

# データ駆動型回帰分析 補足資料

前川 大空 \*

2025 年 6 月 29 日

## 目次

1	回帰分析の課題	1
1.1	回帰分析 . . . . .	2
1.2	線形回帰モデル . . . . .	5
1.3	本書の課題と構成 . . . . .	5
1.4	補論 . . . . .	5
2	変数選択	8
2.1	設定 . . . . .	8
2.2	推定された予測誤差に基づく変数選択 . . . . .	8
2.3	情報量規準 . . . . .	8
2.4	変数選択の一致性と漸近最適性 . . . . .	8
2.5	その他のモデル評価基準 . . . . .	8
2.6	変数選択後の統計的推測の問題 . . . . .	8
付録 A	記法	9
付録 B	数学的準備 (書評: 『計量経済学のための数学』)	9
付録 C	測度論的確率論	11

## 1 回帰分析の課題

■データ駆動 従来の計量経済学では変数選択, ノンパラ, セミパラがこれにあたる. **何を指すのだろう . . .**

■一様妥当性 データ生成の母集団分布についての頑健性, といったことか. **頑健性と同じなのかしら . . .**

\* 一橋大学経済学部 4 年, 五年一貫専修コース公共経済プログラム

## 1.1 回帰分析

■p.1 観測可能性 観測可能性を前提としている。つまり、 $\mathbf{W}$  はコントロール変数。

■p.1 構造モデル 構造モデルはデータ生成過程を表すのみ、観測不可能な部分を誤差に全てまとめているため、内生性などは排除されていない。つまり、構造モデルは平均独立や条件付平均独立を満たすとは限らない。

■p.2 回帰モデル 回帰関数とは、応答変数の条件付期待値関数のことを指す。回帰モデルとは：

$$Y = g(S, \mathbf{W}) + e := \underbrace{\mathbb{E}[Y | S, \mathbf{W}]}_{\text{回帰関数}} + e$$

と回帰関数<sup>\*1</sup> を (1.1) 式に適応させた式である。LIE から条件付平均独立  $\mathbb{E}[e | S, \mathbf{W}] = 0$  が確認できる。

■p.2 回帰関数の識別 回帰関数  $\mathbb{E}[Y | S, \mathbf{W}]$  は  $(Y, S, \mathbf{W})$  の同時分布から一意に定まる。これは、一般の分布について、母集団モーメントは分布が判明することによって一意に定まるためである。条件付分布  $Y |_{S, \mathbf{W}}$  の平均はこの特殊ケースと見なせる。ここで、 $(Y, S, \mathbf{W})$  は全て観測可能である。

## Def: 識別

観測されるデータの同時分布が既知の時、 $\theta$  の値が一意に定まるならば、 $\theta$  は識別されるという。

上の定義から分かるように、回帰関数は識別される。何故ならば、観測可能なデータ  $(Y, S, \mathbf{W})$  の分布が既知の時、上記の議論から回帰関数  $g(S, \mathbf{W}) := \mathbb{E}[Y | S, \mathbf{W}]$  は一意に定まるためである。

■p.2 構造的/記述的な分析 以下のような区別が為されている。

## 構造的/記述的な分析

構造的な分析: 何かしらの決定メカニズムを背後に想定する分析

記述的な分析: 観測される情報のみから識別可能な変数間の関係の分析

つまり、回帰モデルによる分析は記述的な分析といえる。

■p.2 構造モデルの識別 回帰モデルでない構造モデル、つまり、 $\mathbb{E}[e | S, \mathbf{W}] \neq 0$  である場合、 $g$  は回帰関数ではない別の関数になる。関数が特定できないため、このままでは識別できない。構造モデルへの追加的仮定は、識別のために、経済学ならば経済理論に基づいた妥当性が実証データからは検証できない仮定を置く。<sup>\*2</sup>

■p.3 構造モデルにおける誤差項の加法分離性 (1.1) 式において、大卒の因果効果を測るために  $\mathbf{W}$  のみならず本来観測不可能な  $e$  も一定としていることに注意せよ。つまり、因果効果の識別については、(1.1) 式に基づく構造的な分析の文脈においても述べられていない。

■p.3  $\mathbf{W}$  一定では高卒/大卒の賃金の差も一定という制約 検証しておこう。

<sup>\*1</sup> 正確には、回帰関数の内、OLS の下で  $Y$  の最良予測であるもの。他の損失関数を選択すれば他の回帰関数が最良になりうる。例えば、最小絶対誤差法における最良の回帰関数は条件付中央値である。詳しくは『計量経済学のための数学』の素敵な 9 章を参照。

<sup>\*2</sup> **これが構造推定なる分野なのだろうか。**

*Proof* 誤差項が加法分離可能な構造モデル:

$$Y = g(S, \mathbf{W}) + e,$$

について,  $(\mathbf{w}_i, e_i)$  たる個人  $i$  の教育年数の賃金への因果効果は,

$$Y_i |_{S_i=16} - Y_i |_{S_i=12} = [g(16, \mathbf{w}_i) + e_i] - [g(12, \mathbf{w}_i) + e_i] = g(16, \mathbf{w}_i) - g(12, \mathbf{w}_i).$$

$\mathbf{w}_i = \mathbf{w}_j$  なる 2 個人の因果効果は  $g(16, \mathbf{w}_i) - g(12, \mathbf{w}_i) = g(16, \mathbf{w}_j) - g(12, \mathbf{w}_j)$  で同一.  $\square$

■p.3 回帰モデルと因果効果 分かるのはあくまで平均で, 予測に過ぎない. 因果効果とは限らない.

■p.3 因果推論は構造的な分析 先述の例の通り, 因果効果は識別できるとは限らず, 単にメカニズムを記述したのみの, 即ち構造的な分析の範疇であった. しかし理論に基づく様々な仮定を置くことによって識別が可能となり, 因果効果の分析, 因果推論も記述的な分析に落とし込むことが出来る. 特に  $e$  が加法分離可能なケースは直ちに記述的な分析に落とし込める.

*Proof* (1.1) 式の  $e$  が加法分離可能な構造モデルを考える:

$$Y = h(D, \mathbf{X}) + e$$

ここで, 観測可能な説明変数は  $(D, \mathbf{X})$  であり, データが既知ならば LIE より以下のような形となる:

$$\mathbb{E}[Y | D, \mathbf{X}] = h(D, \mathbf{X}) + \mathbb{E}[e | D, \mathbf{X}]$$

ここで, 回帰モデルの必要条件である平均独立の仮定  $\mathbb{E}[e | D, \mathbf{X}] = 0$  を置くと,  $h(D, \mathbf{X})$  は回帰関数に等しくなり, 記述的な分析, さらに言えば回帰分析に落とし込むことが出来た.  $\square$

■p.4 不均一分散の定義 末石計量では, 無条件分散は説明変数が確率の場合必ず定数になるため, 条件付分散を考えるのだ, との説明があったが, 本書では記述すら最早ない. 証明は末石計量の補足資料を参照のこと.

Def: 不均一分散

回帰モデルの仮定の下では無条件分散の不均一分散は実現せず, p.4 の形で定義を行う必要がある.

■p.4 説明変数/応答変数 語の用法として, 回帰モデルでない構造モデルには, この語を使うのは不適切?

■p.4 限界効果 因果効果とは限らない. 先述の通り, 回帰モデルによる分析は記述的な分析であって, 必ずしも構造的な分析だとは限らないためである. 1.1.2 章の内容は, 全体を通じて, 記述的な分析である, 回帰分析についての説明であることを理解しておかねばならない.

■p.4 注 3 の内容 『R による実証分析』は Rubin 流因果推論のフレームワークに基づく説明がなされていた. 具体的には, 記載されているように, Rubin の, 潜在結果モデルによって因果効果を定義していた. ここで重要なのは, (Rubin の) 因果効果は期待値の差分で定義されており, 一貫した関数の構造  $g(X_1, \mathbf{X}_{-1})$  を考える必要がないことだろう. この点で, Rubin の因果推論は, 本書での構造的な分析には当たらない. \*3 一方で, 因果効果を不変の関数構造  $g(X_1, \mathbf{X}_{-1})$  の下での, 状態の変動による出力の変分として捉える点で, 本書

\*3 いわゆる, 『誘導的な分析』.

は STUVA の下での Rubin 流因果推論, ととらえればいいのだろうか. 恐らく補足で説明がある.

■p.4 因果効果としての限界効果 回帰関数  $\mu$  が, 引数に対して一貫した構造を持つ際に, 限界効果は, はじめて因果効果としてみなすことが出来る.

■p.5 コントロール変数 p.1 での記述より, この回帰モデルの説明変数は全て観測可能である. 従って, 説明変数の一種であるコントロール変数にも観測可能性は必須であることには留意せよ.

■p.6 回帰関数は MSPE の意味で最も良い応答変数の予測をもたらす コア計量等で証明したことがあるだろう. 『計量経済学のための数学』 p.190 等にも証明が記載されている.

*Proof* 平均 2 乗予測誤差残差 (損失関数):

$$\text{MSPE} := \mathbb{E}[(Y - f(\mathbf{X}))^2] = \mathbb{E}[\varepsilon^2] \text{ where } \varepsilon = Y - f(\mathbf{X}),$$

の  $f$  による最小化問題を考えると, LIE と変形により:

$$\begin{aligned} \text{MSPE} &= \mathbb{E}[(Y - f(\mathbf{X}))^2] = \mathbb{E}[\mathbb{E}[\varepsilon^2 | \mathbf{X}]] \\ \varepsilon^2 &= [(Y - \mathbb{E}[Y | \mathbf{X}]) - (f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])]^2 \\ &= (Y - \mathbb{E}[Y | \mathbf{X}])^2 + (f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])^2 - 2(Y - \mathbb{E}[Y | \mathbf{X}])(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}]). \end{aligned}$$

ここで,  $\varepsilon^2$  の各項について LIE のバリエーションを用いて変形し:

$$\begin{aligned} \mathbb{E}[\mathbb{E}[(\text{第一項}) | \mathbf{X}]] &= \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y | \mathbf{X}])^2 | \mathbf{X}]] = \mathbb{E}[\mathbb{E}[Y^2 | \mathbf{X}] - (\mathbb{E}[Y | \mathbf{X}])^2] \\ \mathbb{E}[\mathbb{E}[(\text{第二項}) | \mathbf{X}]] &= \mathbb{E}[\mathbb{E}[(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])^2 | \mathbf{X}]] = \mathbb{E}[(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])^2] \\ \mathbb{E}[\mathbb{E}[(\text{第三項}) | \mathbf{X}]] &= \mathbb{E}[\mathbb{E}[-2(Y - \mathbb{E}[Y | \mathbf{X}])(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}]) | \mathbf{X}]] \\ &= -2\mathbb{E}[(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])\mathbb{E}[Y - \mathbb{E}[Y | \mathbf{X}] | \mathbf{X}]] \\ &= -2\mathbb{E}[(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])\mathbb{E}[0]] = 0, \end{aligned}$$

以上を利用して, 以下の MSPE に関する不等式を得る:

$$\begin{aligned} \text{MSPE} &= \mathbb{E}[\mathbb{E}[\varepsilon^2 | \mathbf{X}]] = \mathbb{E}[\mathbb{E}[Y^2 | \mathbf{X}] - (\mathbb{E}[Y | \mathbf{X}])^2] + \mathbb{E}[(f(\mathbf{X}) - \mathbb{E}[Y | \mathbf{X}])^2] \\ &\geq \mathbb{E}[\mathbb{E}[Y^2 | \mathbf{X}] - (\mathbb{E}[Y | \mathbf{X}])^2] \quad (f(\mathbf{X}) = \mathbb{E}[Y | \mathbf{X}] \text{ の時等号成立}). \end{aligned}$$

以上より, minimizer となる  $f$  が条件付期待値関数であることが分かった.  $\square$

■p.6 MSPE と MSE の別 MSE (Mean Squared Error) は, ある点  $(x_1, \dots, x_p)$  におけるモデルの推定値  $\hat{f}(\mathbf{x})$  が真の値  $f(\mathbf{x})$  からどれだけずれているかを測る指標であり, 以下のように定義される:

$$\text{MSE} = \mathbb{E} \left[ \left( \hat{f}(\mathbf{x}) - f(\mathbf{x}) \right)^2 \right]$$

一方, MSPE (Mean Squared Prediction Error) は, モデルの予測値  $\hat{f}(\mathbf{X})$  と実際に観測された応答変数  $Y$  とのずれを測るものであり, 次のように定義される:

$$\text{MSPE} = \mathbb{E} \left[ \left( Y - \hat{f}(\mathbf{X}) \right)^2 \right]$$

MSE は主にモデルの理論的な精度, 特にバイアスと分散に関する解析に用いられるのに対して, MSPE は未知データに対する予測性能の評価に使われる. ... 少なくとも本書では, 『計量経済学のための数学』ではどうやら MPSE も MSE で書いている.

■p.6 バイアスと分散のトレードオフ 不偏性における『バイアス』とは, 任意の点においてではなく, ある点の近辺に限ったものである点で異なる. 天気予報を常に晴れと予測するか, ちゃんと予測しようとするか, とかの例で言われるような予測スコア, 情報量の話? 機械学習の予測の情報量についての議論の文脈では大事な.

$$\text{MSE} = \underbrace{(\mathbb{E}[\hat{f}(\mathbf{X})] - f(\mathbf{X}))^2}_{\text{バイアス}} + \underbrace{\text{Var}[\hat{f}(\mathbf{X})]}_{\text{分散}}$$

■回帰分析の目的 まとめると以下の通り.

#### 回帰分析の目的

1. 興味のある説明変数による, 応答変数への限界効果を調べること. 回帰モデルを構造モデルとして見なせるならば, これは因果効果の測定に他ならない.
2. 応答変数を予測すること.

## 1.2 線形回帰モデル

## 1.3 本書の課題と構成

## 1.4 補論

■p.18 SUTVA 数学的にはたったこれだけの話, 常時暗黙に仮定されていることに留意せよ:

#### Assumption: SUTVA

潜在的結果が以下のように表されることを指す.

$$Y_i = Y_i(D_i) = \begin{cases} Y_i(1) & \text{if } D_i = 1 \\ Y_i(0) & \text{if } D_i = 0 \end{cases} \quad \text{where } D_i \in \{0, 1\}$$

つまり, 以下の 2 条件がこの式で表現されている:

1. 潜在的結果は他者の処置の影響を受けない (引数が  $D_i$  のみ)
2. 処置が均一 (二値関数, デジタルな処置)

■p.18 構造モデルの導入 構造モデルの建付けからも分かるように, 関数  $h$  の形状自体は人によって, そして処置によって変わらないことに注意せよ. SUTVA を満たしていることは容易に確認できるだろう. p.3 の議論を踏まえてか,  $e$  は加法分離可能ではない.

■p.19 構造モデルの処置効果 (1.14) 式の構造モデルにおける Rubin の意味での『処置効果』は:

$$Y_i = h(D_i, \mathbf{X}_i, e_i)$$

$$Y_i(1) - Y_i(0) = h(1, \mathbf{X}_i, e_i) - h(0, \mathbf{X}_i, e_i)$$

しかしこの場合は限界効果とは言えない. これを, 記述的な分析でないことの確認をもって確かめよう.

*Proof* (1.14) 式の形の構造モデルを考える:

$$Y_i = h(D_i, \mathbf{X}_i, e_i)$$

ここで, 観測可能な説明変数は  $(D_i, \mathbf{X}_i)$  であり, データが既知ならば以下のような形となる:

$$\mathbb{E}[Y_i | D_i, \mathbf{X}_i] = \mathbb{E}[h(D_i, \mathbf{X}_i, e_i) | D_i, \mathbf{X}_i]$$

$e_i$  は関数  $h$  に加法分離不可能な形で入るため, 最早 (単純には) LIE は適応不能. 右辺は観測不能な変数が入っているため, 識別不可能であり, 従って追加的な仮定を置かない限り記述的な分析でない.  $\square$

■p.19 Rubin の因果推論の利点  $Y_i(D_i)$  のまま, データ生成過程を特定せずに因果推論が可能となること.

■p.19 経済学における『処置効果』 多くの場合, 経済学における『処置効果』は, Rubin の意味でも, 本書の構造的な分析における因果効果の意味でも, さらに回帰モデルである構造モデルの (二値変数の場合の) 限界効果の意味においても当てはまる. 特に  $e$  が加法分離可能な構造モデルはこれに当てはめられる.

*Proof* 再び, (1.1) 式の  $e$  が加法分離可能な構造モデルを考える:

$$Y_i = h(D_i, \mathbf{X}_i) + e_i$$

$D_i$  はダミー変数 (二値変数) で, 暗黙に SUTVA は満たされる. ここで, 本書の因果効果は:

$$Y_i |_{D_i=1} - Y_i |_{D_i=0} = h(1, \mathbf{X}_i) - h(0, \mathbf{X}_i) \quad (*)$$

一方で Rubin の因果効果も以下のように表現できる:

$$Y_i(1) - Y_i(0) = h(1, \mathbf{X}_i) - h(0, \mathbf{X}_i) \quad (**)$$

先述の通り, 回帰モデルの必要条件である平均独立の仮定  $\mathbb{E}[e_i | D_i, \mathbf{X}_i] = 0$  を置くと,  $h(D_i, \mathbf{X}_i)$  は回帰関数に等しくなり, 記述的な分析, さらに言えば回帰分析に落とし込むことが出来た. 限界効果は:

$$\mathbb{E}[Y_i | 1, \mathbf{X}_i] - \mathbb{E}[Y_i | 0, \mathbf{X}_i] = h(1, \mathbf{X}_i) - h(0, \mathbf{X}_i) \quad (***)$$

ここで (\*), (\*\*), (\*\*\*) は全て同一であることから, 上の記載に合致した.  $\square$

■p.19 セレクションバイアス 主体は意思決定の中で処置を選び取る. 意思決定モデルを考えたとき, 処置はこれの中で選ばれる変数, との意味で内生変数である. 内生変数は一般に, 応答変数との間に, 交絡変数を通じて相関を持つ. 交絡変数 (共変量) が観測不能である際には, これは構造モデルにおける誤差項としてまとめられる. 結果的に, 処置は誤差項との間に相関を持つが, この状態こそ内生性の定義である.

■p.20 交絡変数 Rubin 流の因果推論の文脈で定義される.

■p.21 強い無視可能性 末石計量の 3.2 章でより丁寧な記述がある.

**Assumption: 強い無視可能性**

1. 非交絡の仮定, 条件付独立の仮定:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i$$

2. オーバーラップの仮定:

$$0 < P(D_i = 1 \mid \mathbf{X}_i = \mathbf{x}) < 1 \quad \forall \mathbf{x}$$

■p.20 非交絡の仮定の意味 非交絡の仮定は, 以下が成り立つことに対応する:

$$f(e \mid D, \mathbf{X}) = f(e \mid \mathbf{X})$$

ここで  $f$  は  $e$  の条件付分布. これを示したい.

*Proof* ChatGPT のぼっと出しの証明, 未検証 (1.14) 式の構造モデル:

$$Y_i = h(D_i, \mathbf{X}_i, e_i)$$

において,  $e_i$  は観測不能な説明変数をまとめ上げた誤差項で, 任意の処置水準  $d$  と共変量  $\mathbf{x}$  に対し  $e \mapsto h(d, \mathbf{x}, e)$  が単調増加 (あるいは単調減少) の可逆マッピングであると仮定する. このとき, 任意の  $d$  に対し

$$Y_i(d) = h(d, \mathbf{X}_i, e_i)$$

が定義でき,  $\mathbf{X}_i = \mathbf{x}$  を固定すると

$$e_i \longleftrightarrow (Y_i(1), Y_i(0))$$

は可逆対応になる.

非交絡の仮定より,

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i,$$

すなわち  $\mathbf{X}_i = \mathbf{x}$  の下で

$$f(Y_i(1), Y_i(0) \mid D_i = d, \mathbf{X}_i = \mathbf{x}) = f(Y_i(1), Y_i(0) \mid \mathbf{X}_i = \mathbf{x}) \quad (\forall d = 0, 1)$$

が成り立つ. 可逆変換の性質から, 任意の可測集合  $S \subseteq \mathcal{E}$  (誤差項の値域) について

$$\{e_i \in S\} = \{(Y_i(1), Y_i(0)) \in g(S)\}, \quad g(e) = (h(1, \mathbf{x}, e), h(0, \mathbf{x}, e)).$$

したがって

$$\begin{aligned} P(e_i \in S \mid D_i = d, \mathbf{X}_i = \mathbf{x}) &= P((Y_i(1), Y_i(0)) \in g(S) \mid D_i = d, \mathbf{X}_i = \mathbf{x}) \\ &= P((Y_i(1), Y_i(0)) \in g(S) \mid \mathbf{X}_i = \mathbf{x}) \\ &= P(e_i \in S \mid \mathbf{X}_i = \mathbf{x}), \end{aligned}$$

すなわち確率密度に直すと

$$f(e_i \mid D_i = d, \mathbf{X}_i = \mathbf{x}) = f(e_i \mid \mathbf{X}_i = \mathbf{x}), \quad \forall d, \mathbf{x}.$$

これにより  $f(e_i \mid D_i, \mathbf{X}_i) = f(e_i \mid \mathbf{X}_i)$  が示された. □

## 2 変数選択

### 2.1 設定

### 2.2 推定された予測誤差に基づく変数選択

### 2.3 情報量規準

### 2.4 変数選択の一致性と漸近最適性

### 2.5 その他のモデル評価基準

### 2.6 変数選択後の統計的推測の問題



## 付録 A 記法

■不明点 分からない記述は赤文字を用いて記載する.

■独立性  $\perp$  の記号で表す.

■ベクトル, 行列 共に  $\mathbf{}$  を用いて記載する.

■条件付期待値 本文では  $S = s, \mathbf{W} = \mathbf{w}$  である部分母集団について,  $\mathbb{E}[Y \mid s, \mathbf{w}]$  と記述されている. 任意の  $S, \mathbf{W}$  の標本空間の元についてこの関係が成立する場合, 単に  $\mathbb{E}[Y \mid S, \mathbf{W}]$  と記載することにする.

■繰り返し期待値の法則 Law of Iterated Expectation, LIE と略す.

■対角行列  $\text{diag}(a_1, \dots, a_k)$  と対角成分を具体的に記載して表現する.

■定義  $:=$  を用いて記載する.

## 付録 B 数学的準備 (書評: 『計量経済学のための数学』)

■書評的なもの 本書では数学的補遺が特に準備されていない. 数学的な準備のため, 『計量経済学のための数学』を用いて数学的準備を行うことにした. この本の前書きにはこのような記載がある:

—— 田中 久稔, 計量経済学のための数学, p.i ——

『本書では, 計量経済学の基礎を理解するために必要な最低限のトピックを厳選し, それらについてのみ集中的に解説することを心がけました. 標準的な線形代数や確率論のテキストであれば必ず扱うべき内容(「クラメールの公式」, 「基本変形」, 「ジョルダン標準形」, etc.) であっても, 計量経済学とのかかわりが薄ければ, ばっさりとカットしています.』 (強調は前川)

実際宣言通りの内容である. 計量に使う数学として必要な内容を最短経路で提示し, かつ無駄が極限までそぎ落とされている, 中々に素晴らしい内容の本だ. 1 週間もあれば読破可能なのでぜひ手に取ってほしい.

■線形写像としての行列 上記の内容から分かるように, 本書は基本変形 (掃き出し法, ガウスの消去法) から出発する一次連立方程式システムを中心とした構成ではない. この点で, 自身が把握している線形代数を扱う他の教科書 (三宅 齋藤, Simon & Blume) とは大きな違いがある. 一貫して以下の観点で議論がすすむ:

—— 田中 久稔, 計量経済学のための数学, p.43 ——

『行列とは線形写像です.』

行列の積は合成写像として定義されており, 逆行列の存在は逆写像の存在と同値な全単射と結びつけられる. ランクは像の次元として定義されており, 像とスパンの一致を根拠にして議論が進んでいく.

■射影行列, OLS の幾何的解釈 5 章では, 射影行列によってベクトルが, どのような部分空間に写されるのか (特に射影されるのか) を丁寧に説明してくれる. 図形的には, 1 のようにまとめられる. 他にも例えば, 定理 5.6 (2) は, 2SLS での第一段階回帰を経ても外生変数が保存されることの根拠となっており重要である.

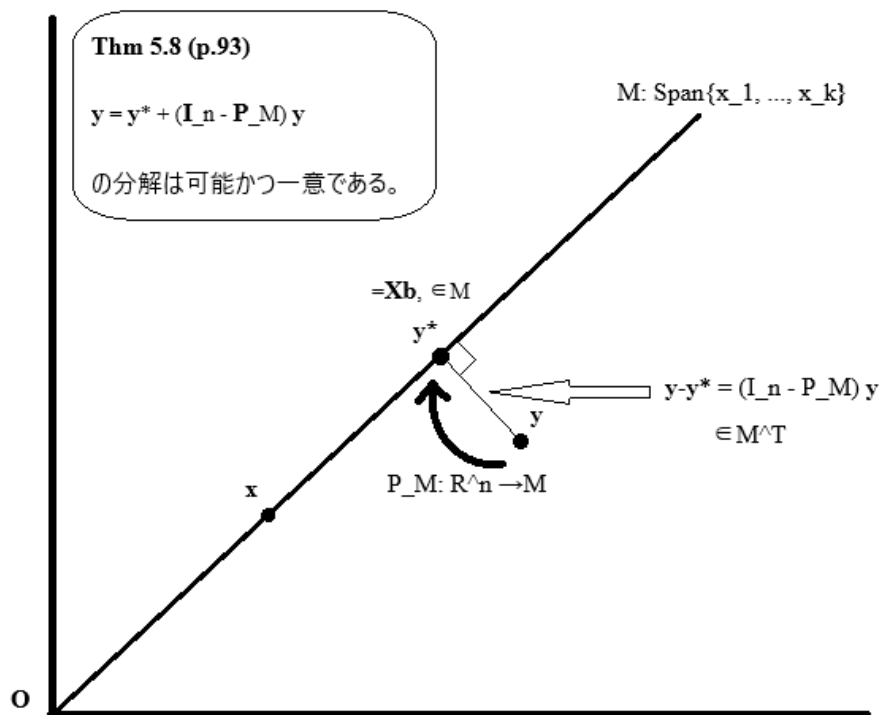


図1 直交分解のイメージ

■行列の平方根 (p.116, 問題 6.6) 行列の平方根は, 存在性については, 練習問題 (すべて解答付き!) で確認することができる. 末石本では更に半正定値行列である平方根行列の一意性が (証明無しで) 言及されている.

Def: 平方根行列

$$B^T B := B^2 = A$$

となるような  $B$  を  $A$  の平方根と呼び,  $A^{1/2}$  と表記する.

Thm: 半正定値行列の平方根行列

$L^T L = I_k$  を満たす  $L$  と, 非負の相異なる実数対角成分を持つ対角行列  $\Lambda$  によって,  $A = L \Lambda L^T$  と表現できる  $k$  次対称行列  $A$  については, 平方根行列  $A^{1/2}$  が常に存在する.

Proof 『計量経済学のための数学』の定理 6.4 で存在が正当化される, 以下の行列の対角化を考える:

$$A = L \text{diag}(\lambda_1, \dots, \lambda_k) L^T$$

$L \in \mathbb{R}^{k \times k}$  (対角化に用いる, 固有ベクトルを結合した行列) は直交行列で,  $L^T L = I_k$  を満たす.  $\text{diag}(\lambda_1, \dots, \lambda_k)$  の対角成分は  $\lambda_i \geq 0$  を満たす. 系 6.2 から  $A$  は半正定値.  $\text{diag}(\lambda_1, \dots, \lambda_k)$  に対し

$$\text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} := \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k})$$

と定義すると、これは平方根行列であることを確認できる。  $\text{diag}(\lambda_1, \dots, \lambda_k)^{1/2}$  も対角成分が非負であるから、半正定値対称行列である。そこで、対角行列の対称性と  $\mathbf{L}^\top \mathbf{L} = \mathbf{I}_k$  を利用して：

$$\begin{aligned} \mathbf{A} &= \mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top \\ &= \mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top \mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top \\ &= [\mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top] [\mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top] \end{aligned}$$

ここで、 $\mathbf{A}^{1/2}$  を以下のように定義する：

$$\mathbf{A}^{1/2} := \mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top .$$

すると、 $\mathbf{A}^{1/2}$  は対称半正定値である。対称行列であるから、先の式変形により、

$$(\mathbf{A}^{1/2})^2 = [\mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top] [\mathbf{L} \text{diag}(\lambda_1, \dots, \lambda_k)^{1/2} \mathbf{L}^\top] = \mathbf{A}$$

が常に存在することが示された。更にこの平方根行列は対称半正定値行列である。  $\square$

■回帰分析 以下の記述は、必ず理解しておくべきだろう。

—— 田中 久稔, 計量経済学のための数学, p.173-174 ——

『回帰分析はよく知られた統計的方法の一つではありますが、改めて「回帰分析とは何か、30 字以内で説明せよ」と問われれば、答えに窮する読者も多いのではないのでしょうか? (中略) 「回帰分析とは何か」と問われて「最小二乗法のことである」と答えることも目的と手段を混同した誤りであるといえます。』

—— 田中 久稔, 計量経済学のための数学, p.174 ——

『回帰分析とは、「ある変数  $X$  の値を手掛かりとして、他の変数  $Y$  の値を推測すること」』

## 付録 C 測度論的確率論

末石計量においても測度論的確率論についての記載は付録にもなく、i.i.d. 性の議論等が困難になっている。いくつかの定義については議論において利用することを避けて通れないため、利用できるものを確認しておく。詳しくは『計量経済学のための数学』の 7 ～ 10 章。

Def:  $\sigma$ -加法族

集合  $\Omega$  上の部分集合族  $\mathcal{F} \subset 2^\Omega$  が、次の 3 条件を満たすとき、 $\mathcal{F}$  を  $\sigma$ -加法族という。

1.  $\Omega \in \mathcal{F}$ .
2.  $A \in \mathcal{F}$  ならば  $A^c := \Omega \setminus A \in \mathcal{F}$ .
3. 任意の列  $\{A_n\}_{n=1}^\infty \subset \mathcal{F}$  に対して、 $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$ .

Def:  $\sigma$ -加法族の生成

$\mathcal{F}_0$  を含む最小の  $\sigma$ -加法族  $\mathcal{F}$  を、 $\mathcal{F}_0$  から生成された  $\sigma$ -加法族と呼び、以下のように表す：

$$\mathcal{F} := \sigma[\mathcal{F}_0]$$

**Def: ボレル集合族**

$\mathbb{R}^k$  上の, あらゆる直方体  $\prod_{i=1}^k (a_i, b_i)$  からなる集合族  $\mathcal{G}$  より生成された (最小の)  $\sigma$ -加法族を, ( $k$  次元) ボレル集合族といい, 以下のように表記される:

$$\mathcal{B} := \sigma[\{\prod_{i=1}^k (a_i, b_i) \mid a_i < b_i \forall i\}].$$

ボレル集合族は標本空間が  $\mathbb{R}^k$  の場合には事象族  $\mathcal{F}$  として利用する.

**Def: 可測空間**

標本空間 (確率空間の諸概念構成に用いられる全体集合)  $\Omega$  とその上の  $\sigma$ -加法族である事象族  $\mathcal{F}$  の組

$$(\Omega, \mathcal{F})$$

を可測空間という. また,  $A \in \mathcal{F}$  である事象  $A$  を可測集合という.

**Def: 確率測度**

可測空間  $(\Omega, \mathcal{F})$  上の写像  $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  が以下を満たすとき,  $\mathbf{P}$  を確率測度という.

1.  $\mathbf{P}\Omega = 1$
2.  $\mathbf{P}(A^C) = 1 - \mathbf{P}A$
3. 互いに素な列  $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$  に対して  $\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}A_n$

**Def: 確率空間**

標本空間  $\Omega$  とその上の事象族  $\mathcal{F}$ , およびこの可測空間  $(\Omega, \mathcal{F})$  上の確率測度  $\mathbf{P}$  の組

$$(\Omega, \mathcal{F}, \mathbf{P})$$

を確率空間という.

**Def: 確率質量関数**

加算集合で表現される標本空間  $\Omega = \{\omega_1, \omega_2, \dots\}$  について, 事象族としてべき集合族  $2^\Omega$  を選択した確率空間  $(\Omega, 2^\Omega, \mathbf{P})$  上で, 各  $\omega \in \Omega$  に対して

$$p(\omega) := \mathbf{P}\{\omega\}$$

で定義される関数  $p : \Omega \rightarrow [0, 1]$  を  $\mathbf{P}$  の確率質量関数という.

**Def: 確率変数**

確率空間  $(\Omega, \mathcal{F})$  上の写像 (関数)  $X : \Omega \rightarrow \mathbb{R}$  が任意の  $t \in \mathbb{R}$  について

$$\{\omega \in \Omega \mid X(\omega) \leq t\} \in \mathcal{F}$$

を満たすとき,  $X$  を確率変数という.

確率測度の定義自体には確率測度は介在しない.

**Def: 確率変数ベクトル**

確率空間  $(\Omega, \mathcal{F})$  上の写像 (ベクトル値関数)  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$  が任意のベクトル  $\mathbf{t} \in \mathbb{R}^k$  について

$$\{\omega \in \Omega \mid \mathbf{X}(\omega) \leq \mathbf{t}\} \in \mathcal{F}$$

を満たすとき,  $\mathbf{X}$  を確率変数ベクトルという.

確率変数は単に関数のこと. 『扱いにくい不確実性は全て標本空間に押し付けられている』.

**Notation: 事象に関する略記**

「関数  $X$  が  $t$  以下の値をとる」ような事象 ( $\times$  状態  $\omega$ )  $\{\omega \in \Omega \mid X(\omega) \leq t\} \in \mathcal{F}$  を以下のように記す:

$$\{X \leq t\}.$$

「関数  $X$  のとる値が  $D$  に属する」ような事象  $\{\omega \in \Omega \mid X(\omega) \in D\} = X^{-1}(D)$  を以下のように記す:

$$\{X \in D\}.$$

$X$  のある実現値  $x$  が判明したとき, 確率変数の関係を辿れば事象  $\{X \leq t\}$  where  $t \in \mathbb{R}$  の成立/不成立が (すべてではないが) 判明する. ここで成立/不成立が判明した事象を集めれば  $\sigma$ -加法族が生成される.

**Def: 確率変数が生成する事象族**

$$\sigma[X] := \sigma[\{X \leq t\} \mid t \in \mathbb{R}]$$

**Def:  $X$  可測な確率変数**

同じ可測空間上で与えられた確率変数  $X, Y$  について以下が成り立つとき,  $Y$  は  $X$  可測であるという.

$$\sigma[Y] \subset \sigma[X]$$

直感的には, 『 $X$  の値が観測できれば  $Y$  の値も分かる』,  $Y$  の情報は  $X$  より劣っている. 条件付期待値やそれに付随する LIE で登場する情報集合の包含関係に関する議論はここにたどり着く.

**Def: 分布関数**

確率空間  $(\Omega, \mathcal{F}, \mathbf{P})$  上の確率変数  $X : \Omega \rightarrow \mathbb{R}$  に対し, その分布関数  $F_X : \mathbb{R} \rightarrow [0, 1]$  を各  $t \in \mathbb{R}$  で:

$$F_X(t) := \mathbf{P}\{\omega \in \Omega \mid X(\omega) \leq t\} = \mathbf{P}\{X \leq t\}$$

で定義する. この関数  $F_X$  は次の性質を満たす:

1. 非減少性:  $x \leq y \implies F_X(x) \leq F_X(y)$ .
2. 境界条件:  $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow +\infty} F_X(x) = 1$ .

確率変数の離散/連続は分布関数の連続性によって区別される.

**Def: 密度関数**

分布関数  $F_X$  を持つ連続確率変数  $X : \Omega \rightarrow \mathbb{R}$  について, 任意の  $t \in \mathbb{R}$  に対して

$$F_X(t) = \int_{-\infty}^t f_X(x) dx, \quad f_X(x) = \frac{dF_X}{dx}(x)$$

が成り立つ可積分関数  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  が存在するとき,  $f_X$  を密度関数という.  $f_X$  は次の性質を満たす:

1. 非負性:  $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. 正規化条件:  $\int_{-\infty}^{\infty} f_X(x) dx = 1$

**Def: 確率変数ベクトルの密度関数**

分布関数  $F_{\mathbf{X}}(\mathbf{t}) = \mathbf{P}\{\mathbf{X} \leq \mathbf{t}\}$  を持つ連続な  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$  を考え, 任意の  $\mathbf{t} \in \mathbb{R}^k$  に対して

$$F_{\mathbf{X}}(\mathbf{t}) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_k} f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_k, \quad f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^k F_{\mathbf{X}}}{\partial x_1 \partial x_2 \cdots \partial x_k}(\mathbf{x})$$

が成り立つ可積分関数  $f_{\mathbf{X}} : \mathbb{R}^k \rightarrow \mathbb{R}$  が存在するとき,  $f_{\mathbf{X}}$  を密度関数という.

密度関数から確率測度を構成することも可能.

**Def: 確率変数の質量関数**

加算集合で表現される実現値  $\mathcal{X} = \{x_1, x_2, \dots\}$  を持つ離散確率変数  $X$  について,

$$p_X(x_j) := \mathbf{P}\{X = x_j\}$$

で定義される関数  $p_X : \mathcal{X} \rightarrow [0, 1]$  を確率変数  $X$  の確率質量関数という.  $p_X$  は次の性質を満たす:

1. 非負性:  $p_X(x) \geq 0 \quad \forall x \in \mathcal{X}$
2. 正規化条件:  $\sum_{j=1}^{\infty} p_X(x_j) = 1$

確率測度に対して定義したものとの差異は定義域程度か. 確率質量関数から確率測度を構成することも可能.

■指示関数 最も単純な構造を持つ確率変数として解釈できる. 『計量経済学のための数学』においては, (8.5) 式での指示関数で表現できる  $X$  と例題 7.8 で定義された  $Y$  は同じ確率関数で, 従ってこれらの確率変数から生成される  $\sigma$ -加法族なども同じであることに気付くだろう.

■単関数 確率変数の集合が線形空間になる事から, 指示関数の線形結合で表せる単関数も確率変数である.

**Def: 単関数**

確率空間  $(\Omega, \mathcal{F}, \mathbf{P})$  に対し, 関数  $X : \Omega \rightarrow \mathbb{R}$  が次の形で表されるとき,  $X$  を単関数という:

$$X(\omega) := \sum_{i=1}^n c_i \mathbb{1}_{A_i}(\omega)$$

ここで,  $c_i \in \mathbb{R}$ ,  $A_i \in \mathcal{F}$  は互いに素な集合,  $\mathbb{1}_{A_i}$  は  $A_i$  の指示関数 ( $\mathbb{1}_{A_i}(\omega) = 1$  if  $\omega \in A_i$ , o.w. 0).

**Def: 単関数の期待値**

確率空間  $(\Omega, \mathcal{F}, \mathbf{P})$  上の単関数  $X(\omega)$  に対して, その期待値  $\mathbb{E}[X]$  を次で定義する:

$$\mathbb{E}[X] := \sum_{i=1}^n c_i \mathbf{P}A_i$$

**Def: 近似単関数**

確率空間  $(\Omega, \mathcal{F}, \mathbf{P})$  上の非負かつ有限な確率変数  $X: \Omega \rightarrow [0, M]$  に対し以下で定義される:

$$X_n := \sum_{i=1}^{2^n} \frac{i-1}{2^n} M \mathbf{1}_{A_{n,i}}$$

$X_n$  を, 近似単関数と呼ぶ. ここで,  $A_{n,i}$  は  $X$  の値域を  $2^n$  等分し, このうち小さいほうから  $j$  番目の小区間に  $X$  が含まれる事象を指す.  $\{X_n\}_{i=1}^\infty$  は上に有界かつ単調増加な数列であるため以下が成り立つ:

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

一般の確率変数では単関数による近似等多様な操作を駆使し, 単関数の期待値の定義に帰着させる.

**Thm: 離散確率変数の期待値**

加算集合で表現される実現値  $\mathcal{X} = \{x_1, x_2, \dots\}$  をとる確率変数  $X: \Omega \rightarrow \mathcal{X}$  に対し, 確率質量関数  $p_X(x_j) = \mathbf{P}\{X = x_j\}$  を用いて, 期待値  $\mathbb{E}[X]$  は次の通り:

$$\mathbb{E}[X] := \sum_{j=1}^{\infty} x_j p_X(x_j)$$

**Thm: 連続確率変数の期待値**

確率密度関数  $f_X$  をもつ連続確率変数  $X: \Omega \rightarrow \mathcal{X} \subset \mathbb{R}$  に, 期待値  $\mathbb{E}[X]$  は次の通り:

$$\mathbb{E}[X] := \int_{\mathcal{X}} x f_X(x) dx$$

確率密度関数の存在と, 分布関数の絶対連続性を前提として, スティルチェス積分の利用により得られる.

**Def: 結合確率質量関数**

離散確率変数ベクトル  $\mathbf{Z} = (Y, \mathbf{X})$  が可算集合  $\mathcal{Z} = \{(y_1, \mathbf{x}_1), \dots\}$  から実現値を取るとき, 結合確率質量関数  $p_{Y, \mathbf{X}}: \mathcal{Z} \rightarrow [0, 1]$  を各  $(y, \mathbf{x}) \in \mathcal{Z}$  に対して以下で定義する:

$$p_{Y, \mathbf{X}}(y, \mathbf{x}) := \mathbf{P}\{Y = y, \mathbf{X} = \mathbf{x}\}$$

**Def: 周辺確率質量関数**

結合分布と同様の状況で周辺確率質量関数を以下で定義する:

$$p_Y(y) := \mathbf{P}\{Y = y\}, \quad p_{\mathbf{X}}(\mathbf{x}) := \mathbf{P}\{\mathbf{X} = \mathbf{x}\}$$

**Thm: 結合分布と周辺分布の関係**

$\mathcal{Y} = \{y_1, y_2, \dots\}$ ,  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$  を値域として以下が成り立つ:

$$p_Y(y) = \sum_{\mathbf{x} \in \mathcal{X}} p_{Y, \mathbf{X}}(y, \mathbf{x}), \quad p_{\mathbf{X}}(\mathbf{x}) = \sum_{y \in \mathcal{Y}} p_{Y, \mathbf{X}}(y, \mathbf{x}).$$

確率測度の  $\sigma$ -加法性から成立する.

**Def: 条件付確率質量関数**

条件付確率質量関数  $p_{Y|\mathbf{X}} : \mathcal{Z} \rightarrow [0, 1]$  を以下で定義する:

$$p_{Y|\mathbf{X}}(y | \mathbf{x}) := \frac{p_{Y, \mathbf{X}}(y, \mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})}$$

**Def: 離散確率変数の条件付期待値**

$\mathbf{X} = \mathbf{x}$  に条件付けられた,  $Y = y$  の,  $\mathcal{X}$  上の条件付期待値を以下で定義する:

$$\mathbb{E}[Y | \mathbf{X} = \mathbf{x}] := \sum_{y \in \mathcal{Y}} y p_{Y|\mathbf{X}}(y | \mathbf{x})$$

■**独立性** 『計量経済学のための数学』の 10 章 1 節を参考にして独立性を定義しておく.

## 参考文献

- [1] 末石 直也 (2024), データ駆動型回帰分析-計量経済学と機械学習の融合, 第 1 版, 日本評論社
- [2] 末石 直也 (2015), 計量経済学 ミクロデータ分析へのいざない, 第 1 版, 日本評論社
- [3] 星野 匡郎, 田中 久稔, 北川 梨津 (2023), R による実証分析: 回帰分析から因果分析へ, 第 2 版, オーム社
- [4] 田中 久稔 (2019), 計量経済学のための数学, 第 1 版, 日本評論社
- [5] InsightEdge, データ駆動型回帰分析を実装してみた, <https://techblog.insightedge.jp/entry/non-semipara>, 2025/06/18 取得