データ駆動型回帰分析 補足資料

前川 大空 *

2025 年 7 月 6 日

目次

* 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 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. データ先行の, 観測者の恣意性が入り込みうるモデルの特定化を避けることを本書の帯書きは強調している.

■一様妥当性 データ生成の母集団分布についての頑健性, といったところ.

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:

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

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

■p.2 回帰関数の識別 回帰関数 $\mathbb{E}[\,Y \mid S, \mathbf{W}\,]$ 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 $Y \mid_{S,\mathbf{W}}$ 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: 識別可能性

観測されるデータの同時分布が既知の時, $\boldsymbol{\theta}$ の値が一意に定まるならば, $\boldsymbol{\theta}$ 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 $g(S, \mathbf{W}) \coloneqq \mathbb{E}[\,Y \mid S, \mathbf{W}\,]$ 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, $\mathbb{E}[\,e \mid S, \mathbf{W}\,] \neq 0$ 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. [*2]

[*1] 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.

[*2] 構造推定の役割の一つだろう.

■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.

Proof 誤差項が加法分離可能な構造モデル:

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

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,

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

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

■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:

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

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:

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

ここで, 回帰モデルの必要条件である平均独立の仮定 $\mathbb{E}[\,e \mid D, \mathbf{X}\,] = 0$ を置くと, $h\left(D, \mathbf{X}\right)$ 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. □

■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的な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, 本書で採用されるのは, STUVA (後述) の下での Rubin 流因果推論, ととらえればいいのだろうか.

■p.4 因果効果としての限界効果 回帰関数 μ 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):

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

の f による最小化問題を考えると, LIE と変形により:

$$\begin{aligned}
\mathrm{MSPE} &= \mathbb{E}[\,(Y - f(\mathbf{X}))^2\,] = \mathbb{E}[\mathbb{E}[\,\varepsilon^2 \mid \mathbf{X}\,]] \\
\varepsilon^2 &= \left[(Y - \mathbb{E}[\,Y \mid \mathbf{X}\,]) - (f(\mathbf{X}) - \mathbb{E}[\,Y \mid \mathbf{X}\,])\right]^2 \\
&= (Y - \mathbb{E}[\,Y \mid \mathbf{X}\,])^2 + (f(\mathbf{X}) - \mathbb{E}[\,Y \mid \mathbf{X}\,])^2 - 2(Y - \mathbb{E}[\,Y \mid \mathbf{X}\,])(f(\mathbf{X}) - \mathbb{E}[\,Y \mid \mathbf{X}\,])\,.
\end{aligned}$$

ここで, ε^2 の各項について LIE のバリエーションを用いて変形し:

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

以上を利用して, 以下の MSPE に関する不等式を得る:

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

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. □

[*3] いわゆる, 『誘導的な分析』.

■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 $\widehat{f}(\mathbf{x})$ 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:

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

一方, MSPE（Mean Squared Prediction Error）は, 母集団からのランダムサンプルについて, モデルの予測値 $\widehat{f}(\mathbf{X})$ 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:

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

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. $\cdots$ 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. 以下の分解は, オーバーフィッティング (過学習) についてのトレードオフを説明するためのものである. 第二項は, $\widehat{f}$ が $\mathbf{x}$ に大きく左右されることを意味している.

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

■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:

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

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:

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

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:

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

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. □

■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:

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

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:

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

一方で Rubin の因果効果も以下のように表現できる:

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

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 $\mathbb{E}[\, e_i \mid D_i, \mathbf{X}_i \,] = 0$ を置くと, $h\left(D_i, \mathbf{X}_i\right)$ 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:

$$\mathbb{E}[\, Y_i \mid 1, \mathbf{X}_i \,] - \mathbb{E}[\, Y_i \mid 0, \mathbf{X}_i \,] = h\left(1, \mathbf{X}_i\right) - h\left(0, \mathbf{X}_i\right) \tag{$***$}$$

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

■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, 処置はこの中で選び取られる変数, 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:
$$0 < \mathrm{P}(D_i = 1 \mid \mathbf{X}_i = \mathbf{x}) < 1\ \forall \mathbf{x}$$

■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. 三輪くんが示してくれた方法を確認してみよう.

Proof ChatGPT のぱっと出しの証明, 未検証 (1.14) 式の構造モデル:
$$Y_i \;=\; h\big(D_i,\ \mathbf{X}_i,\ e_i\big)$$
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
$$Y_i(d) \;=\; h\big(d, \mathbf{X}_i, e_i\big)$$
が定義でき, $\mathbf{X}_i = \mathbf{x}$ を固定すると
$$e_i \;\longleftrightarrow\; \big(Y_i(1),\, Y_i(0)\big)$$
は可逆対応になる.
非交絡の仮定より,
$$\big(Y_i(1),\, Y_i(0)\big) \;\perp\!\!\!\perp\; D_i \;\big|\; \mathbf{X}_i,$$
すなわち $\mathbf{X}_i = \mathbf{x}$ の下で
$$f\big(Y_i(1), Y_i(0) \mid D_i = d, \mathbf{X}_i = \mathbf{x}\big) = f\big(Y_i(1), Y_i(0) \mid \mathbf{X}_i = \mathbf{x}\big) \quad (\forall\, d = 0, 1)$$
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
$$\big\{e_i \in S\big\} \;=\; \big\{(Y_i(1), Y_i(0)) \in g(S)\big\}, \quad g(e) = \big(h(1, \mathbf{x}, e),\, h(0, \mathbf{x}, e)\big).$$
したがって
$$\begin{aligned}\mathrm{P}\big(e_i \in S \mid D_i = d, \mathbf{X}_i = \mathbf{x}\big) &= \mathrm{P}\big((Y_i(1), Y_i(0)) \in g(S) \mid D_i = d, \mathbf{X}_i = \mathbf{x}\big)\\ &= \mathrm{P}\big((Y_i(1), Y_i(0)) \in g(S) \mid \mathbf{X}_i = \mathbf{x}\big)\\ &= \mathrm{P}\big(e_i \in S \mid \mathbf{X}_i = \mathbf{x}\big),\end{aligned}$$

すなわち確率密度に直すと

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

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　$\perp\!\!\!\perp$ の記号で表す.

■事象　大文字を用いて記載する.

■ベクトル, 行列　共に `mathbf` を用いて記載する.

■確率変数の値域, 集合族, 事象族　カリグラフィー `mathcal` 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　$\mathrm{diag}(a_1, \ldots, a_k)$ 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みてほしい.

■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, 図 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.

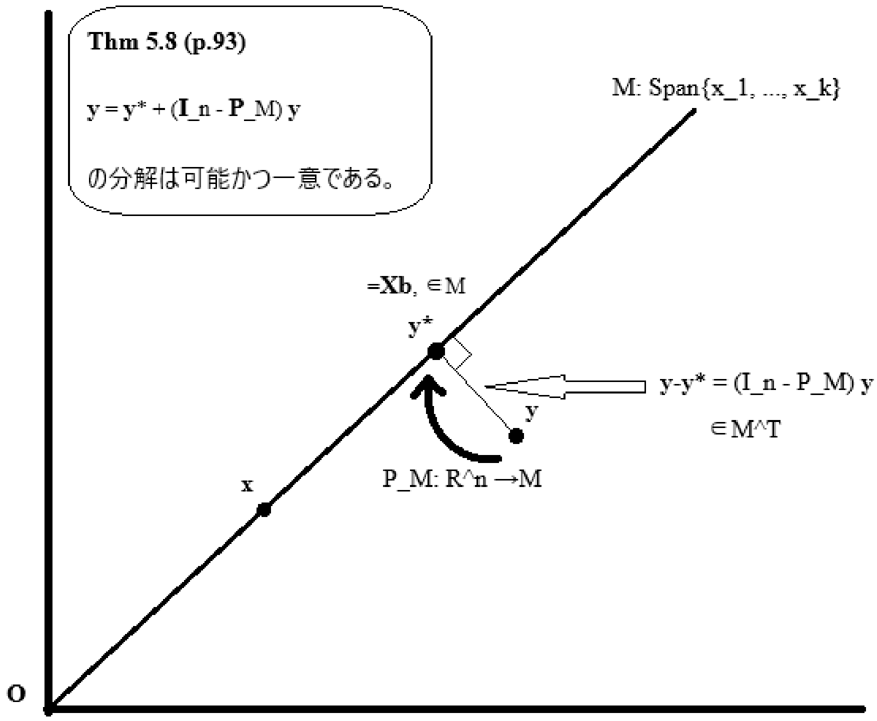

図 1　直交分解のイメージ

■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

$$\mathbf{B}^\top \mathbf{B} \coloneqq \mathbf{B}^2 = \mathbf{A}$$

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

Thm: 半正定値行列の平方根行列

$\mathbf{L}^\top \mathbf{L} = \mathbf{I}_k$ を満たす $\mathbf{L}$ と, 非負の相異なる実数対角成分を持つ対角行列 $\mathbf{\Lambda}$ によって, $\mathbf{A} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^\top$ と表現できる k 次対称行列 $\mathbf{A}$ については, 平方根行列 $\mathbf{A}^{1/2}$ 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:

$$\mathbf{A} \;=\; \mathbf{L}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_k)\,\mathbf{L}^\top$$

$\mathbf{L} \in \mathbb{R}^{k\times k}$ (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, $\mathbf{L}^\top\mathbf{L} = \mathbf{I}_k$ を満たす. $\mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ の対角成分は $\lambda_i \geq 0$ を満たす. 系 6.2 から $\mathbf{A}$ は半正定値. $\mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ に対し

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_k)^{1/2} \coloneqq \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k})$$

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. $\mathrm{diag}(\lambda_1, \ldots, \lambda_k)^{1/2}$ 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:

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

ここで, $\mathbf{A}^{1/2}$ を以下のように定義する:

$$\mathbf{A}^{1/2} \coloneqq \mathbf{L}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_k)^{1/2}\,\mathbf{L}^\top\ .$$

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,

$$(\mathbf{A}^{1/2})^2 = \left[\mathbf{L}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_k)^{1/2}\,\mathbf{L}^\top\right]\left[\mathbf{L}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_k)^{1/2}\,\mathbf{L}^\top\right] = \mathbf{A}$$

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. □

■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」』

■回帰係数の識別　まず識別の定義を確認する.

Def: 識別

興味の対象である θ を, 観測可能な量によって表現すること.

損失関数として $L(u) = u^2$ を採用し, かつ $\mathbb{E}[Y \mid \mathbf{X}] = \mathbf{X}^\top\boldsymbol{\beta}$ が成り立つ線形回帰モデルにおいて, 興味の対象である回帰係数 $\boldsymbol{\beta}$ は以下のように識別される:

Them 9.10: 回帰係数の識別

線形回帰モデルにおいて, $\mathbb{E}[\mathbf{X}\mathbf{X}^\top] \in \mathbb{R}^{k\times k}$ が正則であるならば, 以下が成り立つ:

$$\boldsymbol{\beta} = [\mathbb{E}[\mathbf{X}\mathbf{X}^\top]]^{-1}\mathbb{E}[\mathbf{X}Y]$$

■**OLSE の構成**　興味のある対象 θ を推定するために, 推測統計学では, 母集団上の確率変数を用いて推定量 $S_n = S(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ を構成する. 線形回帰モデルにおける $\boldsymbol{\beta}$ の推定量である, OLS 推定量 (OLSE) の構成の方法が本書では 2 通り示されている. 一つ目の方法は, 識別した回帰係数の期待値を標本対応させるというもの. 末石本では以下で説明する二つ目の方法を利用して OLSE を定義する. 特にこの二つ目の方法は, **MSE の記法に文献によって揺れがある理由**を与えてくれるため, 理解のため議論を整理しておこう.

Thm 9.11: 回帰係数の最良線形予測

線形回帰モデルで $\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ が正則ならば, MSE($\mathbf{b}$) を最小にする $\mathbf{b}$ は真の回帰係数 $\boldsymbol{\beta}$ に一致する, 即ち:

$$\arg\min_{\mathbf{b}} \mathbb{E}\big[(Y - \mathbf{X}^\top \mathbf{b})^2\big] = \boldsymbol{\beta}.$$

この結果を手掛かりに, LLN を利用して $\boldsymbol{\beta}$ を推定する方法が以下である.

Def: OLSE の間接的な構成

$\mathrm{MSE}(\mathbf{b}) = \mathbb{E}\big[(Y - \mathbf{X}^\top \mathbf{b})^2\big]$ をこの標本平均により**近似**したものを最小化するような $\mathbf{b}$:

$$\arg\min_{\mathbf{b}} \widehat{M}_n(\mathbf{b}) = \arg\min_{\mathbf{b}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i^\top \mathbf{b})^2, \text{ where } \widehat{M}_n(\mathbf{b}) \coloneqq \sum_{i=1}^{n}(Y_i - \mathbf{X}_i^\top \mathbf{b})^2$$

を OLSE $\widehat{\boldsymbol{\beta}}_n$ として定義する. $\mathbf{Y} \coloneqq (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$, $\mathbf{X} \coloneqq (\mathbf{X}_1, \ldots, \mathbf{X}_n)^\top \in \mathbb{R}^{n\times k}$ とベクトルを結合して表記すると, $\widehat{\boldsymbol{\beta}}_n$ は $\mathbf{X}^\top\mathbf{X}$ が正則ならば以下で与えられる:

$$\widehat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^\top\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{X}_i Y_i\right) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$$

OLSE を構成するために用いる目的関数は, MSE ではないことが分かった. 最終的には本書でも, 定義として採用されたのは後者の方法.

■**$\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ が正則ならば正定値**　線形回帰モデルにおける $\boldsymbol{\beta}$ の解釈を与える, **Thm 9.11** において, SOSC の確認のため利用されている事実だが, 特に明示して証明されていなかったため, ここで示しておく.

Thm

$\mathbf{X} \in \mathbb{R}^k$ とするとき, $\mathbb{E}[\mathbf{X}\mathbf{X}^\top] \in \mathbb{R}^{k\times k}$ が正則であるならば正定値である.

Proof 任意の非零ベクトル $\mathbf{v} \in \mathbb{R}^k$ に対して

$$\mathbf{v}^\top A\,\mathbf{v} = \mathbf{v}^\top\,\mathbb{E}[\mathbf{X}\mathbf{X}^\top]\,\mathbf{v} = \mathbb{E}\big[\mathbf{v}^\top(\mathbf{X}\mathbf{X}^\top)\mathbf{v}\big] = \mathbb{E}\big[(\mathbf{v}^\top\mathbf{X})^2\big] \;\geq\; 0$$

が成り立つので, $\mathbf{A}$ は半正定値である. A は対称かつ半正定値であるので, 直交行列 $\mathbf{Q}$ を用いて

$$\mathbf{Q}^\top\mathbf{A}\,\mathbf{Q} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$$

と対角化でき, 半正定値性から各固有値 λ_i は $\lambda_i \geq 0$ である. ここで正則性 $\det\mathbf{A} \neq 0$ から

$$\det\mathbf{A} = \prod_{i=1}^{k}\lambda_i \neq 0$$

より, 全ての i に対して $\lambda_i > 0$ が得られる. 以上より, 任意の非零ベクトル $\mathbf{v} \in \mathbb{R}^k$ に対して

$$\mathbf{v}^\top \mathbf{A}\,\mathbf{v} = \mathbf{v}^\top \mathbf{Q}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_k)\,\mathbf{Q}^\top \mathbf{v} = \sum_{i=1}^{k} \lambda_i\, w_i^2 > 0$$

(ただし $\mathbf{w} = \mathbf{Q}^\top \mathbf{v} \neq \mathbf{0}$) が成り立つので, $\mathbf{A}$ は正定値である. □

一般に, 正定値性は正則性を意味する. だが, 今回の例で示された逆の関係 (正則ならば正定値行列) は, 一般の正則な対称行列には成り立たない関係であることを注意せよ. 半正定値かつ正則ならば, 正定値である, これに関しては証明後半と同様の議論によって, 一般に同値関係が成り立つ. 例えばここを参照せよ: **証明**

Thm: 半正定値行列と正則性

半正定値行列が正定値行列であるための必要十分条件は, 正則なことである.

これを利用して, 以下のように完全な多重共線性 (フルランク性) の排除の意味を考えることが出来る.

Thm: $\mathbf{X}^\top\mathbf{X}$ に課される正則性の意味

$\mathbf{X} \in \mathbb{R}^{n\times k}$ とするとき, $\mathbf{X}^\top\mathbf{X} \in \mathbb{R}^{k\times k}$ が正則であるならば正定値である.

Proof まず, $\mathbf{X}^\top\mathbf{X}$ が半正定値であることを示す. 任意の非ゼロベクトル $\mathbf{v} \in \mathbb{R}^k$ に対して,

$$\mathbf{v}^\top\mathbf{X}^\top\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})^\top(\mathbf{X}\mathbf{v}) = \|\mathbf{X}\mathbf{v}\|^2 \geq 0$$

が成り立つため, $\mathbf{X}\mathbf{X}^\top$ は半正定値. さらに上の定理と, 正則性の仮定より題意は満たされた. □

Thm 6.8: 二次形式の最小化における十分条件 (SOSC)

関数 $f : \mathbb{R}^k \to \mathbb{R}$:

$$f(\mathbf{b}) = \frac{1}{2}\mathbf{b}^\top\mathbf{A}\mathbf{b} + \mathbf{a}^\top\mathbf{b} + c$$

で, $\mathbf{A}$ は所与の対称行列, $\mathbf{a}$ は定数ベクトル, c は定数とする. このとき, 一階条件の $\mathbf{D}f(\mathbf{b}) = \mathbf{0}$ を満たす $\widehat{\mathbf{b}}$ が $f(\mathbf{b})$ を最小化するための十分条件は, $\mathbf{A}$ が半正定値なこと. 二階の十分条件 (SOSC) という.

計量経済学で登場する大概の最適化問題の十分条件として, 上の定理から半正定値性を確認すれば十分. 実際, OLS や GLS に関しては, 一階条件が実際に目的関数を最小化することは, この確認さえすれば十分である.

■OLSE や GLSE と正定値性 OLSE や GLSE を考える際に, $\mathbf{A}$ にあたる部分が半正定値になるように構成されているわけだが, この部分は推定量を表す際に逆行列をとられる. これは正則性が必要な操作であり, 従ってこの部分は各推定量の定義をした時点で, 正定値であることが確定する.

Def 6.2: マハラノビス距離

ベクトル $\mathbf{b} \in \mathbb{R}^k$ と, k 次対称 (半) 正定値行列 $\mathbf{A}$ に対し, $\mathbf{A}$ をウェイトとするノルムを:

$$\|\mathbf{b}\|_{\mathbf{A}} \coloneqq \sqrt{\mathbf{b}^\top \mathbf{A} \mathbf{b}}$$

と定義し, ベクトル $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^k$ に対して, マハラノビス距離 $d_{\mathbf{A}}(\mathbf{b}_1, \mathbf{b}_2)$ は次で定義される:

$$d_{\mathbf{A}}(\mathbf{b}_1, \mathbf{b}_2) \coloneqq \|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{A}}$$

Def: GLS の目的関数

正定値行列 $\mathbf{W} \in \mathbb{R}^{n \times n}$ に対し, 任意のベクトル $\mathbf{e} \in \mathbb{R}^n$ に対する一般化ノルムは次で定義される:

$$\|\mathbf{e}\|_{\mathbf{W}} \coloneqq \sqrt{\mathbf{e}^\top \mathbf{W} \mathbf{e}}$$

GLS では, 残差ベクトル $\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{b}$ のこの一般化ノルムを最小化することによって推定値を求める.

重み付け行列が正定値であることについては, 末石計量の補足資料, GLS の効率性の章で触れた. 重み付け行列として不均一分散の下で, 共分散行列の逆行列を利用するのが一般的だが, これは正則性を課していることと同値で, 共分散行列の半正定値性から, 正定値性を仮定していることに他ならない. 正定値の逆行列も正定値であるため, この一般形にあたる $\mathbf{W}$ は初めから正定値として仮定されているのだと考えられる. なお, 『計量経済学のための数学』においても, 重み付け行列は対角行列に直ちに限定されて GLSE は定義される.

■GLSE の一致性の条件　GLS の一致性の条件からも得られる含意がある. 確認してみよう.

Thm 10.8, 11.6: GLS 推定量の条件

線形回帰モデル $\mathbb{E}[Y \mid \mathbf{X}] = \mathbf{X}^\top \boldsymbol{\beta}$ について, 以下の (i)〜(iii) が満たされるとする:

1. $\mathbb{E}[\mathbf{X}^\top \mathbf{X}] < \infty, \quad \mathbb{E}[Y^2] < \infty,$
2. W は $\mathbf{X}$ 可測, $\quad \mathbf{P}\{W > 0\} = 1, \mathbb{E}[W^2 \mathbf{X}^\top \mathbf{X}] < \infty, \mathbb{E}[\varepsilon^2 W^2 \mathbf{X}^\top \mathbf{X}] < \infty,$
3. $\mathbf{X}_n^\top W_n \mathbf{X}_n, \mathbb{E}[W \mathbf{X} \mathbf{X}^\top]$ は正則

このとき, GLS 推定量について一致性 $\widehat{\boldsymbol{\beta}}_n^w \xrightarrow{p} \boldsymbol{\beta}$ と漸近正規性が成立する.

期待値の有限性は期待値の線形性を経由して LLN に利用される. W の $\mathbf{X}$ 可測は LIE で利用されるが, 具体例として, 不均一分散の場合も当てはまる. 11 章で後述されるが, これは, (条件付の) 共分散行列が対角行列となり, その対角成分が $\mathbf{X}$ の関数 $\sigma(\cdot)$ として表される状況である. **Thm 9.4** での議論から, これは $\mathbf{X}$ 可測である. 正則性は先述の通りで, GLSE の表現に必要となる.

Thm 11.7: 一般化コーシー・シュワルツの不等式

任意の確率変数ベクトル $\mathbf{X} \in \mathbb{R}^k, \mathbf{Y} \in \mathbb{R}^l$ について, $\mathbb{E}[\mathbf{X}\mathbf{X}^\top], \mathbb{E}[\mathbf{Y}\mathbf{Y}^\top]$ が正則なとき, 以下が成立する:

$$\mathbb{E}[\mathbf{Y}\mathbf{X}^\top][\mathbb{E}[\mathbf{X}\mathbf{X}^\top]]^{-1}\mathbb{E}[\mathbf{X}\mathbf{Y}^\top] \leq \mathbb{E}[\mathbf{Y}\mathbf{Y}^\top]$$

ここで, $\mathbf{A} \leq \mathbf{B}$ は差分 $\mathbf{B} - \mathbf{A}$ が半正定値であることを意味する.

この不等式を利用して, GLSE における効率的な重み付け行列を特定化できる.

Thm 11.8: GLS 推定量の最適な重み付け行列

Thm 11.6 を満たす任意の確率変数 W について GLSE の漸近分散を $\mathbf{V}(W)$ として, 以下が成り立つ:

$$\mathbf{V}(\frac{1}{\sigma_\varepsilon^2(\mathbf{X})}) = \big[\mathbb{E}[\frac{\mathbf{X}\mathbf{X}^\top}{\sigma_\varepsilon^2(\mathbf{X})}]\big]^{-1} \leq \mathbf{V}(W)$$

つまり, 条件付分散 $\sigma_\varepsilon^2(\mathbf{X}_i) = \mathbb{E}[\varepsilon^2 \mid \mathbf{X}_i]$ を対角成分にもつ対角行列の逆数, 若しくはこの逆数を対角成分にもつ対角行列が最良の重み付け行列となる.

■**誤差項の** i.i.d. **性** (p.224)　当然視される気もするがあまり自明とは思えない事実. 実際, 自分は独立性のみが成立すると認識していた. ここでは, **付録 C** の測度論的確率論の諸定義, 諸定理を活用しながら証明を行う.

Thm: 誤差項の i.i.d. 性

確率変数ベクトル列 $\{\mathbf{Z}\} = \{[Y, \mathbf{X}^\top]^\top\}$ が独立同一であるならば, $\varepsilon = Y - \mathbb{E}[Y \mid \mathbf{X}]$ によって与えられる $\{\varepsilon\}$ も独立同一に分布する確率変数になる.

Proof まず, 各標本 $i \in \{1, 2, \ldots\}$ に対して $\mathbf{Z}_i$ の実現値 (データ) $\mathbf{z}_i = (y_i, \mathbf{x}_i)$ が得られた場合:

$$\varepsilon_i = g(\mathbf{Z}_i), \quad g(\mathbf{Z}_i) \coloneqq Y_i - \mathbb{E}[\,Y_i \mid \mathbf{X}_i\,],$$
$$g(\mathbf{Z}_i = \mathbf{z}_i) = y_i - \mathbb{E}[\,Y_i \mid \mathbf{X}_i = \mathbf{x}_i\,]$$

と定義できる. ここで, ε_i もまた確率変数である. このとき, **Def 9.5** より条件付期待値である第二項の $\mathbf{X}_i$ 可測性が言えて, $\mathbf{Z}_i$ 可測であることも分かる. $g(\mathbf{Z}_i)$ は $\mathbf{Z}_i$ 可測. 実際,

$$\mathbb{E}[g(\mathbf{Z}_i) \mid \mathbf{Z}_i] = \mathbb{E}[Y_i - \mathbb{E}[\,Y_i \mid \mathbf{X}_i\,] \mid \mathbf{Z}_i] = \mathbb{E}[Y_i - \mathbb{E}[\,Y_i \mid \mathbf{X}_i\,] \mid Y_i, \mathbf{X}_i] = Y_i - \mathbb{E}[\,Y_i \mid \mathbf{X}_i\,] = g(\mathbf{Z}_i)$$

の関係が, **Thm 9.8** から成り立つことが確認できる. **付録 C.1** の $\mathbf{Z}_i$ 可測の定義から, 確率変数によって生成される σ-加法族の関係性として以下が成立している:

$$\sigma[\varepsilon_i] \ \subset \ \sigma[\mathbf{Z}_i]$$

(1) 独立性.　任意の $n \in \mathbb{N}$ に対して $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ が独立であるとは, **Def 10.2.2** より:

$$\sigma[\mathbf{Z}_1], \ldots, \sigma[\mathbf{Z}_n] \quad \text{が互いに独立}$$

であることをいう. 各 $\sigma[\varepsilon_i]$ は $\sigma[\mathbf{Z}_i]$ の部分集合族であるので, 集合族の独立性 **Def 10.2.1** から, 集合族 $\mathcal{G}_i \coloneqq \sigma[\varepsilon_i] \subset \sigma[\mathbf{Z}_i]$ から選ばれた任意の事象 $A_i \in \mathcal{G}_i \subset \sigma[\mathbf{Z}_i]$ も常に独立である. よって:

$$\sigma[\varepsilon_1], \ldots, \sigma[\varepsilon_n] \quad \text{も互いに独立}$$

となり, すなわち任意の $n \in \mathbb{N}$ に対して $\varepsilon_1, \ldots, \varepsilon_n$ は互いに独立である.

(2) 同一分布性.　$\{\mathbf{Z}_i\}$ が同一分布とは, ある分布関数 $F_{\mathbf{Z}}(\cdot)$ が存在して, 全ての $\mathbf{t} \in \mathbb{R}^k$ について

$$\mathbf{P}\{\mathbf{Z}_i \leq \mathbf{t}\} = F_{\mathbf{Z}}(\mathbf{t}) \quad (i = 1, 2, \ldots)$$

であること. $\varepsilon_i = g(\mathbf{Z}_i)$ は同一の可測写像 g による変換であるから, 全ての $\tau \in \mathbb{R}$ に対して

$$\mathbf{P}\{\varepsilon_i \leq \tau\} = \mathbf{P}\{\,g(\mathbf{Z}_i) \leq \tau\} = F_\varepsilon(\tau) \quad (i = 1, 2, \ldots)$$

となり, $\{\varepsilon_i\}$ は同一分布である.

以上 (1)(2) より, **付録 C.4** の定義と整合し, $\{\varepsilon_i\}$ が i.i.d. であることが示された. □

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こう. 詳しくは『計量経済学のための数学』の 7 ∼ 11 章を参照のこと. 定義・定理番号は教科書に対応する.

C.1　確率空間 (7 章)

Def 7.1: σ-加法族

集合 Ω 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 σ-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 7.4: σ-加法族の生成

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

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

Def 7.5: ボレル集合族

$\mathbb{R}^k$ 上の, あらゆる直方体 $\Pi_{i=1}^{k}(a_i, b_i)$ からなる集合族 $\mathcal{G}$ より生成された (最小の) σ-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:

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

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) Ω とその上の σ-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 7.2: 確率測度

可測空間 $(\Omega, \mathcal{F})$ 上の写像 $\mathbf{P}: \mathcal{F} \to [0, 1]$ 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 $\mathbf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbf{P}A_n$

Def: 確率空間

標本空間 Ω 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 7.3: 確率質量関数

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

$$p\,(\omega) \coloneqq \mathbf{P}\{\omega\}$$

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

Def 7.6: 確率変数

可測空間 $(\Omega, \mathcal{F})$ 上の写像 (関数) $X : \Omega \to \mathbb{R}$ 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.

確率変数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

可測空間 $(\Omega, \mathcal{F})$ 上の写像 (ベクトル値関数) $\mathbf{X} : \Omega \to \mathbb{R}^k$ 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 ω) $\{\omega \in \Omega \mid X\,(\omega) \leq t\} \in \mathcal{F}$ 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 σ-加法族が生成される.

Def: 確率変数が生成する事象族

$$\sigma\big[X\big] \coloneqq \sigma\big[\{\{X \leq t\} \mid t \in \mathbb{R}\}\big]$$

値域が加算集合の離散確率変数については, 以下のように表現が可能.

Thm 7.2: 離散確率変数が生成する事象族

X の値域 $\mathcal{X} = X(\Omega)$ が加算集合であるとき, 以下が成り立つ:

$$\sigma[X] = \{\{X \in \mathcal{X}_0\} \mid \mathcal{X}_0 \subset \mathcal{X}\}$$

証明前半は σ-加法族であることを確認しているが, 後半は何をやっている?

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 7.7 & Thm 7.3, 7.4: 分布関数

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の確率変数 $X : \Omega \to \mathbb{R}$ に対し, その分布関数 $F_X : \mathbb{R} \to [0,1]$ 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: $\lim_{x\to-\infty} F_X(x) = 0,\ \lim_{x\to+\infty} F_X(x) = 1$.

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 7.8: 密度関数

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

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

が成り立つ可積分関数 $f_X : \mathbb{R} \to \mathbb{R}$ 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 7.9: 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 $\mathbf{X} : \Omega \to \mathbb{R}^k$ を考え, 任意の $t \in \mathbb{R}^k$ に対して

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

が成り立つ可積分関数 $f_{\mathbf{X}} : \mathbb{R}^k \to \mathbb{R}$ 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 7.10: 確率変数の質量関数

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

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

で定義される関数 $p_X : \mathcal{X} \to [0,1]$ 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: $p_X(x) \geq 0 \ \ \forall x \in \mathcal{X}$
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.

C.2 積分と期待値 (8 章)

■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 σ-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 8.1: 単関数

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

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

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 8.2: 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:

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

Def 8.3: 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 $X : \Omega \to [0, M]$ に対し以下で定義される:

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

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:

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

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 8.2: 離散確率変数の期待値

加算集合で表現される実現値 $\mathcal{X} = \{x_1, x_2, \ldots\}$ をとる確率変数 $X : \Omega \to \mathcal{X}$ 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:

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

Thm 8.3: 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 $X : \Omega \to \mathcal{X} \subset \mathbb{R}$ に, 期待値 $\mathbb{E}[\,X\,]$ は次の通り:

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

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.

C.3　条件付期待値と回帰分析 (9 章)

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 $\mathcal{Z} = \{(y_1, \mathbf{x}_1), \ldots\}$ から実現値を取るとき, 結合確率質量関数 $p_{Y,\mathbf{X}} : \mathcal{Z} \to [0,1]$ を各 $(y, \mathbf{x}) \in \mathcal{Z}$ に対して以下で定義する:

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

まずは加算集合から実現値をとる場合の, 条件付確率にまつわる諸概念を考えていこう.

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:

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

Thm 9.1: 結合分布と周辺分布の関係

$\mathcal{Y} = \{y_1, y_2, \ldots\}$, $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ 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 σ-加法性から成立する.

Def 9.1: 条件付確率質量関数

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

$$p_{Y|\mathbf{X}}(y \mid \mathbf{x}) \coloneqq \frac{p_{Y,\mathbf{X}}(y, \mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})}$$

Def 9.2: 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:

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

■条件付期待値は $\mathcal{X}$ 上の関数　なぜこのように言えるのだろうか. 条件付 pmf は実数値をとるため, とる値は定義を見る限りでは実数値に過ぎない. 確認してみよう. $\mathbf{X} = \mathbf{x} \in \mathcal{X}$ に条件付けられた条件付期待値を以下のように $\mathbf{x}$ の関数 ϕ として表現する:

$$\phi(\mathbf{x}) := \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$$

ここで, $\mathbf{x}$ は $\mathcal{X}$ 上の実現値をとることが分かっているため, $\phi : \mathcal{X} \to \mathbb{R}$ である. これは $\mathcal{X}$ 上の関数である. 自身の誤解は, 『$\mathcal{X}$ 上の関数』という言葉を, 値域が $\mathcal{X}$ である, と誤解したことによるものだった.

Notation: 条件付期待値の略記

実現値 $\mathbf{x}$ を特定しない場合は以下のように記す:

$$\mathbb{E}[Y \mid \mathbf{X}] \, .$$

これは各 $\omega \in \Omega$ に対して $\mathbb{E}[Y \mid \mathbf{X} = \mathbf{X}(\omega)]$ を対応させる, $\Omega \to \mathbb{R}$ の関数, 即ち**確率変数**と見なせる.

Thm 9.2: 繰り返し期待値の法則 (LIE)

加算集合 $\mathcal{Y} = \{y_1, y_2, \ldots\}$, $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ 上の離散確率変数 $Y, \mathbf{X}$ に対して以下が成り立つ:

$$\mathbb{E}[\mathbb{E}[Y \mid \mathbf{X}]] = \mathbb{E}[Y] \, .$$

証明では外側の期待値について, 条件付期待値を確率変数と見なせることによって, **Thm 8.2** を使用できる. 更に, 各実現値を考えるにあたっては条件付期待値は $\mathcal{X}$ 上の関数としても考えられたため, 証明 (p.178) の一行目の変形が得られる. 後は明らか.

Thm 9.3: 条件付期待値の性質

加算集合 $\mathcal{Y} = \{y_1, y_2, \ldots\}$, $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ 上の離散確率変数 $Y, \mathbf{X}$, 任意の関数 $g(\mathbf{X})$ に対して:

$$\mathbb{E}[g(\mathbf{X})Y \mid \mathbf{X}] = g(\mathbf{X})\mathbb{E}[Y \mid \mathbf{X}] \, .$$

$\mathcal{Y}_0 \subset \mathcal{Y}$ と Y の取る値を限定している部分はテクニカル.

Thm 9.4: 条件付期待値と X 可測性

条件付期待値 $\mathbb{E}[Y \mid \mathbf{X}]$ は $\mathbf{X}$ 可測である. さらに, 任意の $\mathbf{X}$ 可測集合 $A \in \sigma[\mathbf{X}]$ で以下が成り立つ:

$$\mathbb{E}[\mathbb{1}_A \, \mathbb{E}[Y \mid \mathbf{X}]] = \mathbb{E}[\mathbb{1}_A \, Y] \, .$$

証明には教科書で触れられていない行間が存在する. 任意の $\mathbf{X}$ の関数 Y が $\mathbf{X}$ 可測であることを示さねばならないはずだが, 教科書では特に記載がない. 問題 7.6 でも特定の関数について可測性を示すにとどまっている. 後半の証明は **Thm 7.2**, **Thm 9.3**, **Thm 9.2** による.

■一般の条件付期待値　いよいよ離散の場合に限らない, 一般の条件付期待値を考える.

Def 9.3: 部分 σ-加法族

σ-加法族 $\mathcal{F}$ の部分集合族 $\mathcal{G} \subset \mathcal{F}$ が, σ-加法族の条件を満たすとき, $\mathcal{G}$ を $\mathcal{F}$ の部分 σ-加法族という.

Def 9.4: 一般の条件付期待値

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の確率変数 Y と, $\mathcal{F}$ の部分 σ-加法族 $\mathcal{G} \subset \mathcal{F}$ に対して, 以下の性質を持つ確率変数 W を, Y の $\mathcal{G}$ に条件付けられた期待値といい, $W = \mathbb{E}[Y \mid \mathcal{G}]$ と書く:

1. $\mathbb{E}[X \mid \mathcal{G}]$ は $\mathcal{G}$-可測, 即ち任意の実数 t について $\{W \leq t\} \in \mathcal{G}$
2. 任意の $A \in \mathcal{G}$ に対して以下が成立する:

$$\mathbb{E}[\mathbb{1}_A W] = \mathbb{E}[\mathbb{1}_A Y]$$

とくに, $A = \Omega$ とするとき $\mathbb{E}[W] = \mathbb{E}[Y]$

$\mathcal{G}$-可測はここで初めて定義された言葉. **Thm 9.4** の一般化を定義として議論が始まっている.

■部分 σ-加法族の意味　記述を眺めてみよう. いい解説.

田中 久稔, 計量経済学のための数学, p.183

『事象族 $\mathcal{F}$ には「試行の結果を観測することによって明らかになる情報の全て」という意味があったことを思い出しましょう. とくに (中略) 事象 $\{Y = t\}$ はつねに $\mathcal{F}$ に含まれました. したがって, **情報 $\mathcal{F}$ のすべてにアクセス可能なら, そのときには Y の実現値が観測可能ということになります.** 』 (強調は前川)

田中 久稔, 計量経済学のための数学, p.183

『 $\mathcal{F}$ よりも粗い情報である $\mathcal{G} \subset \mathcal{F}$ のもとでは, Y の実現値を直接観察できるとは限りません. そのときには, **手に入る情報 $\mathcal{G}$ を活用して Y の代用品を見つける**しかありません. 条件付き期待値 $\mathbb{E}[Y \mid \mathcal{G}]$ が, その代用品だというわけです. 』(強調は前川)

以上の議論は, 以下のように結論づけられる.

田中 久稔, 計量経済学のための数学, p.183

『定義 9.4 は, 「条件付き期待値 $\mathbb{E}[Y \mid \mathcal{G}]$ は, 限られた情報 $\mathcal{G}$ のもとでの Y の代用品である」 ということを述べています.』

■「合理的期待」　さらに例として, マクロ経済学の合理的期待を例示して **Def 9.4** の意味が説明される.

Assumption: 合理的期待

$$\mathbb{E}\big[\mathbb{E}[Y_{t+1} \mid \mathcal{G}_t]\big] = \mathbb{E}[Y_{t+1}]$$

t 期の principle にとって, 『観測によって明らかになる情報のすべて』 $\mathcal{F}$ は未来の情報も含み, 完全には入手できない. 代わりに利用できるのは t 期までの情報 $\mathcal{G}_t \subset \mathcal{F}$ に限られる. 期待値 $\mathbb{E}[Y_{t+1} \mid \mathcal{G}_t]$ が $\mathcal{G}_t$ 可測であることが **条件 1** だが, これはより細かな情報 $\mathcal{G}_{t+1}$ を principle が利用することを, $\mathcal{G}_t \subset \mathcal{G}_{t+1}$, $\mathcal{G}_{t+1} \not\subset \mathcal{G}_t$ の包含関係から禁止していると読み取れる. 次に **条件 2** についてもその意味を確認しよう. $A = \Omega$ のケースにおいて, $\mathbb{E}[\mathbb{E}[Y_{t+1} \mid \mathcal{G}_t]] = \mathbb{E}[Y_{t+1}]$ が条件の書き換えだが, 将来の変数 Y_{t+1} の期待値を, 現時点 t の情報 $\mathcal{G}_t$

に基づいて見積もり, その見積もりの平均をとったものが左辺にあたるが, これは右辺の将来の変数そのものの期待値と一致する, ということである. これは合理的期待の核心的な前提であり, 「予想は体系的に偏らない」ことを意味する. この点で, **条件 2** は, 「条件付き期待値が "整合的" である」ことを数学的に保証しており, 経済学的には「期待が体系的に誤っていない」という仮定と一致する. 合理的期待の仮定とは, このように確率論的に定義された条件付き期待値の性質を, 経済主体の行動に当てはめることである.

■一般化繰り返し期待値の法則　いよいよ LIE の一般化を示す.

Thm 9.6: Tower Property (塔の性質), 一般化 LIE

$\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ をすべて σ-加法族とし, 以下が成り立つ:

$$\mathbb{E}\big[\mathbb{E}[Y \mid \mathcal{G}] \mid \mathcal{H}\big] = \mathbb{E}\big[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}\big] = \mathbb{E}[Y \mid \mathcal{H}]$$

Proof $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}]\mathcal{H}] = \mathbb{E}[Y \mid \mathcal{H}]$ を証明したい. まず $W = \mathbb{E}[Y \mid \mathcal{G}]$ とおく. **Def 9.4** によって W は $\mathcal{G}$ 可測 ($\because$ **条件 1**) かつ任意の $A \in \mathcal{G}$ について $\mathbb{E}[\mathbb{1}_A W] = \mathbb{E}[\mathbb{1}_A Y]$ を満たす ($\because$ **条件 2**). 次に $V = \mathbb{E}[W \mid \mathcal{H}]$ とおく. 同様に V は $\mathcal{H}$ 可測 ($\because$ **条件 1**) かつ任意の $B \in \mathcal{H}$ について $\mathbb{E}[\mathbb{1}_B V] = \mathbb{E}[\mathbb{1}_B W]$ が成り立つ ($\because$ **条件 2**). ここで仮定 $\mathcal{H} \subset \mathcal{G}$ より任意の $B \in \mathcal{H}$ は $B \in \mathcal{G}$.
以上から:

$$\mathbb{E}[\mathbb{1}_B V] = \mathbb{E}[\mathbb{1}_B W] = \mathbb{E}[\mathbb{1}_B Y].$$

よって $\mathbb{E}[Y \mid \mathcal{H}] = V = \mathbb{E}\big[W \mid \mathcal{H}\big] = \mathbb{E}\big[\mathbb{E}[Y \mid \mathcal{G}] \mid \mathcal{H}\big]$ が得られる.
次に, $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[Y \mid \mathcal{H}]$ を示したい. $\mathcal{H} \subset \mathcal{G}$ であるから, $Z = \mathbb{E}[Y \mid \mathcal{H}]$ は $\mathcal{G}$ 可測. ここで, $X = Z$ について, X も $\mathcal{G}$ 可測である (**条件 1**). さらに, 任意の $A \in \mathcal{G}$ について $\mathbb{E}[\mathbb{1}_B X] = \mathbb{E}[\mathbb{1}_B Z]$ が成立する (**条件 2**). 以上から X は条件付期待値としての条件を満たし, **Def 9.4** から:

$$\mathbb{E}[Y \mid \mathcal{H}] = Z = X = \mathbb{E}[Z \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}]$$

が成立する. よって, 目標の三つの条件付期待値が等式で結ばれる. □

最も粗い情報 $\mathcal{H}$ によって条件付期待値は左右されている.

Thm 9.7: $\mathbb{E}[Y \mid \mathcal{G}]$ の性質

Y, W を確率空間上の確率変数, $\mathcal{G}$ を $\mathcal{F}$ の部分 σ-加法族とするとき, 以下が成立する:

1. 定数 c について $\mathbb{E}[c \mid \mathcal{G}] = c$,
2. 任意の定数 a, b について, $\mathbb{E}[aY + bW \mid \mathcal{G}] = a\mathbb{E}[Y \mid \mathcal{G}] + b\mathbb{E}[W \mid \mathcal{G}]$,
3. W が $\mathcal{G}$-可測であるとき, $\mathbb{E}[YW \mid \mathcal{G}] = W \cdot \mathbb{E}[Y \mid \mathcal{G}]$.

3 の証明には優収束定理が必要らしい. (証明: **問題 9.5**, 未確認)

Def 9.5: 確率変数への条件付期待値

$(\Omega, \mathcal{F}, P)$ 上の確率変数 Y と, 同空間上の確率変数ベクトル $\mathbf{X} : \Omega \to \mathbb{R}^k$ に対し, $\mathcal{G} = \sigma[\mathbf{X}] \subset \mathcal{F}$ 上に条件付けられた, 条件付期待値 $\mathbb{E}[Y \mid \mathcal{G}]$ を Y の $\mathbf{X}$ に条件付けられた期待値といい, $\mathbb{E}[Y \mid \mathbf{X}]$ と書く.

未確認

Thm 9.8: 条件付期待値の性質

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の確率変数 Y, W と確率変数ベクトル $\mathbf{X}$ に対し, $\mathbb{E}[\,Y \mid \mathbf{X}\,]$ は以下を満たす:

1. 定数 c について $\mathbb{E}[\,c \mid \mathbf{X}\,] = c$
2. 任意の定数 a, b について $\mathbb{E}[\,aY + bZ \mid \mathbf{X}\,] = a\,\mathbb{E}[\,Y \mid \mathbf{X}\,] + b\,\mathbb{E}[\,Z \mid \mathbf{X}\,]$
3. W が $\mathbf{X}$ 可測のとき $\mathbb{E}[\,W \cdot Y \mid \mathbf{X}] = W \cdot \mathbb{E}[\,Y \mid \mathbf{X}\,]$
4. $\mathbb{E}\big[\,\mathbb{E}[\,Y \mid W, \mathbf{X}] \mid \mathbf{X}\,\big] = \mathbb{E}[\,Y \mid \mathbf{X}\,]$
5. $\mathbb{E}\big[\mathbb{E}[Y \mid \mathbf{X}]\big] = \mathbb{E}[Y]$

上の諸性質は, **Thm 9.7** に確率変数から生成された σ-加法族を対応させたものとして得られる. とくに 3 については, 適当な関数 g を用いて $W = g(\mathbf{X})$ と表現されるものは $\mathbf{X}$ 可測なので, **Thm 9.3** に対応する結果となる $\mathbb{E}[\,g(\mathbf{X}) \cdot Y \mid \mathbf{X}] = g(\mathbf{X}) \cdot \mathbb{E}[\,Y \mid \mathbf{X}\,]$ が得られる.

C.4 大数の法則と推定量の一致性 (10 章)

■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.

Def 10.1: 事象の独立性

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の事象 $A_1, \ldots, A_n \in \mathcal{F}$ が互いに独立であるとは, 以下が成り立つことをいう:

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbf{P}A_i$$

Def 10.2.1: 集合族の独立性

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の部分集合族 $\mathcal{G}_1, \ldots, \mathcal{G}_n \subset \mathcal{F}$ が互いに独立であるとは, これらの集合族から選ばれた任意の事象 $A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$ がつねに独立になることをいう.

Def 10.2.2: 確率変数ベクトルの独立性

確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上の確率変数 $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ が独立であるとは, これらが生成する σ-加法族 $\sigma[\mathbf{Z}_i] \subset \mathcal{F}$ に関して, $\sigma[\mathbf{Z}_1], \ldots, \sigma[\mathbf{Z}_n]$ が独立であることを指す. とくに, $\mathbf{Z}_1, \mathbf{Z}_2$ が独立なことを以下のように表す:

$$\mathbf{Z}_1 \perp\!\!\!\perp \mathbf{Z}_2$$

Thm 10.1: 確率変数ベクトルの独立性と確率測度

$\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ が独立である必要十分条件として, 任意の $\mathbf{t}_1, \ldots, \mathbf{t}_n$ で以下が成立することがある:

$$\mathbf{P}\{\mathbf{Z}_1 \leq \mathbf{t}_1, \ldots, \mathbf{Z}_n \leq \mathbf{t}_n\} = \mathbf{P}\{\mathbf{Z}_1 \leq \mathbf{t}_1\} \cdots \mathbf{P}\{\mathbf{Z}_n \leq \mathbf{t}_n\}$$

Corr 10.1: 連続確率変数における独立性

連続確率変数 $X_1, \ldots, X_n$ と, これを成分とするベクトル $\mathbf{X}$ が, 周辺密度 $f_{X_i}(t_1)$ と, 結合密度関数 $f_{\mathbf{X}}(\mathbf{t})$ を持つとする. このとき, 次が成り立つことが確率変数の独立性の必要十分条件である:

$$f_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^{n} f_{X_i}(t_i)$$

Def: 独立同一な確率収束列, i.i.d. 性

確率変数列 $\{\mathbf{Z}_n\}_{n=1}^{\infty}$ が以下を満たすとき, これは独立同一に分布する (i.i.d.) 確率変数列という:

1. 任意の $n \in \mathbb{N}$ に対して, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ は独立である.
2. 同一の分布関数 $F : \mathbb{R}^k \to [0,1]$ が存在して, 全ての $\mathbf{t} \in \mathbb{R}^k$ について, 以下が成立する:

$$\mathbf{P}\{\mathbf{Z}_1 \leq \mathbf{t}\} = \mathbf{P}\{\mathbf{Z}_2 \leq \mathbf{t}\} = \cdots = F(\mathbf{t})$$

Def 10.3: 確率収束 (ベクトル列)

確率変数ベクトル列 $\{\mathbf{W}_n\}$ が $\mathbf{a} \in \mathbb{R}^k$ に確率収束するとは, 任意の $\varepsilon > 0$ に対して

$$\lim_{n\to\infty} \mathbf{P}\big\{\|\mathbf{W}_n - \mathbf{a}\| < \varepsilon\big\} = 1$$

が成立することであり, 確率収束を $\mathbf{W}_n \xrightarrow{p} \mathbf{a}$ と表記する.

Thm 10.3: 連続写像定理 (CMT)

確率変数ベクトル列 $\mathbf{W}_n \xrightarrow{p} \mathbf{a}$ と $\mathbf{a} \in \mathbb{R}^d$ において連続な関数 $f : \mathbb{R}^k \to \mathbb{R}$ について, 以下が成り立つ:

$$f(\mathbf{W}_n) \xrightarrow{p} f(\mathbf{a})$$

関数の連続性, 確率測度の単調性, 確率収束の定義により示される.

Def: 確率収束 (行列)

確率行列列 $\{\mathbf{A}_n\} \subset \mathbb{R}^{k\times l}$ が行列 $\mathbf{A}$ に確率収束するとは, $\mathbf{A}_n$ の各成分 $a_{n,ij}$ がそれぞれ $\mathbf{A}$ の各成分 a_{ij} に確率収束することを指し, 以下のように表記する:

$$\mathbf{A}_n \xrightarrow{p} \mathbf{A}$$

Thm: 逆行列の確率収束

確率行列列 $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ かつ $\det(\mathbf{A}) \neq 0$ のとき,

$$\mathbf{A}_n^{-1} \xrightarrow{p} \mathbf{A}^{-1}$$

CMT によって示される. ただし, 確率変数ベクトルの確率収束とその各成分の確率収束が必要十分条件の関係にあることを利用する必要がある筈, 教科書では特に記載がない.

Lemma 10.1: チェビシェフの不等式

任意の確率変数 X と, 任意の $\varepsilon > 0$ に関して, 以下が成り立つ:

$$\mathbf{P}\big\{|X| \geq \varepsilon\big\} \leq \frac{\mathbb{E}[X^2]}{\varepsilon^2}$$

チェビシェフの不等式によって以下の定理が示される.

Thm 10.5: 大数の法則 (WLLN, LLN)

i.i.d. な確率変数列 $\{Z_i\}_{i=1}^n$, $Var(Z) < \infty$ について, 標本平均 $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ は以下を満たす:

$$\bar{Z}_n \xrightarrow{p} \mathbb{E}[Z]$$

Thm 10.6: 大数の法則 (ベクトル列)

i.i.d な確率変数ベクトル列 $\{\mathbf{Z}_i\} \subset \mathbb{R}^k$ について $\mathbb{E}|Z_1|, \ldots, \mathbb{E}|Z_k| < \infty$ であるとき, $\overline{\mathbf{Z}}_n$ は以下を満たす:

$$\overline{\mathbf{Z}}_n \xrightarrow{p} \mathbb{E}[\mathbf{Z}]$$

標本平均も確率変数である. 一般の推定量も定義より確率変数である.

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