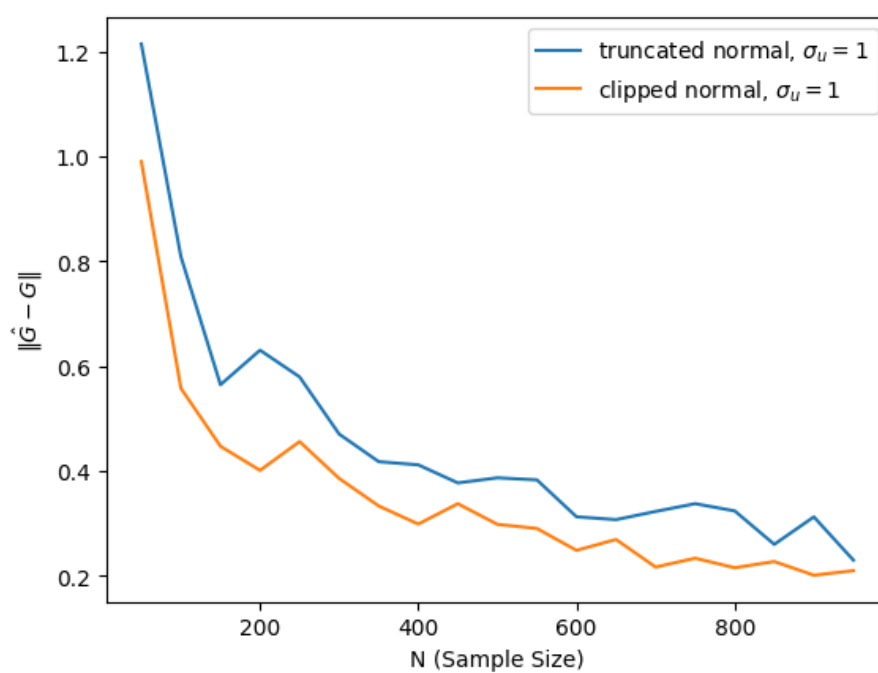


図3.2 切断正規分布と丸めた正規分布の比較

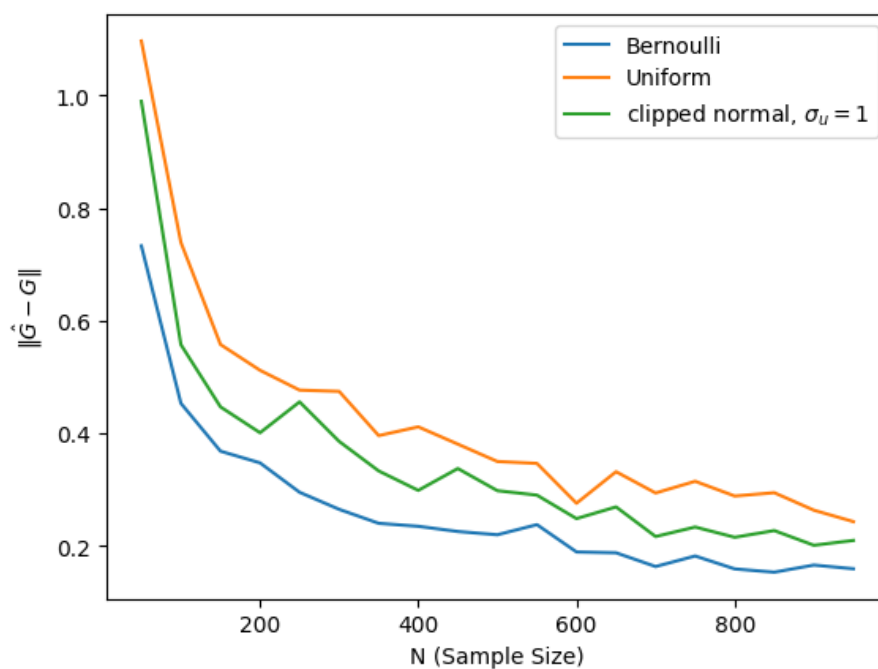


図3.3 一様分布, ベルヌーイ分布, 正規分布の比較

次の図3.4では,

$$y = \frac{2k+1}{2}x^{2k}, \quad (-1 \leq x \leq 1)$$

の確率分布と, ベルヌーイ試行について推定誤差を比較している. この分布は $k \rightarrow \infty$ で $\{-1, +1\}$ を等確率でとるベルヌーイ試行に収束する. $k = 1, 2$ の場合である $f(x) = \frac{3}{2}x^2, f(x) = \frac{5}{2}x^4$ と $\{-1, +1\}$ を等確率でとるベルヌーイ試行を比較した.

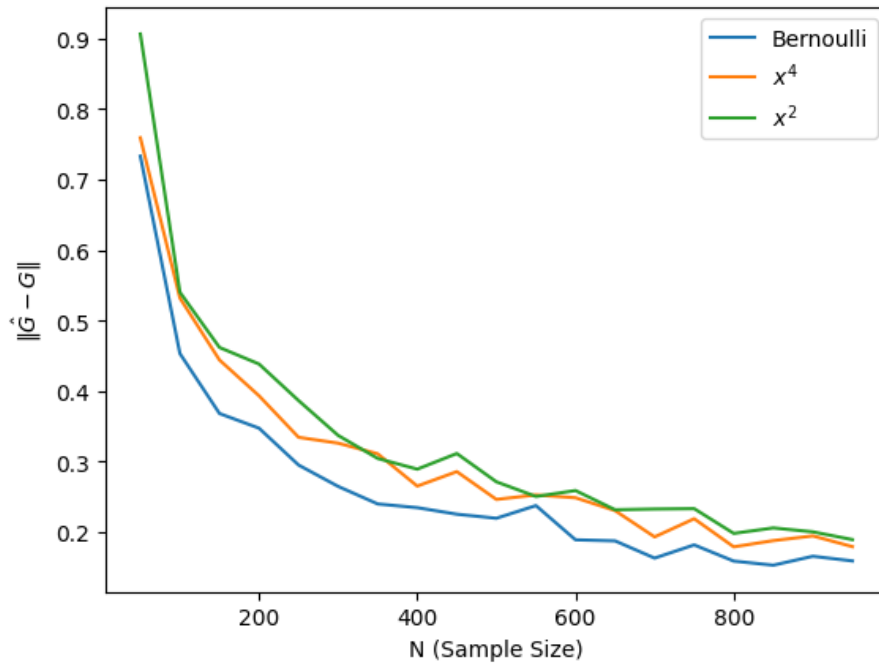


図3.4 $f(x) = \frac{3}{2}x^2, f(x) = \frac{5}{2}x^4$, ベルヌーイ分布の比較

以上の比較から, $\{-1, +1\}$ に近い値を高確率でとる分布が推定誤差を小さくしていることがわかる.

実際に T, N を固定した低次元の場合について, 問題1を解いた場合の結果を以下に示す. 入力の次元は, $p = 1$ として, 各時刻の入力に対し, $|u_t| \leq 1, (t = 1, 2, \dots, \bar{N})$ とした. $T = 4, \bar{N} = 7$ の場合について, 問題1の最適解の1つは,

$$U = \begin{bmatrix} 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

であり, 最適値は $\sigma_{\min}(U) = 2$ である.

また, $T = 4, \bar{N} = 8$ の場合について, 問題1の最適解の1つは,

$$U = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

であり, 最適値は $\sigma_{\min}(U) = 2$ である. この結果からも $\{-1, +1\}$ を等確率でとる分布を選択するのが良いことが分かる.

実際に, 補題3.1のアルゴリズムに対して, $\{-1, +1\}$ を等確率でとる確率変数を用いて入力を設計し, その入力に基づいて得られた最小二乗推定値と, 先行研究の入力が正規分布に従い生成される場合の最小二乗推定値を比較した. 用いたサンプル数は, $N = 100$ で, $\epsilon = \delta = 0.0001$ としたときに補題3.1を満たす M として, $M = 100000$ とした. 表 3.1 から, 入力に制約のない場合よりも推定精度が高いことがわかる.

表3.1 入力をマルコフパラメータの最小二乗推定値

	D	CB	CAB	CA^2B
真値	1	2	1.3	0.86
$\mathcal{N}(0, 1)$	1.0923	2.0340	1.4414	0.9217
提案手法	1.0443	2.0321	1.3640	0.8664

3.3 推定誤差の上界

入力を設計した場合の行列 G の最小二乗推定値と真値の誤差のスペクトルノルムの上界について, 以下の定理を示す.

定理 3.1. $\bar{N} = N + T - 1$ 時間の入出力データが与えられたとき, 任意の $\delta \in (0, 1)$ について G の最小二乗推定値 \hat{G} について, $1 - 3\delta$ 以上の確率で

$$\|\hat{G} - G\| \leq \frac{R_w \|F\| + R_e + R_z}{\sigma_{\min}(U)} \quad (3.6)$$

が成立する。ただし,

$$\begin{aligned}
 R_w &= \sigma_w \sqrt{2 \max\{N, pT\} \log \left(\frac{N + pT}{\delta} \right)} \\
 R_e &= \|C\| \eta^{T-1} \sum_{t=T}^{\bar{N}} \sum_{i=1}^{t-T-i} \eta^{t-T-i} \|Bu_i\| \\
 &\quad + N \frac{\sigma_w \eta^{T-1}}{\sqrt{1-\eta^2}} \|C\| \sqrt{m + \sqrt{8m \log(2/\delta)}} \\
 R_z &= \sigma_z (\sqrt{N} + \sqrt{m} + \sqrt{\log 1/\delta})
 \end{aligned}$$

である.

証明. まず, ノイズの項 $\|Z\|$ の上界を与える. 付録Aの補題A.1より, $1 - \delta$ 以上の確率で

$$\|Z\| \leq \sigma_z (\sqrt{N} + \sqrt{m} + \sqrt{\log 1/\delta})$$

が成り立つ.

次に, $\|W\|$ の上界を与える. まず, w_t がスカラーである場合 ($p = 1$) を考える.

$$W^* = \begin{bmatrix} W_T & W_{T+1} & \cdots & W_{\bar{N}} \\ W_{T-1} & W_T & \cdots & W_{\bar{N}-1} \\ \vdots & \vdots & \ddots & \cdots \\ W_1 & W_2 & \cdots & W_N \end{bmatrix} \in \mathbb{R}^{T \times N}$$

であるから,

$$B_1 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 0 & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}, \cdots, B_{\bar{N}} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

と定義すると,

$$W^* = w_1 B_1 + w_2 B_2 + \cdots + w_{\bar{N}} B_{\bar{N}}$$

と表すことができる．ここで，補題A.2から， W の分散に関する統計量 $v(W)$ を，

$$\begin{aligned} v(W) &= \sigma_w^2 \max \left\{ \sum_k B_k B_k^*, \sum_k B_k^* B_k \right\} \\ &= \sigma_w^2 \max\{NI, TI\} \\ &= \sigma_w^2 \max\{N, T\} \end{aligned}$$

とする． $w_t \sim \mathcal{N}(0, \sigma_w^2 I)$ であるから， w_t がベクトル($p > 1$)の場合も同様に考えることができ，このとき $v(W) = \sigma_w^2 \max\{N, pT\}$ であるから，

$$\mathbb{P}\{\|W\| \geq t\} \leq \exp\left(\frac{-t^2}{2\sigma_w^2 \max\{N, pT\}}\right)$$

$\exp\left(\frac{-t^2}{2\sigma_w^2 \max\{N, pT\}}\right) = \delta$ とおき， δ について解くと，

$$\mathbb{P}\left\{\|W\| \leq \sigma_w \sqrt{2 \max\{N, pT\} \log\left(\frac{N + pT}{\delta}\right)}\right\} \geq 1 - \delta$$

が成り立つ．

次に， $\|E\|$ の項の上界について考える．

$$\begin{aligned} e_t &= CA^{T-1} x_{t-T+1} \\ &= CA^{T-1} \begin{bmatrix} B & AB & \dots & A^{t-T-1}B \end{bmatrix} \begin{bmatrix} u_{t-T} \\ \vdots \\ u_1 \end{bmatrix} \\ &\quad + CA^{T-1} \begin{bmatrix} I & A & \dots & A^{t-T-1} \end{bmatrix} \begin{bmatrix} w_{t-T} \\ \vdots \\ w_1 \end{bmatrix} \end{aligned}$$

と展開できる．

$$W_t = CA^{T-1} \sum_{i=0}^{t-T-1} A^i w_{t-T-i}$$

とすると， e_t の共分散行列は $\mathbb{E}[W_t W_t^*]$ で表されるので， $w_t \sim \mathcal{N}(0, \sigma_w^2 I)$ であること

から,

$$\begin{aligned} X_t &= CA^{T-1} \sum_{i=1}^{t-T} A^{t-T-i} B u_i \\ \Sigma(e_t) &= \mathbb{E}[W_t W_t^*] \\ &= CA^{T-1} \left\{ \sum_{i=0}^{t-T-1} \sigma_w^2 A^i (A^*)^i \right\} (A^{T-1})^* C^* \end{aligned}$$

とすると, e_t は, $\mathcal{N}(X_t, \Sigma(e_t))$ に従う. また $\rho(A) \leq \eta$ であるから,

$$\begin{aligned} \Sigma(e_t) &\leq CA^{T-1} \left\{ \sum_{i=0}^{\infty} \sigma_w^2 A^i (A^*)^i \right\} (A^{T-1})^* C^* \\ &\leq \frac{\sigma_w^2 \eta^{2(T-1)}}{1 - \eta^2} CC^* \\ &= \Sigma_e \end{aligned}$$

とすると上界を導出するにあたって, $e_t \sim \mathcal{N}(X_t, \Sigma_e)$ と見積もってもよい. したがって, $\tilde{e}_t = (\Sigma_e)^{-1/2}(e_t - X_t)$ は $\mathcal{N}(0, I_n)$ に従う. これを e_t について解くと, $e_t = (\Sigma_e)^{1/2}(\tilde{e}_t) + X_t$ となるので,

$$\begin{aligned} \|e_t\| &= \|(\Sigma_e)^{1/2} \tilde{e}_t + X_t\| \\ &\leq \|\Sigma_e\|^{1/2} \|\tilde{e}_t\| + \|X_t\| \end{aligned}$$

が成り立つ. したがって, (2.13)式より,

$$\begin{aligned} \|E\| &\leq \|e_T\| + \|e_{T+1}\| + \cdots + \|e_{\bar{N}}\| \\ &\leq N \|\Sigma_e\|^{1/2} \|\tilde{e}_t\| + \sum_{t=T}^{\bar{N}} \|X_t\| \end{aligned}$$

が成り立つ. ここで,

$$\begin{aligned} \sum_{t=T}^{\bar{N}} \|X_t\| &= \sum_{t=T}^{\bar{N}} \|CA^{T-1} \sum_{i=1}^{t-T} A^{t-T-i} B u_i\| \\ &\leq \|C\| \eta^{T-1} \sum_{t=T}^{\bar{N}} \sum_{i=1}^{t-T} \eta^{t-T-i} \|B u_i\| \end{aligned}$$

である.

また, $\|\tilde{e}_t\|^2$ は自由度 m の χ 二乗分布に従うので, 付録Aの補題A.3より,

$$\mathbb{P} \left[\left| \frac{1}{m} \|\tilde{e}_t\|^2 - 1 \right| \geq t \right] \leq 2e^{-nt^2/8}$$

が成り立つ. 右辺を δ として整理すると, $1 - \delta$ 以上の確率で

$$\|\tilde{e}_t\| \leq \sqrt{m + \sqrt{8m \log(2/\delta)}}$$

が成り立つことが分かる.

以上を組み合わせると G の推定誤差の上界(3.6)式が得られる. \square

また, 今回の問題設定では入力の大きさに上界 $|u_{tk}| \leq a$, ($\forall t = 1, 2, \dots, k = 1, 2, \dots, p$)を課している. 付録Bから, 上界, 下界のある確率分布から入力を設計しているため, 設計した入力はsub-Gaussian性を持つことが分かる. 2章で紹介した定理2.1は, 入力がsub-Gaussian性を持つ確率分布から生成されていれば用いることができる.

正規分布 $\mathcal{N}(\mu, \sigma^2)$ には従う変数 X は, パラメータ σ のsub-Gaussianである. また, 補題3.1を用いて入力を設計する際に選択した $\{-a, +a\}$ を等確率でとる確率分布から生成される確率変数も, 例B.1から, パラメータ a のsub-Gaussianであることが分かる. したがって, 定理2.1や, 補題2.1において入力の分散である σ_u をsub-Gaussianのパラメータに置き換えて考えることができる. 十分大きな N における推定誤差の上界について, 以下の定理を示す.

定理 3.2. $c, C > 0$ は十分大きな定数である. $\rho(A) < 1$ であり, $\frac{N}{\log^2(Np)} \geq cpT \log^2(pT)$ であるとする. $q = p + n$, $N_w = cTq \log^2 Tq \log^2 Nq$ とする. また, 入力がパラメータ σ のsub-Gaussianである確率変数に従い生成されれるとする. このとき, 高い確率でマルコフパラメータ行列 G の最小二乗推定値 \hat{G} について,

$$\|\hat{G} - G\| \leq \frac{R_w \|F\| + R_z + R_e}{\sigma \sqrt{N}}$$

が成り立つ．ここで，

$$\begin{aligned} R_z &= 8\sigma_z\sqrt{Tp+m} \\ R_w &= \sigma_w(\sqrt{N_w} + N_w/\sqrt{N}) \\ R_e &= C\sigma_e\sqrt{\left(1 + \frac{mT}{N(1-\rho(A)^{T/2})}\right)(Tp+m)} \end{aligned}$$

である．また，

$$\begin{aligned} \Phi(A) &= \sup_{\tau \geq 0} \frac{\|A^\tau\|}{\rho(A)^{\tau/2}} \\ \Gamma_\infty &= \sum_{i=0}^{\infty} \sigma_w^2 A^i (A^*)^i + \sigma^2 A^i B B^* (A^*)^i \end{aligned}$$

とすると，

$$\sigma_e = \Phi(A) \|CA^{T-1}\| \sqrt{\frac{T\|\Gamma_\infty\|}{1-\rho(A)^T}}$$

である．

第4章

結論

線形時不変システム同定の有限時間での解析においては、様々な観点から研究が行われている。本研究では、特に部分観測の線形時不変システムにおけるマルコフパラメータの推定に焦点を当てた。従来の研究では、入力信号が正規分布に従い生成されるという理想的な前提のもとで設計されていた。しかし、実際の応用環境では、アクチュエータの出力制約などにより、入力信号に様々な制約が課されることが一般的である。

そこで本研究では、より現実的な問題設定として、アクチュエータの物理的制約により入力信号の大きさに上限がある状況下で、システムの入出力関係を示すマルコフパラメータを推定する手法を提案し、その有効性を数値実験により検証した。具体的には、補題3.1のアルゴリズムを用いる際に、異なる確率分布を用いて入力を設計し、推定誤差を比較した。その結果、一様分布や正規分布と比較して、分散の大きい丸めた正規分布や $\{-1, +1\}$ を等確率でとるベルヌーイ試行など、上限値に高確率で近づく確率分布を用いることで、推定精度が向上することが明らかになった。また、推定するマルコフパラメータの数や用いる入出力データセットの数が少ない場合においても、問題1を解いた結果から、ベルヌーイ試行を用いた入力設計が有効であることが確認された。さらに、設計した入力を用いてマルコフパラメータの最小二乗推定を行った際の推定誤差のノルムについて、その理論的な誤差上界を導出した。これにより、制約のある環境下でも適切な入力設計を行うことで推定精度が保証されることを数理的に示した。

今後の課題として、同じ問題設定のもとでより厳密な推定誤差の上界を導出するこ

とが挙げられる．また，本研究ではアクチュエータの物理的制約に基づいた入力設計を行ったが，今後は電力消費が制約されるデバイスにおけるエネルギー制約を考慮した入力設計へと発展させることが求められる．さらに，実際の計測環境ではノイズが理想的な正規分布に従わない場合があるため，より現実的なノイズモデルを考慮した解析も重要な課題である．

謝辞

本研究は，著者が大阪大学工学部応用自然科学科応用物理学科目在学中に大阪大学大学院情報科学研究科情報数理学専攻計画数理学講座において行われたものである．未筆ながら，本研究を進めるにあたり，御多忙の中，細部にいたるまで御指導と御助言を賜りました本学准教授 和田孝之先生には深く感謝の意を表すと共に，厚く御礼を申し上げます．また，研究を行う上で，終始適切な御助言と御助力を賜りました本学教授 藤崎泰正先生，事務補佐員 桑原奈美氏には厚く御礼申し上げます．最後に，貴重な御意見を賜りました計画数理学講座内諸氏に御礼申し上げます．

参考文献

- [1] 足立: システム同定の基礎, 東京電機大学出版局 (2009)
- [2] L. Ljung: *System Identification Theory for the User*, Prentice-Hall, inc. (1999)
- [3] R. Vershynin: *High-Dimensional Probability An Introduction with Applications in Data Science*, Cambridge University Press (2018)
- [4] M. J. Wainwright: *High-Dimensional Statistics: A Non-asymptotic Viewpoint*, Cambridge University Press (2019)
- [5] T. Hastie, R. Tibshirani, and J. Friedman: *The Elements of Statistical Learning*, Springer New York (2009)
- [6] T. Anastasios, Z. Ingvar, M. Nikolai, and G. J. Pappas: Statistical Learning Theory for Control: A Finite-Sample Perspective, *IEEE Control Systems Magazine* **43**-6, 67/97 (2023)
- [7] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis: Finite time identification in unstable linear systems, *Automatica*, **96**, 342/353 (2018)
- [8] Y. Jedra and A. Proutiere: Finite-Time Identification of Linear Systems: Fundamental Limits and Optimal Algorithms, *IEEE Transactions on Automatic Control*, **68**-5, 2805/2820 (2023)
- [9] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu: On the Sample Complexity of the Linear Quadratic Regulator, *Foundations of Computational Mathematics*, **20**, 633/679 (2020)
- [10] S. Oymak and N. Ozay: Revising Ho-Kalman-based System Identification: Robustness and Finite-Sample Analysis, *IEEE Transactions on Automatic Control*, **67**-4, 1914/1928 (2022)
- [11] Y. Zheng and N. Li: Non-Asymptotic Identification of Linear Dynami-

- cal Systems Using Multiple Trajectories *IEEE Control Systems Letters*, **5-5**, 1683/1698 (2021)
- [12] R. Tempo, G. Calafiore, and F. Dabbene: *Randomized Algorithms for Analysis and Control of Uncertain Systems*, Springer London (2005)
- [13] R. Vershynin: *Introduction to the non-asymptotic analysis of random matrices*, Cambridge University Press (2012)
- [14] J. A. Tropp: *An Introduction to Matrix Concentration Inequalities*, Foundations and Trends in Machine Learning (2015)
- [15] D. J. H. Garling: *Inequalities: A journey into Linear Analysis*, Cambridge University Press (2010)

付録A

定理1の証明に用いる補題

補題 A.1. [13] Z を各要素が $\mathcal{N}(0, \sigma_z^2)$ に従う $N \times m$ の行列とする. このとき, 任意の $\delta \in (0, 1)$ について $1 - \delta$ 以上の確率で

$$\|Z\| \leq \sigma_z(\sqrt{N} + \sqrt{m} + \sqrt{\log 1/\delta})$$

が成り立つ.

補題 A.2. [14] 各行列の次元が $d_1 \times d_2$ である複素行列の n 個の有限列 $\{B_k\}, k = 1, 2, \dots, n$ を考える. また $\{\gamma_k\}$ を標準正規分布に従う有限列であるとする. このとき, $Z = \sum_k \gamma_k B_k$ とする. Z の分散に関する統計量を

$$\begin{aligned} v(Z) &= \max\{\|\mathbb{E}(ZZ^*)\|, \|\mathbb{E}(ZZ^*)\|\} \\ &= \max\left\{\left\|\sum_{k=1}^n B_k B_k^*\right\|, \left\|\sum_{k=1}^n B_k B_k^*\right\|\right\} \end{aligned}$$

とする. このとき, 任意の $t \geq 0$ について,

$$\mathbb{P}\{\|Z\| \geq t\} \leq \left(d_1 + d_2 \exp\left(\frac{-t^2}{2v(z)}\right)\right)$$

が成り立つ.

補題 A.3. [4] $z_k \sim \mathcal{N}(0, 1)$ に従う独立な変数であるとき, $Y = \sum_{k=1}^n z_k^2$ は自由度 n の χ^2 乗分布に従う. このとき,

$$\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \geq t\right] \leq 2e^{-nt^2/8}$$

が成り立つ.

付録B

Sub-Gaussian性をもつ確率変数の性質

ここでは本論文で用いるSub-Gaussianの基本的性質をまとめる.

定義 B.1. 平均が $\mathbb{E}[X] = \mu$ である確率変数 X は, ある正の数 σ が存在し, 任意の $\lambda \in \mathbb{R}$ について,

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\sigma^2 \lambda^2}{2}}$$

が成り立つとき, X はパラメータ σ のsub - Gaussianであるという.

例 B.1. $\{-a, +a\}$ を等確率でとる確率変数を ϵ とする. このとき,

$$\begin{aligned} \mathbb{E}[e^{\lambda \epsilon}] &= \frac{1}{2} \{e^{-a\lambda} + e^{a\lambda}\} \\ &= \frac{1}{2} \left\{ \sum_{k=0}^{\infty} \frac{(-a\lambda)^k}{k!} + \sum_{k=0}^{\infty} \frac{(a\lambda)^k}{k!} \right\} \\ &= \sum_{k=0}^{\infty} \frac{(a\lambda)^{2k}}{(2k)!} \\ &\leq e^{a^2 \lambda^2 / 2} \end{aligned} \tag{B.1}$$

であるから, 確率変数 ϵ はパラメータ a のsub - Gaussianである.

例 B.2. 一般に, 有界区間 $[a, b]$ にのみ値をとる $\mathbb{E}[X] = 0$ である確率変数 X について, X' を X と同じ分布に従う独立な別の確率変数であるとする. $\mathbb{E}[X'] = 0$, 凸関数

$f(x) = e^x$ に対して, Jensenの不等式($f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$)/15)を適用すると,

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}_X[e^{\lambda(X - \mathbb{E}_{X'}[X'])}] \leq \mathbb{E}_{X, X'}[e^{\lambda(X - X')}]$$

が成り立つ. さらに, $\{-1, +1\}$ を等確率でとる確率変数を ϵ として導入すると,

$$\begin{aligned} \mathbb{E}_\epsilon[e^{\lambda \epsilon(X - X')}] &= \frac{1}{2}e^{\lambda(X - X')} + \frac{1}{2}e^{-\lambda(X - X')} \\ &= e^{\lambda^2(X - X')^2/2} \end{aligned}$$

が成り立つ. X, X' は, 同じ分布に従う独立な変数であるため, $X - X', X' - X$ を新たな確率変数と置くと同じ分布に従うからである. また, (B.1)式において, $\lambda \rightarrow \lambda(X - X')$ とすると

$$\mathbb{E}_\epsilon[e^{\lambda \epsilon(X - X')}] \leq \mathbb{E}_{X, X'}[e^{\lambda^2(X - X')^2/2}]$$

が成り立つ. $X, X' \in [a, b]$ であるから, $|X - X'| \leq |b - a|$ であるので, .

$$\begin{aligned} \mathbb{E}[e^{\lambda X}] &\leq \mathbb{E}_{X, X'}[e^{\lambda^2(X - X')^2/2}] \\ &\leq e^{\lambda^2(b-a)^2/2} \end{aligned}$$

が成り立つ. したがって, X は少なくともパラメータ $b - a$ のsub - Gaussianである.

sub-Gaussian性は, 正規分布と同様の集中不等式を適用できるため期待値周辺の集中性やサンプル数と精度の関係を評価する際に広く利用される.