

令和6年度 中級計量経済学
講義ノート付録 数学補論（行列と確率）

このノートでは、計量経済学で使う、基本的な線形代数の知識を紹介する。また、確率変数の行列に関する議論も少し紹介する。計量経済学の学習に必要な数学としては、他に基本的な確率論と統計学の知識があるが、これらは中級の統計学などの授業すでに学習していることを前提とする。これらの数学の知識は、計量経済学の理論を理解するのに必要であるのはもちろんのこと、実際にデータの分析をする際に使用する統計ソフトの説明書やヘルプファイルを読む際にも必要になる。

1.1 ベクトルと行列

この講義では、ベクトルといえば、特に断りのない限り、列ベクトル（縦長のもの）をさす。 \mathbf{a} を p 次元のベクトルとする。 a_1, \dots, a_p をベクトル \mathbf{a} の要素として、次のように表記する。

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}. \quad (1)$$

行ベクトルを表記するときは、転置の記号 $(')$ を使う。上で定義した \mathbf{a} のとき、

$$\mathbf{a}' = (a_1, a_2, \dots, a_p). \quad (2)$$

となる。

行列とは、長方形の形に数字を並べたものである。 A を $n \times k$ の行列とする。 A の i 行 j 列の要素を a_{ij} として、

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \quad (3)$$

と表記する。 A の転置行列は

$$A' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1k} & a_{2k} & \dots & a_{nk} \end{pmatrix} \quad (4)$$

である。 $\mathbf{a}_i = (a_{1i}, a_{2i}, \dots, a_{ni})'$ とすると、 A という行列は、

$$A = (\mathbf{a}_1, \dots, \mathbf{a}_k) \quad (5)$$

のように、書くことができる。

行列の和と差 A と B を、 $n \times k$ の行列とする。すると、行列の和と差は、各要素の和と差を探ることにより定義することができる。

$$A + B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{pmatrix} \quad (6)$$

$$= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1k} + b_{1k} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2k} + b_{2k} \\ \dots & \dots & \dots & \dots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \dots & a_{nk} + b_{nk} \end{pmatrix}. \quad (7)$$

$$A - B = \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \dots & a_{1k} - b_{1k} \\ a_{21} - b_{21} & a_{22} - b_{22} & \dots & a_{2k} - b_{2k} \\ \dots & \dots & \dots & \dots \\ a_{n1} - b_{n1} & a_{n2} - b_{n2} & \dots & a_{nk} - b_{nk} \end{pmatrix}. \quad (8)$$

ベクトルの積 a と b を p 次元のベクトルとする。次の用語はよく使われる。

- 内積: $a'b = \sum_{i=1}^p a_i b_i$.
- もし、 $a'b = 0$ なら、 a と b は直交している。
- 外積: ab' 。これは、 $p \times p$ の行列であり、その ij 番の要素は、 $a_i b_j$ 。

行列の積 A と B を、それぞれ $n_a \times k_a$ と $n_b \times k_b$ の行列とする。もし、 $k_a = n_b$ なら、行列の積 AB を定義することでき、 AB の ij 番の要素は $\sum_{m=1}^{k_a} a_{im} b_{mj}$ である。

$A = (\mathbf{a}_1, \dots, \mathbf{a}_n)'$ と $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)'$ とすると、

$$A'B = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}'_i \quad (9)$$

と書ける。この行列の積の表現は、計量経済学でよく使用される。

転置の性質 次の計算が可能な A, B について以下の性質が成り立つ。

- $(A')' = A$
- $(A + B)' = A' + B'$
- $(AB)' = B'A'$

用語

- すべての要素が 0 であるベクトル（行列）をゼロベクトル（ゼロ行列）という。
- $n = k$ なら正方行列である。

- 正方行列 A が $A = A'$ なら、対称という。
- 正方行列 A が、すべての $i \neq j$ について、 $a_{ij} = 0$ なら、対角行列という。
- すべての対角要素が 1 である対角行列は、恒等行列(単位行列)といい、 I で表す。
- 対称行列 A が、ゼロベクトル以外のすべてのベクトル \mathbf{x} について、 $\mathbf{x}'A\mathbf{x} > 0$ なら、正值定符号といい。正值定符号な行列は、逆行列をとれる。
- 対称行列 A が、すべてのベクトル \mathbf{x} について、 $\mathbf{x}'A\mathbf{x} \geq 0$ なら、半正值定符号といい。
- n 次元の正方行列 A について、 $tr(A) = \sum_{i=1}^n a_{ii}$ を A のトレースという。

線形独立と階数 k 個のベクトル $(\mathbf{a}_1, \dots, \mathbf{a}_k)$ は次の条件を満たすとき、線形独立(一次独立)という。その条件とは、 c_1, \dots, c_k を変数とした

$$c_1\mathbf{a}_1 + \dots + c_k\mathbf{a}_k = 0 \quad (10)$$

という連立方程式の解が $c_1 = \dots = c_k = 0$ だけであるというものである。

- あるベクトルの組で張られる空間とは、それらのベクトルの線形結合で書けるベクトルの集合である。ある空間の次元とは、その空間を張るのに最小限必要なベクトルの数である。
- 行列の列階数とは、その行列を構成する列ベクトルで張られる空間の次元である。
- ある行列は、もし、列階数が行列の列数と同じならば、列フルランク行列といい。
- 行階数も同じように定義できる。行階数と、列階数はおなじである。
- もし、 $rank(A) = \min(n, k)$ なら、 A はフルランク行列である。

公式:

$$rank(A) = rank(A'A) = rank(AA'). \quad (11)$$

逆行列 A, B を正方行列とする。もし、 $AB = BA = I$ であるなら、 B は A の逆行列であるといい。 A の逆行列は、常に存在するとは限らない。もし、 A が正方行列で、フルランクなら、逆行列は存在し、また一意である。

- もし、 A の逆行列がとれるなら、 A を正則といい。
- A の逆行列は、 A^{-1} と表記する。

公式:

$$(A^{-1})' = (A')^{-1}, \quad (12)$$

$$(AB)^{-1} = B^{-1}A^{-1}. \quad (13)$$

行列における微分 ベクトルや行列を微分することがあるが、これは、単に要素ごとに微分を行うだけである。ここでは、行列の微分に関するいくつかの公式をまとめておく。

$$\frac{\partial \mathbf{c}' \mathbf{x}}{\partial \mathbf{x}} = \mathbf{c}, \quad (14)$$

$$\frac{\partial A \mathbf{x}}{\partial \mathbf{x}'} = A, \quad (15)$$

$$\frac{\partial \mathbf{x}' A \mathbf{x}}{\partial \mathbf{x}} = (A + A') \mathbf{x}. \quad (16)$$

1.2 確率変数のベクトルと行列

ここでは、中級の統計学程度の確率論は、すでに学習が終わっているものとし、確率変数であるベクトルと行列についてのみ議論するものとする。

\mathbf{Y} を確率変数のベクトルとする。

- 期待値: $E(\mathbf{Y})$ は、 \mathbf{Y} の i 番目の要素の確率変数の期待値を i 番目の要素にもつベクトルである。
- 分散: $var(\mathbf{Y}) = E\{(\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})'\}$ 。ただし、 $\mu_{\mathbf{Y}} = E(\mathbf{Y})$ である。分散は行列で、その対角要素は \mathbf{Y} の各要素の分散となっており、その非対角要素は \mathbf{Y} の二つの要素の共分散である。また $var(\mathbf{Y})$ の行列を分散共分散行列と呼ぶ。
- 共分散: $\sigma_{\mathbf{XY}} = E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})']$ 。

公式 \mathbf{X} を確率変数のベクトルとし、 \mathbf{A} を固定された（非確率的な）行列とする。

$$E(\mathbf{AX}) = \mathbf{AE}(\mathbf{X}), \quad (17)$$

$$var(\mathbf{AX}) = \mathbf{A} var(\mathbf{X}) \mathbf{A}'. \quad (18)$$

- 繰り返し期待値の法則: 2組の確率変数 \mathbf{X} と \mathbf{Y} について、 $E(\mathbf{Y}) = E\{E(\mathbf{Y}|\mathbf{X})\}$ が成り立つ。右辺の2つめの期待値は \mathbf{X} を条件として \mathbf{Y} の期待値をとっており、1つめの期待値は \mathbf{X} についての期待値をとったものである。これは最小二乗推定量や操作変数推定量の統計的性質を明らかにする際に繰り返し使われる重要な性質である。

1.2.1 正規分布とカイ²乗分布

正規分布 正規分布は計量経済学で最も重要な分布である。ベクトルの場合も正規分布を定義できる。

- 正規分布は、平均のベクトル μ と、分散共分散行列 Σ という二つのパラメーターにより決まる。
- 平均が μ 、分散が Σ の正規分布を $N(\mu, \Sigma)$ と表記する。
- 重要な特徴: もし、 $\mathbf{Y} \sim N(\mu, \Sigma)$ であるなら、 A を固定された行列としたとき、 $A\mathbf{Y} \sim N(A\mu, A\Sigma A')$ である。
- もし、 $(X, Y)'$ が、正規分布で X と Y が無相関 ($\sigma_{XY} = 0$) なら、 X と Y は独立である。

カイ²乗分布 自由度 m のカイ²乗分布とは、 m 個の独立な標準正規確率変数の二乗和の分布であり、 χ_m^2 と表記する。なお、 $\mathbf{Y} \sim N(0, I_p)$ とすると、 $\mathbf{Y}'\mathbf{Y} \sim \chi_p^2$ である。また、 $\mathbf{Y} \sim N(0, \Sigma)$ で Σ が正值定符号なら、 $\mathbf{Y}'\Sigma^{-1}\mathbf{Y} \sim \chi_p^2$ である。

なお、カイ²乗確率変数を自由度で割ったものの分布を、 $F_{m,\infty}$ 分布という。つまり、 $X \sim \chi_m^2$ のとき、 $X/m \sim F_{m,\infty}$ である。

1.3 標本平均の分布と、漸近(大標本)理論

母集団から無作為に選ばれた n 個の観測値を大きさ n の無作為標本という。

無作為標本 $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ は、次の性質を持つ。

- \mathbf{Y}_i は、すべての $i = 1, \dots, n$ について、同じ分布を持つ。これを、 $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ は同一に分布しているという。
- $\mathbf{Y}_i, i = 1, \dots, n$ は独立に分布している。

独立で同一に分布していることを、i.i.d.(independently and identically distributed) という略称で表記する。

この講義では、最後のトピックである時系列解析の部分を除き、無作為標本つまり、i.i.d. 標本を仮定する。

\mathbf{Y} の平均を $\mu_{\mathbf{Y}}$ 、分散を $\Sigma_{\mathbf{Y}}$ とする。

標本平均 標本平均は、最も基本的な統計量であり、ベクトルの場合にも定義できる。それは、

$$\bar{\mathbf{Y}} = \frac{1}{n}(\mathbf{Y}_1 + \mathbf{Y}_2 + \dots + \mathbf{Y}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i \quad (19)$$

である。

計量経済学あるいは統計学一般においては、 $\bar{\mathbf{Y}}$ の分布を知ることが重要になる。 $\bar{\mathbf{Y}}$ の分布を、 $\bar{\mathbf{Y}}$ の標本分布という。

$\bar{\mathbf{Y}}$ の平均

$$E(\bar{\mathbf{Y}}) = \frac{1}{n} \sum_{i=1}^n E(\mathbf{Y}_i) = \mu_{\mathbf{Y}}. \quad (20)$$

$\bar{\mathbf{Y}}$ の分散

$$var(\bar{\mathbf{Y}}) = var\left(\frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i\right) = \frac{1}{n^2} \sum_{i=1}^n var(\mathbf{Y}_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n cov(\mathbf{Y}_i, \mathbf{Y}_j) = \frac{1}{n} \Sigma_{\mathbf{Y}}. \quad (21)$$

ここで、i.i.d. の仮定により、 $cov(\mathbf{Y}_i, \mathbf{Y}_j) = 0$ となることを使っている。

標本平均の分布を求めるには、強い仮定を置く必要がある。特に \mathbf{Y}_i の分布を仮定しないと求めることができない。しかし、かなり緩い仮定のもとで、分布の近似を求めることはできる。計量経済学では、通常、大標本理論を用いた近似を行う。大標本理論は、

1. 比較的緩い条件のもとで成り立つ。

2. したがって、一般的である。つまり、同じ結果を多くの違った状況のもとで得ることができる。
3. しかし、標本数が少ない場合には、近似の精度は高くないかもしれない。

大数の法則 (LLN) $\bar{\mathbf{Y}}$ は、 n が大きいときに、 $\mu_{\mathbf{Y}}$ にとても近くなることが証明できる。

以下を仮定する。

1. $\mathbf{Y}_i, i = 1, \dots, n$ は i.i.d..
2. \mathbf{Y}_i の分散は、有限。 $(\|\Sigma_{\mathbf{Y}}\| < \infty)$

すると、任意の $\epsilon > 0$ について、 $n \rightarrow \infty$ のとき、

$$\Pr(\|\bar{\mathbf{Y}} - \mu_{\mathbf{Y}}\| > \epsilon) \rightarrow 0 \quad (22)$$

である。なお $\|\cdot\|$ はユークリッド距離である ($\|A\| = \sqrt{\text{tr}(A'A)}$)。

$\bar{\mathbf{Y}} \xrightarrow{p} \mu_{\mathbf{Y}}$ と表記し、 $\bar{\mathbf{Y}}$ は $\mu_{\mathbf{Y}}$ に確率収束するという。

中心極限定理 (CLT) $\bar{\mathbf{Y}}$ の分布は、 n が大きいとき、正規分布で近似できる。

以下を仮定する。

1. $\mathbf{Y}_i, i = 1, \dots, n$ は i.i.d..
2. \mathbf{Y}_i の分散は、有界であり、十分に 0 から離れている。 $(0 < \lambda_{\min}(\Sigma_{\mathbf{Y}})$ であり、 $\|\Sigma\| < \infty$ である。)

すると、 $\sqrt{n}(\bar{\mathbf{Y}} - \mu_{\mathbf{Y}})$ の分布関数がすべての点で、正規分布の分布関数に収束する。

この性質を $\sqrt{n}(\bar{\mathbf{Y}} - \mu_{\mathbf{Y}}) \xrightarrow{d} N(0, \Sigma_{\mathbf{Y}})$ と表記する。なおこのとき、 $\sqrt{n}(\bar{\mathbf{Y}} - \mu_{\mathbf{Y}})$ は $N(0, \Sigma_{\mathbf{Y}})$ に分布収束するという。また、 $\bar{\mathbf{Y}}$ は漸近正規である、ともいう。

ただし、 $\mathbf{Y}_i \sim N(\mu_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$ であるなら、この結果は、近似ではなく厳密に成り立つ。

それでは、どの程度の標本数があれば、 $\bar{\mathbf{Y}}$ を正規分布で近似しても問題ないのであろうか？この問いには、状況によるとしか答えようがない。ここで考えている標本平均の場合、 $n = 100$ であれば、十分と思われるが、もっと複雑な統計量の場合は、より多くの観測値が必要になる可能性が高い。

Slutsky の補題 $\mathbf{X}_n, \mathbf{X}, \mathbf{Y}_n$ を、確率変数からなるベクトルあるいは行列とする。ここで、 $n \rightarrow \infty$ のとき、 $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ 、ある固定されたベクトルと行列 c にあって、 $\mathbf{Y}_n \xrightarrow{p} c$ と仮定する。

1. $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{d} \mathbf{X} + c;$
2. $\mathbf{Y}_n \mathbf{X}_n \xrightarrow{d} c \mathbf{X};$
3. $\mathbf{Y}_n^{-1} \mathbf{X}_n \xrightarrow{d} c^{-1} \mathbf{X}$ 。 $(\mathbf{Y}_n^{-1}$ と c^{-1} の存在を仮定している。)

Slutsky の補題は、統計量の漸近分布を求める際に必ずと言ってほいほど使用される重要な定理である。