計量経済学のための行列入門

別所俊一郎

1　はじめに

計量経済学を勉強するときにはいろいろな場面で行列が登場します．行列を回避していても統計パッケージのプログラムが書ければ分析ができないことはないのですが，それでは心許ないので，最低限の知識は身につけておきましょう．ここで想定する読者は，行列を高校時分に習ったが楽しい学部生活のうちに忘れてしまった，あるいはベクトルは見たことがあるかもしれないが行列は聞いたことがないという文系学部の卒業生です[1]．

2　回帰分析のどこに行列が必要か

いまのところ，計量経済分析でもっとも頻繁に利用されている手法の一つが回帰分析（regression analysis）です．標準的な回帰分析では，ある変数 X_1 が他の変数 Y に与える因果効果[2]を線形に仮定（近似）し[3]，他の変数 $X_2, ..., X_k$ が Y に与える効果も同じように線形と仮定したうえで，これらの変数の関係を以下の式のように仮定します．

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \tag{2.0.1}$$

最後に足されている u は誤差項（**error term**）と呼ばれ，検討の対象となっている変数 $X_1, ..., X_k$ 以外の要因でも Y が変動することを表現しています．係数 $\beta_0, ..., \beta_k$ はそれぞれ定数で，各変数が与える効果の大きさを示しています．じっさい，

$$\frac{\partial Y}{\partial X_i} = \beta_i \tag{2.0.2}$$

が成り立ちますね．この $\beta_0, ..., \beta_k$ は定数であって，$Y, X_1, ..., X_k, u$ をすべて確率変数とみなしたときの同時分布の特性値のひとつと考えることができます[4]．ですが，この $\beta_0, ..., \beta_k$ は私たちが直接に観察することのできる定数ではありません．ちょうど，母集団の母平均が直接観測できなかったのと同じです．私たちが実際に観察することができるのは，$(Y, X_1, ..., X_k)$

---

[1] これを書いている本人がこの条件に合致しているという止むを得ない理由により，数学的にあいまいだったり，間違っていたりすることがあるかもしれないので先に謝っておきます．ごめんなさい．

[2]「因果関係」を想定して分析を進めることが多いですが，回帰分析が「因果」関係を証明するかというとそうでもありません．

[3] 対数線形や 2 次式など，線形に近い関係も扱うことができます．これはまたあとの話．

[4] $X_1, ..., X_k$ は定数とみなして話を進めても最初のうちはそれほど話が変わったりはしません．

の実現値の組み合わせだけです．誤差項 u も観測することはできません．このような状況で $\beta_0, ..., \beta_k$ を統計的に推測する，というのが回帰分析のやりたいことです．

さて，いま手許にサンプルサイズ n のデータセットがあるとしましょう．つまり，$(Y, X_1, ..., X_k)$ の実現値の組み合わせ（観測値 **observation**）が n 個あるとしましょう．このそれぞれの観測値が（2.0.1）の関係を満たしているのですから，それらすべてを書いてみると，適当に添え字を割り当てると

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \\ Y_2 = \beta_0 + \beta_1 X_{22} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n \end{cases} \tag{2.0.3}$$

という連立方程式となります．もし，$u_1, ..., u_n$ がすべて観測できていたら，この連立 1 次方程式は，未知数が k 個，式の数が n 個なので，$k = n$ のときにはふつうはちゃんと解くことができます[5]．しかし，ここでは $\beta_0, ..., \beta_k, u_1, ..., u_n$ が観測されていない未知数ですから，未知数の数が式の数よりも明らかに多く，ふつうの連立 1 次方程式を解くように解くことはできません．そのため，いくつかの前提条件を置いて適当な $\beta_0, ..., \beta_k$ を推定（estimate）することになります．上に述べたように，$\beta_0, ..., \beta_k$ は $Y, X_1, ..., X_k, u$ をすべて確率変数とみなしたときの同時分布の特性値のひとつですから，サンプルが変われば推定される $\beta_0, ..., \beta_k$ の値も変わります．つまり，推定される $\beta_0, ..., \beta_k$ もまた確率変数であり，その平均，分散，共分散を考える必要があります．

さて，この推定の手続きのために連立方程式をいじくらなければならないわけですが，(2.0.3）の形のままでは扱いにくいので，簡単に表現しなおしたほうがなにかと便利です．たとえば（2.0.3）は

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{U} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{2.0.4}$$

と定義してやると，

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U} \tag{2.0.5}$$

と書き直すことができます．このように書き直すための，あるいは書き直したあとのさまざまな変形に必要なお約束を一通り知っておく，というのがここでの目的です．

[5]一般に誤差項 u_i は「小さい」値を取ると想定されるので，$k = n$ のときは，$u_i = 0$ と仮定することで β_j を求めることができます．さらに，$k > n$ なら，一般にはこの連立方程式は解を持つとは限りません．つまり，知りたい係数の数よりも観測値の数のほうが小さければ回帰分析を行うことはできません．

3　ベクトル

3.1　大きさと向きを持った量

ベクトル（vector）とは「大きさと向きを持った量」であり，向きを持たないただの数のことをスカラー（scalar）と呼ぶのだ，ということは高校数学で習ったかもしれません．ベクトルというのはある点からある点を結んだ矢印だ，ということですね．このような考え方では幾何的なイメージを持ちやすいのですが，計量経済学で使うベクトルはしばしば次元が大きいので想像を超えた[6]ベクトルが登場します．そこで，ベクトルとは「いくつかの数の組」だというふうにも考えておいたほうがよいでしょう．どちらの考え方にも慣れれば理解が進むことには違いありません．

ベクトルは大きさと向きを持ちますが，この「大きさ」のことをベクトルの絶対値あるいは長さといいます．ベクトル $\mathbf{a}$ の絶対値を $|\mathbf{a}|$ と表記します．矢印としてのベクトルを想像すると，矢印の長さがベクトルの絶対値になります．ベクトルの長さはスカラーです．また，向きを変えずに大きさだけを伸ばしたり縮めたりしたものをそのベクトルの定数倍（スカラー倍）と呼びます．ベクトル $\mathbf{a}$ とスカラー k に対してスカラー倍は $k\mathbf{a}$ と書きます．$k < 0$ であれば，矢印は 180 度逆の方向を向きます．

矢印としてのベクトルを考えるとき，ベクトルの始点はどこにおいてもかまいません．長さがゼロでない 2 つのベクトル $\mathbf{a}, \mathbf{b}$ を考え，ベクトル $\mathbf{b}$ の始点をベクトル $\mathbf{a}$ の終点におくと，ベクトル $\mathbf{a}$ の始点からベクトル $\mathbf{b}$ の終点へ至る新しいベクトルを考えることができます．この新しいベクトルを 2 つのベクトルの和と呼び，$\mathbf{a} + \mathbf{b}$ で表現します．同じことですが，2 つのベクトルの始点をそろえ，2 つのベクトルを 2 辺とする平行四辺形を考え，そろえられた始点からの対角線に相当するベクトルもベクトルの和になります．この矢印のイメージから，ベクトルの和についてのいくつかの性質が導かれます．第 1 は，

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} \tag{3.1.1}$$

という性質で，交換則と呼ばれます．第 2 は結合則であり

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}) = \mathbf{a} + \mathbf{b} + \mathbf{c} \tag{3.1.2}$$

が成り立ちます．第 3 は分配則です．実数 k, ℓ に対して，

$$(k + \ell)\mathbf{a} = k\mathbf{a} + \ell\mathbf{a} \tag{3.1.3}$$

$$k(\mathbf{a} + \mathbf{b}) = k\mathbf{a} + k\mathbf{b} \tag{3.1.4}$$

[6] 想像を超えた，というのは，とても難しくて，ということではなくて，3 次元の世界に住んでいるわれわれにはイメージしにくいよね，といったくらいの意味です．しかし，われわれの世界も 10 次元だか 11 次元だか，という説もあるそうですが．

となります．スカラー倍したベクトルを足し合わせることを線形結合といいます．また，三角不等式と呼ばれる次の関係が成立します．

$$|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}|+|\mathbf{b}| \tag{3.1.5}$$

これは，3 角形の 2 辺の長さの和は他の 1 辺よりも長い，という当たり前の関係式をいいます．

長さがゼロでない 2 つのベクトルについて，一方が他方のスカラー倍になっているとき，この 2 つのベクトルは平行である，といいます．また，2 つのベクトルのなす角が 90 度であるとき，2 つのベクトルは直交している（**orthogonal**）といいます．直交という概念は計量経済学では極めて重要な概念となっていますが，平面上のベクトルであればベクトルのなす角はわかりやすいものの，そうでない場合には幾何的には想像しにくいかもしれません．

さて，実数の 2 次式 $(a+b)^2 = a^2+b^2+2ab$ という関係式に倣って，

$$|\mathbf{a}+\mathbf{b}|^2 = |\mathbf{a}|^2+|\mathbf{b}|^2+2\mathbf{a}\cdot\mathbf{b} \tag{3.1.6}$$

が成り立つようにベクトルの積を定義します．このような積のことをとくに内積（inner product）と呼びます．$\mathbf{a}=\mathbf{b}$ とおくと，

$$|\mathbf{a}+\mathbf{a}|^2 = |\mathbf{a}|^2+|\mathbf{a}|^2+2\mathbf{a}\cdot\mathbf{a}, \quad |\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a} \tag{3.1.7}$$

となります．また，余弦定理から，

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta \tag{3.1.8}$$

が成り立ちます[7]．内積の定義としてはこちらのほうが有名 (?) かもしれません．$\theta = 0$ のとき $\cos\theta = 0$ ですから，

$$\text{ベクトル } \mathbf{a}，\mathbf{b} \text{ が直交している} \iff \mathbf{a}\cdot\mathbf{b} = 0 \tag{3.1.9}$$

です．三角関数が思い出せない人も，この関係は覚えておきましょう．

練習問題： $\mathbf{y}$ をゼロでない任意のベクトルとし，$\mathbf{e}$ を任意の単位ベクトル（$|\mathbf{e}|=1$ を満たすベクトル）とします．いま，

$$\mathbf{y_p} = (\mathbf{e}\cdot\mathbf{y})\mathbf{e}, \quad \mathbf{y_o} = \mathbf{y}-\mathbf{y_p} \tag{3.1.10}$$

と定義します．このとき，$\mathbf{e}$ と $\mathbf{y_o}$ が直交していることを証明しなさい．このような分解をしたときの $\mathbf{y_p}$ を，$\mathbf{y}$ のベクトル $\mathbf{e}$ 方向への射影（projection）と呼びます．

[7]「まさかここまできて三角関数が…」と思ったかもしれませんが，今後ほとんど出てこないので心配しなくても大丈夫です．でも，時系列解析をやるとまた出てくるらしいです．

ここで線形独立性（linear independence）という概念を示しておきましょう．線形独立性の定義： k 個のベクトル $\mathbf{x_1}, ..., \mathbf{x_k}$ が線形独立であるとは，

$$\beta_1 \mathbf{x_1} + \cdots + \beta_k \mathbf{x_k} = \mathbf{0} \tag{3.1.11}$$

を満たすような実数の組 $(\beta_1, \cdots, \beta_k)$ が $\beta_1 = \cdots = \beta_k = 0$ しか存在しないことをいう．

k 個のベクトルが線形独立でないとき，線形従属（1 次従属，linear dependent）である，といいます．（3.1.11）が $\beta_1 = \cdots = \beta_k = 0$ 以外の組み合わせで成り立っている，つまり線形従属の関係にあるとき，いずれかのベクトルを右辺に移項してやれば，移項されたベクトルは他のベクトルの線形結合で表現されていることになります．逆に，あるベクトルが他のベクトルの線形結合で表現されているとき，右辺がゼロベクトルになるように移項してやれば，（3.1.11）を $\beta_1 = \cdots = \beta_k = 0$ 以外の組み合わせで作ることができることになります．したがって，k 個のベクトル $\mathbf{x_1}, ..., \mathbf{x_k}$ が線形従属であることは，このうちの少なくとも 1 つのベクトルが他の $k-1$ 個のベクトルの線形結合で表現できることと同値です．もし，あるベクトルが他のベクトルのスカラー倍になっているとき，その 2 つのベクトルは 1 次従属の関係にあります．

3.2 数の組み合わせ

ベクトルは矢印のイメージで考えることもできますが，始点を原点にとって終点の座標を示せばそのベクトルを表現できるように，数の組み合わせと考えることもできます．以下では，数の組み合わせとしてのベクトルや，ベクトルをさらに組み合わせた行列を考えていきますが，矢印としてのベクトルで考えることも理解の助けになるかもしれません．

まず，いくつかの用語を導入しましょう．n 個の数 $b_1, b_2, ..., b_n$ を縦に並べたベクトル $\mathbf{b}$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{3.2.12}$$

を n 次元の列ベクトル（**column vector**）[8]，縦ベクトルと呼び，b_i はベクトル $\mathbf{b}$ の第 i 成分（要素，element）と呼びます．また，n 個の数 $c_1, c_2, ..., c_n$ を縦に並べたベクトル $\mathbf{c}$

$$\mathbf{c} = \begin{bmatrix} c_1, & c_2, & \cdots, & c_n \end{bmatrix} \tag{3.2.13}$$

[8]いくつの要素が並んでいるかということです（ルパン 3 世とは関係ない）．

を n 次元の行ベクトル（**row vector**），横ベクトルと呼びます[9]．縦ベクトルの縦横をひっくりかえして横ベクトルにしたり，横ベクトルの縦横をひっくりかえして縦ベクトルにしたりすることを転置（**transpose**）と呼び，$\mathbf{b}'$ のように表現します[10]．つまり，

$$\mathbf{b}' = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}' = \begin{bmatrix} b_1, & b_2, & \cdots, & b_n \end{bmatrix} \tag{3.2.14}$$

となります．縦横をひっくりかえしてひっくりかえすと元に戻りますから，

$$(\mathbf{b}')' = \mathbf{b} \tag{3.2.15}$$

が成り立ちます．

さて，ベクトルの和やスカラー倍は，2 次元のばあいを考えると自然にわかるように，以下のように定義します．たとえば縦ベクトルのばあい，

$$k\mathbf{b} + \ell\mathbf{c} = k \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \ell \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \equiv \begin{bmatrix} kb_1 + \ell c_1 \\ kb_2 + \ell c_2 \\ \vdots \\ kb_n + \ell c_n \end{bmatrix} \tag{3.2.16}$$

次元が違うベクトルは足したりスカラー倍したりできません．このように定義すると，前に述べたベクトルの和の交換則，結合則，分配則が成り立ちます．ベクトルをスカラー倍して和をとることを線形結合と呼びましたが，k 個の縦ベクトル $\mathbf{x_1}, \cdots, \mathbf{x_k}$ を，定数 $\beta_1, \cdots, \beta_k$ を用いて線形結合し，新しいベクトル $\mathbf{y}$ を作る，つまり，

$$\mathbf{y} = \beta_1\mathbf{x_1} + \cdots + \beta_k\mathbf{x_k} \tag{3.2.17}$$

---

[9] 横ベクトルの成分の間にはカンマを打たないのですが，混乱しそうなときには書いたほうが便利だったりします．

[10] 理数系では $\mathbf{b}^t$ とか $\mathbf{b}^T$ とか書く場合もありますが，計量経済学の文脈では $\mathbf{b}'$ と書いて「b プライム」とか読むことも多いです．

を成分表示すると，

$$
\begin{aligned}
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} &= \beta_1 \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} + \cdots + \beta_k \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kn} \end{bmatrix} \\
&= \begin{bmatrix} \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_k x_{k1} \\ \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_k x_{k2} \\ \vdots \\ \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{kn} \end{bmatrix}
\end{aligned} \tag{3.2.18}
$$

となります．このように書くと，(2.0.3) は，各データからなるベクトルを適当に線形結合して，被説明変数のデータベクトルに「近づける」というふうに解釈することができます．

次に，内積（inner product）を定義しましょう．矢印としてのベクトルの内積を定義するときには長さを使ったり余弦を使ったりしましたが，ここでは以下のように定義します．

内積の定義： n 次元の横ベクトル $\mathbf{b}$ と n 次元の縦ベクトル $\mathbf{c}$ を考える．この 2 つのベクトルを この順番で 並べたものについて内積を以下のように定義する．

$$
\mathbf{bc} \equiv \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = b_1 c_1 + b_2 c_2 + \cdots + b_n c_n = \sum_{i=1}^{n} b_i c_i \tag{3.2.19}
$$

内積がゼロになる 2 つのベクトルは，互いに直交している（**orthogonal**）という．

この定義では，同じ次元のベクトルを，横・縦の順番に並べた場合にだけ内積を定義することができます．次元が同じであっても，縦・横の順に並べると大きな行列ができあがってしまうので注意しましょう．内積がゼロになることと，ベクトルが直交していることは，矢印のイメージからは想像しにくいとおもうので，ここでは「そんなものだ」と思って覚えてしまいましょう．でもじっさい，2 次元のケースでは直交しているベクトルの内積がゼロになることが簡単に確認できます．たとえば，2 つのベクトル $(0,1),(1,0)$ を考えてみましょう．明らかにこの 2 つのベクトルのなす角は 90 度ですが，内積もちゃんとゼロになっています．

ベクトルの長さ（絶対値）も，2 次元のベクトルのときの定義をそのまま拡張して考えます．2 次元のベクトルの長さは，各成分を 2 乗したものの和の 2 乗根でした（三平方の定理ですね）から，n 次元のベクトル $\mathbf{b}$ の長さは以下のように定義します．

$$
|\mathbf{b}| = \sqrt{b_1^2 + b_2^2 + \cdots + b_n^2} = \sqrt{\mathbf{bb}} \tag{3.2.20}
$$

4 行列

4.1 基礎的な演算

数を縦に n 個，横に m 個ならべて 1 セットとしたもののことを n 行 m 列の（$n \times m$ の）行列（**matrix**）と呼びます[11]．「$n \times m$」のような行列の次元を表すときには，縦に並んでいる個数を最初に，横に並んでいる個数を次に書きます．このお約束を守っていると，行列の積の定義のときなどにちょっと便利です．上から i 番目に並んでいる横ベクトルのことを第 i 行，左から j 番目に並んでいる縦ベクトルのことを第 j 列と呼びます．行列の縦横はよく混乱するので気をつけましょう．

n 行 k 列の行列はまた，n 次元の縦ベクトルが横に k 個ならんだものと考えることもできます．n 次元の縦ベクトルを $\mathbf{x_i}$ で表現すると，$n \times k$ の行列 $\mathbf{X}$ は，

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_k} \end{bmatrix} \tag{4.1.1}$$

と書くことができます．このように考えると，行列の和とスカラー倍も次のように自然に定義されます．いま，$n \times k$ の行列 $\mathbf{X}$，$\mathbf{Y}$ の (i, j) 要素を $x_{i,j}, y_{i,j}$ とし，k を定数として，

$$\begin{aligned}
\mathbf{X} + \mathbf{Y} &= \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix} + \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,k} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,k} \end{bmatrix} \\
&\equiv \begin{bmatrix} x_{1,1} + y_{1,1} & x_{1,2} + y_{1,2} & \cdots & x_{1,k} + y_{1,k} \\ x_{2,1} + y_{2,1} & x_{2,2} + y_{2,2} & \cdots & x_{2,k} + y_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} + y_{n,1} & x_{n,2} + y_{n,2} & \cdots & x_{n,k} + y_{n,k} \end{bmatrix} \qquad (4.1.2) \\
k\mathbf{X} &\equiv \begin{bmatrix} kx_{1,1} & kx_{1,2} & \cdots & kx_{1,k} \\ kx_{2,1} & kx_{2,2} & \cdots & kx_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ kx_{n,1} & kx_{n,2} & \cdots & kx_{n,k} \end{bmatrix} \qquad (4.1.3)
\end{aligned}$$

定義から明らかなように，次元の異なる行列は足すことができません．

[11] matrix は「母体，鋳型」などの意味も持っています．Keanu Reeves 主演の映画は行列とはなんの関係もありませんね．

$n \times k$ の行列 $\mathbf{X}$ の縦と横をひっくり返したものを，ベクトルと同じように，転置行列（**transpose**）と呼びます．上記の行列を例にとると，

$$\mathbf{X}' = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,k} & x_{1,k} & \cdots & x_{k,n} \end{bmatrix} = \begin{bmatrix} \mathbf{x_1'} \\ \mathbf{x_2'} \\ \vdots \\ \mathbf{x_k'} \end{bmatrix} \tag{4.1.4}$$

となります．

練習問題： $n \times k$ の行列 $\mathbf{X}$，$\mathbf{Y}$，$\mathbf{Z}$ について，以下が成り立つことを確認しなさい．

$$\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X} \tag{4.1.5}$$

$$(\mathbf{X} + \mathbf{Y}) + \mathbf{Z} = \mathbf{X} + (\mathbf{Y} + \mathbf{Z}) \tag{4.1.6}$$

$$(\mathbf{X} + \mathbf{Y})' = \mathbf{X}' + \mathbf{Y}' \tag{4.1.7}$$

次に，行列の積を定義します[12]．ベクトルの内積は n 次元の横ベクトルと n 次元の縦ベクトルをこの順番で並べたものに対して定義されました．また，行列は縦ベクトルを横に並べたもの，あるいは横ベクトルを縦に積み重ねたものでもあります．そこでたとえば，ベクトルの内積の左側の横ベクトルをそのままにして，右から縦ベクトルを順々にかけていき，その内積をかけた順番に左から並べて新たな横ベクトルを定義することは自然な拡張のように思えます．すなわち，n 次元の横ベクトル $\mathbf{a}$ に，n 次元の縦ベクトル $\mathbf{b_i}$ を m 個並べた行列 $\mathbf{B}$ を右から掛けて，その積 m 個を左から順に並べた横ベクトルを考えます．

$$\mathbf{aB} = \mathbf{a} \begin{bmatrix} \mathbf{b_1} & \mathbf{b_2} & \cdots & \mathbf{b_m} \end{bmatrix} = \begin{bmatrix} \mathbf{ab_1} & \mathbf{ab_2} & \cdots & \mathbf{ab_m} \end{bmatrix} \tag{4.1.8}$$

さらに，左にある横ベクトルが縦に ℓ 個積み重なって行列 $\mathbf{A}$ ができているとすれば，内積を順番に上から積み重ねるのが自然でしょう．つまり，

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_\ell} \end{bmatrix} \begin{bmatrix} \mathbf{b_1} & \mathbf{b_2} & \cdots & \mathbf{b_m} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a_1b_1} & \mathbf{a_1b_2} & \cdots & \mathbf{a_1b_m} \\ \mathbf{a_2b_1} & \mathbf{a_2b_2} & \cdots & \mathbf{a_2b_m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_\ell b_1} & \mathbf{a_\ell b_2} & \cdots & \mathbf{a_\ell b_m} \end{bmatrix} \end{aligned} \tag{4.1.9}$$

[12] 掛けられる行列のそれぞれの要素に，掛ける行列を掛けるという，積となる行列の次元がやたらと大きくなる掛け算も考えることがあって，そういうのをクロネッカー積といいますが，さしあたって使わないので気にしないことにしましょう．

これが行列の積として定義されます．すなわち，$\ell \times n$ の行列 $\mathbf{A}$ と $n \times m$ の行列 $\mathbf{B}$ の積 $\mathbf{AB}$ の (i, j) 要素は，

$$(\mathbf{AB})_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j} \tag{4.1.10}$$

となります．定義からわかるように，行列の積を求めるには，2 つの行列の次元がうまく一致する（confortable）必要があります．たとえば，$m \times n$ の行列 $\mathbf{A}$ と $n \times m$ の行列 $\mathbf{B}$ を考えると，

$$\underbrace{\mathbf{A}}_{m\times n} \times \underbrace{\mathbf{B}}_{n\times m} = \underbrace{\mathbf{AB}}_{m\times m}, \quad \underbrace{\mathbf{B}}_{n\times m} \times \underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{BA}}_{n\times n} \tag{4.1.11}$$

となります．したがって一般に，

$$\mathbf{AB} \neq \mathbf{BA} \tag{4.1.12}$$

です．行列の次元を $m \times n$ というふうに書くことにすれば，隣り合った行列の積が定義できるかどうかは，隣り合った列・行の次元が一致しているかどうかで判定でき，積の行列の次元はその「外側」によって表現されます．つまり，$\ell \times n$ の行列に右から $n \times m$ の行列をかけた積は $\ell \times m$ の行列になります．

練習問題： 行列の積が定義できるとき，以下が成り立つことを確認しなさい．

$$(\mathbf{XY})\mathbf{Z} = \mathbf{X}(\mathbf{YZ}) \tag{4.1.13}$$

$$(\mathbf{X} + \mathbf{Y})\mathbf{Z} = \mathbf{XZ} + \mathbf{YZ} \tag{4.1.14}$$

$$(\mathbf{XY})' = \mathbf{Y}'\mathbf{X}' \tag{4.1.15}$$

行列の積は，掛け合わせる順番によって異なる値を取り，定義ができないことすらありますから，縦と横の次元の等しい正方行列$\mathbf{A}$，$\mathbf{B}$ については

$$\begin{aligned}(\mathbf{A} + \mathbf{B})^2 = (\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B}) &= \mathbf{A}^2 + \mathbf{BA} + \mathbf{AB} + \mathbf{B}^2 \\ &\neq \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2\end{aligned} \tag{4.1.16}$$

となります．また，n 次の縦ベクトルに右から m 次の横ベクトルを掛けると，その積は $n \times m$ の行列になります．いま，絶対値が 1 の n 次の縦ベクトル $\mathbf{e}$ について，$n \times n$ の行列 $\mathbf{P_e}$ を

$$\mathbf{P_e} \equiv \mathbf{ee}' \tag{4.1.17}$$

と定義するとき，この行列 $\mathbf{P_e}$ を $\mathbf{e}$ への射影行列（projection matrix）と呼びます．じっさい，n 次の任意の縦ベクトル $\mathbf{x}$ に対して，

$$\mathbf{P_e x} = \mathbf{ee}'\mathbf{x} = (\mathbf{e}'\mathbf{x})\mathbf{e} \tag{4.1.18}$$

となり，$\mathbf{e}$ のスカラー倍になります．

4.2　名前のある行列

いくつかの行列は，名前を持っています．ここでは，計量経済学で出てきそうないくつかについてご紹介．

正方行列 square matrix 先ほど出てきましたが，縦と横の次元が等しい行列のことです．

対称行列 symmetric matrix 正方行列のうち，第 (i,j) 成分と第 (j,i) 成分の値が等しい行列を対称行列といいます．第 (i,i) 成分のことを対角成分といいます．k 個の確率変数を考え，その分散を対角成分に，非対角成分に共分散を置いた行列のことを共分散行列と呼びますが，共分散行列はその定義からあきらかに，対称行列です．

対角行列 diagonal matrix 正方行列のうち，非対角成分がすべてゼロである行列を対角行列といいます．つまり，

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & 0 & \cdots & 0 \\ 0 & x_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{k,k} \end{bmatrix} \tag{4.2.19}$$

のような行列です．k 個の互いに無相関な確率変数を考えると，その共分散行列は対角行列であり，対角成分には各確率変数の分散が入ります．

練習問題： $\mathbf{A}, \mathbf{B}$ が対角行列であるとき，$\mathbf{AB} = \mathbf{BA}$ が成り立つことを示しなさい[13]．

単位行列 identity matrix 対角行列のうち，対角成分がすべて 1 である行列をとくに単位行列といいます．つまり，

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{4.2.20}$$

が単位行列です．単位行列には $\mathbf{I}$ という記号をあてます[14]．

練習問題： 正方行列 $\mathbf{X}$ に対して $\mathbf{XI} = \mathbf{IX} = \mathbf{X}$ が成り立つことを示しなさい．

[13] 対角行列でなくても，$\mathbf{AB} = \mathbf{BA}$ が成り立つような正方行列 $\mathbf{A}, \mathbf{B}$ はあります．そのような行列は可換 commutable といいます．

[14] 次元を明記して $\mathbf{I_n}$ と書くこともあります．

直交行列 **orthogonal matrix** 正方行列のうち，その転置との積が単位行列となる行列を直交行列といいます．つまり，$\mathbf{X}$ が直交行列であれば，

$$\mathbf{X}'\mathbf{X} = \mathbf{I}, \tag{4.2.21}$$

が成り立ちます．また，任意のベクトル $\mathbf{a}, \mathbf{b}$ に対して，以下が成り立ちます．

$$\mathbf{a} \cdot \mathbf{b} = (\mathbf{X}\mathbf{a}) \cdot (\mathbf{X}\mathbf{b}) \tag{4.2.22}$$

4.3 逆行列

逆行列の定義：正方行列 $\mathbf{X}$ に対して，

$$\mathbf{X}\mathbf{Y} = \mathbf{Y}\mathbf{X} = \mathbf{I} \tag{4.3.23}$$

となる行列 $\mathbf{Y}$ が存在するとき，$\mathbf{Y}$ を $\mathbf{X}$ の逆行列（inverse）と呼び，$\mathbf{Y} = \mathbf{X}^{-1}$ と書く．この定義から明らかなように，逆行列を考えることができるのは正方行列に対してのみであり，また，正方行列であっても逆行列が存在しない行列があります．逆行列が存在しない正方行列のことを特異行列（singular matrix）と呼ぶので，逆行列が存在する行列は非特異（nonsingular）とか可逆（invertible）とか正則（regular）とか呼ばれます．

練習問題： 正方行列 $\mathbf{X}, \mathbf{Y}$ が逆行列をもつとき，定数 k にたいして以下が成り立つことを示しなさい．

$$(\mathbf{X}^{-1})^{-1} = \mathbf{X} \tag{4.3.24}$$
$$(k\mathbf{X})^{-1} = k^{-1}\mathbf{X}^{-1} \tag{4.3.25}$$
$$(\mathbf{X}\mathbf{Y})^{-1} = \mathbf{X}^{-1}\mathbf{Y}^{-1} \tag{4.3.26}$$

逆行列を左から掛けることは，直感的にはスカラーでの割り算に対応します．たとえば，$\mathbf{x}$, $\mathbf{y}$, $\mathbf{u}$ をベクトルとし，$\mathbf{A}$, $\mathbf{B}$ を正則行列として

$$\mathbf{A}\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{u} \tag{4.3.27}$$

という関係が成り立っているとすれば，左から逆行列を掛けると

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{y} = \mathbf{A}^{-1}(\mathbf{B}\mathbf{x} + \mathbf{u}) \quad \Rightarrow \quad \mathbf{y} = \mathbf{A}^{-1}(\mathbf{B}\mathbf{x} + \mathbf{u}) \tag{4.3.28}$$

となります．

さて，逆行列があるかないかについては以下のような性質があります．難しいので証明は省略．

逆行列の存在：正方行列 $\mathbf{X}$ を縦ベクトルを横に並べたもの（横ベクトルを縦に積んだもの）とみなすとき，縦ベクトルが 1 次従属であることと，逆行列が存在しないことは同値であり，このとき $\mathbf{X}$ の行列式（determinant）の値はゼロ．

行列式（**determinant**）というのも線形代数では非常に重要な概念なのですが今回は省略（すいません）．行列をベクトルを並べたものとみなすときに，ベクトルが 1 次従属であることと逆行列が存在しないことが関係ありそうなのは次のような例からわかるでしょう．いま，k 個の変数 $(x_1, \cdots, x_k)$ からなる連立 1 次方程式

$$\begin{cases} b_{11}x_1 + b_{21}x_2 + \cdots + b_{k1}x_k = b_{01} \\ b_{12}x_1 + b_{22}x_2 + \cdots + b_{k2}x_k = b_{02} \\ \vdots \\ b_{1k}x_1 + b_{2k}x_2 + \cdots + b_{kk}x_k = b_{0k} \end{cases} \tag{4.3.29}$$

を考えます．未知数が k 個で式が k 個ありますから，「普通であれば」この連立 1 次方程式は解をもちます．しかしたとえば，いずれかの方程式が他の方程式を単に 2 倍したものだったりしたら，連立方程式の解は定まりません．この「解が定まらない」という状況が，逆行列が存在しないという状況に対応します．上記の連立方程式を行列表現すると，

$$\mathbf{Bx} = \mathbf{b_0} \tag{4.3.30}$$

と表現できますから，もし $\mathbf{B}^{-1}$ が存在すれば，両辺に左から $\mathbf{B}^{-1}$ を乗じることで

$$\mathbf{x} = \mathbf{B}^{-1}\mathbf{b_0} \tag{4.3.31}$$

として解を求めることができます．もし逆行列が存在しなければ，この演算が定義できなくなり，解が定まりません．

一般に，ベクトルを並べたものとしての行列に対して，1 次独立なベクトルの個数のことをその行列のランク（**rank**）といいます．$n \times m$ の行列 $\mathbf{X}$ について，そのランクが列の数 m に等しいとき，full column rank と呼び，m より小さいときにはランク落ちと言ったりします．$n \times m$ の行列 $\mathbf{X}$ が full column rank であるとき，$\mathbf{X'X}$ は逆行列を持ちます．また，正方行列について，行列を構成するベクトルが互いに 1 次独立であれば逆行列が存在するわけですから，逆行列の存在する行列のランクはその次元に等しくなります．

4.4　定値行列

n 個の変数 $x_1,...,x_n$ と $a_{i,j}$（ただし $a_{i,j}=a_{j,i}$ とする[15]）について，

$$\sum_{i,j=1}^{n} a_{i,j}x_ix_j \tag{4.4.32}$$

を $x_1,...,x_n$ の 2 次形式（quadratic form）と呼びます．これまでの行列とベクトルの表現を用いると，$x_1,...,x_n$ を要素とする n 次の縦ベクトル $\mathbf{x}$ と，$a_{i,j}$ を要素とする対称行列 $\mathbf{A}$ をもちいて，

$$\sum_{i,j=1}^{n} a_{i,j}x_ix_j = \mathbf{x}'\mathbf{A}\mathbf{x} \tag{4.4.33}$$

と表現できます．ここで，対称行列 $\mathbf{A}$ に注目しましょう．任意の n 次ベクトル $\mathbf{x}$ に対して 2 次形式 $\mathbf{x}'\mathbf{A}\mathbf{x}$ がつねに正の値をとる，すなわち $\mathbf{x}'\mathbf{A}\mathbf{x}>0$ が任意の $\mathbf{x}$ に対して成り立つような行列 $\mathbf{A}$ を正定値（**positive definite**）行列といいます[16]．同様に，$\mathbf{x}'\mathbf{A}\mathbf{x}\geq 0$ が任意の $\mathbf{x}$ に対して成り立つような行列 $\mathbf{A}$ を半正定値（**positive semidefinite**）行列といいます．正定値行列はかならず逆行列を持ちます．

4.5　固有値，固有ベクトル

n 次の縦ベクトル $\mathbf{x}$ をもってきて，それに左から $n\times n$ の行列 $\mathbf{A}$ を掛けるとします．左から掛けているので，積として出てくるのはやっぱり n 次の縦ベクトルになり，できあがったベクトルともとの $\mathbf{x}$ は向きも大きさも違うベクトルになるのが普通でしょう．でも，$\mathbf{x}$ と $\mathbf{A}$ をうまいこと持ってくると，積として出てくるベクトルが元のベクトルのスカラー倍になることがあるかもしれません．このような状況を式で表現すると，

$$\mathbf{A}\mathbf{x}=\lambda\mathbf{x} \tag{4.5.34}$$

となります．もちろん，$\mathbf{x}$ がゼロベクトルであればこの関係は常に成り立ちますが，それはあんまりおもしろくありません．もし，行列 $\mathbf{A}$ に対してうまいこと選んだ $\mathbf{x}$ が (4.5.34) の関係式をみたすとき，このベクトル $\mathbf{x}$ を行列 $\mathbf{A}$ の固有ベクトル（**eigenvector**）といい，対応する λ を固有値（**eigenvalue**）といいます．また，上の関係式のことを固有方程式と呼びます．ベクトル $\mathbf{x}$ のスカラー倍した $k\mathbf{x}(k\neq 0)$ を考えると，行列 $\mathbf{A}$ に対して $k\lambda$ も固有方程式を満たしますから，通常はベクトルの絶対値が 1 であるような $\mathbf{x}$ を固有ベクトルといい，対応する λ を固有値と呼んでいます．

---

[15] この仮定はじつは必要ないのですけど，しばしばこのような状況をあつかいますので．

[16] つねに $\mathbf{x}'\mathbf{A}\mathbf{x}<0$ となる行列は負定値行列といいます．

固有方程式（4.5.34）の右辺を左辺に移項して，単位行列を使って書き直すと，

$$(\mathbf{A}-\lambda\mathbf{I_n})\mathbf{x}=\mathbf{0_n} \tag{4.5.35}$$

となります．右辺の $\mathbf{0_n}$ は n 次のゼロベクトルです．ここで，$(\mathbf{A}-\lambda\mathbf{I_n})$ が逆行列を持つとすると，その逆行列を両辺に左から掛けて，

$$(\mathbf{A}-\lambda\mathbf{I_n})^{-1}(\mathbf{A}-\lambda\mathbf{I_n})\mathbf{x}=(\mathbf{A}-\lambda\mathbf{I_n})^{-1}\mathbf{0_n} \quad \Rightarrow \mathbf{x}=\mathbf{0_n} \tag{4.5.36}$$

となってしまいますから，固有値 λ に対して $(\mathbf{A}-\lambda\mathbf{I_n})$ は逆行列を持たない，ということになります．ということは，行列式がゼロになるので，固有値は n 個存在することになります（証明省略．n 個のうち，同じ値をとるものがあるかもしれません）．

さて，さきほど述べた正定値行列と固有値・固有ベクトルのあいだには次のような関係が成り立つことが知られています．いま，$n\times n$ の行列 $\mathbf{A}$ が正方で対称な正定値行列であるとします．この正定値行列は実数の固有値を n 個もち，それに対応する固有ベクトルが n 個あります．固有値を $\lambda_1,...,\lambda_n$ とし，対応する固有ベクトルを $\mathbf{q}_1,...,\mathbf{q}_n$ とし，これらを使って次のような行列を定義します．

$$\mathbf{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}, \quad \mathbf{Q}=\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{bmatrix} \tag{4.5.37}$$

つまり，$\mathbf{\Lambda}$ は固有値を対角成分に並べた対角行列であり，$\mathbf{Q}$ は対応する順番で固有ベクトルを並べたもの，ということです．さて，これらの行列を使うと，もとの $n\times n$ の行列 $\mathbf{A}$ は

$$\mathbf{A}=\mathbf{Q\Lambda Q'} \tag{4.5.38}$$

と表現されます．また，固有ベクトルは互いに直交していることが知られていて，

$$\mathbf{Q'Q}=\mathbf{0} \tag{4.5.39}$$

となります．

「なんのこっちゃ」というところですが[17]，ここまでの固有値のはなしを 2 次元で考えてみましょう．2 次元のあるベクトルに 2×2 の行列を左から掛けると新しい 2 次元のベクトルを作ることができます．こういう手続きを 1 次変換と呼んで，むかしは文系の高校数学でも扱ったものです[18]．さて，1 次変換を表す行列を用意すると，2 次元の（平面上の）

[17]実際にこれらの性質を使ってなにかを証明しよう，ということはこの授業の範囲ではなかったりします．

[18]高校数学のカリキュラムの改正があって 1 次変換は姿を消し，かわりに極座標などが入ったそうです．たぶん．

ベクトルはあらぬ方向に向きを変えてしまうわけですが，ある特定のベクトルはその行列がおこなう1次変換に対して大きさは変わっても向きを変えません．このようなベクトルが固有ベクトルであり，大きさの変化分のことが固有値です．2次元ですから，固有ベクトルはふつうは2つあります．もし1次変換を示す行列が対称な正定値行列であれば[19]，2つの固有ベクトルが直交する，ということになります．

さてこの固有値，計量経済学のどこらへんに出てきそうでしょうか．対称行列というと，確率変数の共分散行列が対称行列です．推定の結果を評価するときには，推定量の分散や共分散をつかって仮説検定を行います．仮説検定というのは，ある決まった手続きに従って産出した統計量が，既知の確率分布に従うという性質を使います．既知の分布というと正規分布か χ^2 分布あたりがよく使われますが，いずれにしても共分散が0でなくて，それぞれが相関をもつような状況では評価が難しいでしょう．もし確率変数が互いに相関を持たなければ既知の分布に近づけることができるかもしれません．確率変数が互いに相関をもたないとき，共分散行列は対角行列になっています．上で述べたように，対称行列は固有ベクトルによって対角行列に転換することができますから，この性質を統計量の計算に用いることができるというわけです．

5　データ行列と標本平均，標本分散

ここまで扱ってきた行列の手法を用いて，データから標本平均や標本分散を求めてみましょう．いま手許にサンプルサイズ n のデータセットがあるとしましょう．つまり，$(X_1, ..., X_k)$ の実現値の組み合わせ（観測値 observation）が n 個あるとします．このデータを行列に格納します．横方向に1つの観測値のセットを並べて，それを縦に積み重ねて行列を定義しましょう．変数 X_1 の i 番目の実現値を x_{1i} とすると，

$$\mathbf{X} = \underbrace{\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}}_{n\times k} \tag{5.0.1}$$

[19] $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ とすると，$b = c$ かつ $a + b > 0, c + d > 0$.

という行列を定義できます．この行列をデータ行列と呼んだりします．ここで，各要素が 1 であるような n 次元の縦ベクトルを考え，太字で $\mathbf{1}$ と表記します[20]．

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{5.0.2}$$

5.1 標本平均

データ行列を使って，標本平均を表してみましょう．さきほど導入した縦ベクトルは $n \times 1$ なので，$n \times k$ のデータ行列 $\mathbf{X}$ を転置して左から掛けることができます．

$$\begin{aligned} \mathbf{X}'\mathbf{1} &= \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix}}_{k\times n} \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{n\times 1} \\ &= \underbrace{\begin{bmatrix} x_{11}+x_{12}+\cdots+x_{1n} \\ x_{21}+x_{22}+\cdots+x_{2n} \\ \vdots \\ x_{k1}+x_{k2}+\cdots+x_{kn} \end{bmatrix}}_{k\times 1} = \begin{bmatrix} \sum_{i=1}^n x_{1i} \\ \sum_{i=1}^n x_{2i} \\ \vdots \\ \sum_{i=1}^n x_{ki} \end{bmatrix} \end{aligned} \tag{5.1.3}$$

このようにして求められた $k \times 1$ の縦ベクトルの各成分は，各変数の総和となっています．総和をサンプルサイズ n で割れば標本平均が求められます．スカラーの割り算は各成分の割り算なので，

$$\frac{1}{n}\mathbf{X}'\mathbf{1} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{1i} \\ \sum_{i=1}^n x_{2i} \\ \vdots \\ \sum_{i=1}^n x_{ki} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_{1i} \\ \frac{1}{n}\sum_{i=1}^n x_{2i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n x_{ki} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_k \end{bmatrix} \tag{5.1.4}$$

標本平均ベクトルを $\bar{\mathbf{X}}$ と書くと，ここまでの計算は

$$\bar{\mathbf{X}} = \frac{1}{n}\mathbf{X}'\mathbf{1} \tag{5.1.5}$$

と書くことができます．

[20] 単位行列のときと同じく，次元を明記するために $\mathbf{1}_n$ と書くこともあります．

5.2　標本分散

次に，標本分散を考えてみましょう．標本分散は，各観測値からそれぞれの標本平均を引いたものの 2 乗の和を $n-1$ で割って求めるのでした．まず，各観測値からそれぞれの標本平均を引いたデータを作ってみましょう．標本平均はベクトル $\bar{\mathbf{X}}$ の各成分となっていますから，データ行列と引き算ができるように次のような掛け算を考えます．

$$\mathbf{1_n}\bar{\mathbf{X}}' = \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{n\times 1} \underbrace{\begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \end{bmatrix}}_{1\times k} = \underbrace{\begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \\ \bar{x}_1 & \bar{x}_2 & \vdots & \bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \end{bmatrix}}_{n\times k} \tag{5.2.6}$$

これをデータ行列から引いてやると各成分に偏差が入った行列を得ることができます．この平均偏差行列を $\mathbf{V}$ と表記すると，

$$\begin{aligned} \mathbf{V} &= \mathbf{X} - \mathbf{1_n}\bar{\mathbf{X}}' \\ &= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \\ \bar{x}_1 & \bar{x}_2 & \vdots & \bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \end{bmatrix} \\ &= \begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{k1}-\bar{x}_k \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{k2}-\bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{kn}-\bar{x}_k \end{bmatrix} \end{aligned} \tag{5.2.7}$$

偏差を得ることができたので，次にその 2 乗和を計算します．行列の積は横方向と縦方向に掛けて足し合わせたものですから，$\mathbf{V}$ の転置行列を左から乗じてやれば，対角成分に 2

乗和が登場します．計算してみると，

$$
\begin{aligned}
\underbrace{\mathbf{V}'\mathbf{V}}_{k\times k} &= \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_1 & \cdots & x_{1n}-\bar{x}_1 \\ x_{21}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{2n}-\bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1}-\bar{x}_k & x_{k2}-\bar{x}_k & \cdots & x_{kn}-\bar{x}_k \end{bmatrix} \begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{k1}-\bar{x}_k \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{k2}-\bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{kn}-\bar{x}_k \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 & \sum_{i=1}^n (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) & \cdots & \sum_{i=1}^n (x_{1i}-\bar{x}_1)(x_{ki}-\bar{x}_k) \\ \sum_{i=1}^n (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) & \sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 & \cdots & \sum_{i=1}^n (x_{2i}-\bar{x}_2)(x_{ki}-\bar{x}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n (x_{1i}-\bar{x}_1)(x_{ki}-\bar{x}_k) & \sum_{i=1}^n (x_{2i}-\bar{x}_2)(x_{ki}-\bar{x}_k) & \cdots & \sum_{i=1}^n (x_{ki}-\bar{x}_k)^2 \end{bmatrix}
\end{aligned}
$$

これを標本平均のときと同じように $n-1$ で割ってやれば標本分散が求まるのですが，この行列 $\mathbf{V}'\mathbf{V}$ の非対角成分をよくよくみると，標本共分散の定義になっていることがわかります．変数 X_1 の標本分散を s_1^2 と表し，変数 X_1 と X_2 の標本共分散を s_{12} と表すようにすれば，

$$
\frac{1}{n-1}\mathbf{V}'\mathbf{V} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1k} \\ s_{21} & s_2^2 & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{k1} & s_{k2} & \cdots & s_k^2 \end{bmatrix}
$$

共分散については $s_{ij} = s_{ji}$ が成り立ちますから，この行列は対称行列です．この行列を標本共分散行列[21]といい，$\mathbf{S}$ で表記すると，

$$
\mathbf{S} = \frac{1}{n-1}\mathbf{V}'\mathbf{V} \tag{5.2.8}
$$

5.3 ベクトルと確率変数

確率変数の線形和もまた確率変数になりますから，確率変数そのものをベクトルと考えることもできます[22]．ただし，確率変数が大きさや向きを持つというわけではなくて（大きさや向きが確率的に決まる，確率変数のベクトルは次に扱います），そのようなアナロジーで考えることができる，といったほどの意味です．

確率変数から定数を引いても分散の値は変わらないので，平均値（期待値）がゼロの確率変数に絞って考えてみましょう．雰囲気をつかむために，離散確率変数を考えます[23]．

[21]分散-共分散行列（variance-covariance matrix）と呼んだりもします．

[22]たとえば，小倉久直．1998．確率過程入門．森北出版，29 ページあたりからそのようなことが書いてあります．

[23]以下の説明はあくまでもイメージです．

確率変数がとりうる値を成分とするベクトルを作り，各成分に対応する確率は同じとします．たとえば，確率変数 X がふつうのサイコロを振ったときに出る目を表すとすれば，$X = (1, 2, 3, 4, 5, 6)'$ です．もしどれかの値に割り振られた確率が大きいなら，その値を何回か繰り返してベクトルを作ってみましょう．このベクトルの長さは (3.2.20) で与えられますから，

$$|\mathbf{X}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{5.3.9}$$

となりますが，この確率変数の平均はゼロとしていますから，この「長さ」は標準偏差に他なりません．言い換えると，「長さ」の 2 乗は分散を表します．

平均値がゼロの確率変数 Y について同じようなベクトルを考えて X と次元を揃えてやると，ベクトルの内積を (3.2.19) で考えることができます．

$$\mathbf{XY} \equiv \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i \tag{5.3.10}$$

2 つの確率変数の平均がともにゼロであることから，この内積は 2 つの確率変数の共分散を表します．内積は

$$X \text{ と } Y \text{ の内積} = X \text{ の長さ} \times Y \text{ の長さ} \times \cos\theta$$

でしたから，確率変数の概念で書いてみると

$$X \text{ と } Y \text{ の共分散} = X \text{ の標準偏差} \times Y \text{ の標準偏差} \times \cos\theta$$

となります．変形すると，

$$\cos\theta = \frac{X \text{ と } Y \text{ の共分散}}{X \text{ の標準偏差} \times Y \text{ の標準偏差}} = X \text{ と } Y \text{ の相関係数} \tag{5.3.11}$$

となり，2 つのベクトルのなす「角度」θ は相関係数に対応することになります．$\cos\theta$ は-1 から 1 までの値しかとりませんが，相関係数もまた絶対値で 1 以下であったことを思い出しましょう．2 つの確率変数の相関係数あるいは共分散がゼロのとき，その 2 つの確率変数はたがいに無相関だという言い方をしました．$\cos\theta = 0$ となるとき 2 つのベクトルは直交している（$\theta = 90$ 度）わけですから，確率変数の無相関はベクトルの直交に対応していることになります．計量経済学では，2 つの確率変数が相関を持たないという条件のことをしばしば直交条件（**orthogonality condition**）といいますが，その名前はここらへんに由来しています．

5.4 確率変数の平均，分散

各成分が確率変数であるようなベクトルや行列を考えることもあります．確率変数 $(X_1, ..., X_k)$ を成分とする確率変数ベクトル $\mathbf{x}$ を考えましょう．このとき，各成分の母平均を縦に並べた $k \times 1$ の実数ベクトルを μ と書き，

$$E[\mathbf{x}] = \mu = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_k] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} \tag{5.4.12}$$

と表します．行列やベクトルの期待値は，期待値の行列やベクトルで定義します．標本分散のときとおなじく，確率変数ベクトル $\mathbf{x}$ とその期待値ベクトル μ に対して偏差ベクトル $\mathbf{v}$ を $\mathbf{v} = \mathbf{x} - \mu$ と定義して，この転置積の期待値を考えると，

$$\begin{aligned} E[\mathbf{v}\mathbf{v}'] &= E\left[\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_k-\mu_k \end{bmatrix} \begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 & \cdots & x_k-\mu_k \end{bmatrix}\right] \\ &= \begin{bmatrix} E(x_1-\mu_1)^2 & E(x_1-\mu_1)(x_2-\mu_2) & \cdots & E(x_1-\mu_1)(x_k-\mu_k) \\ E(x_1-\mu_1)(x_2-\mu_2) & E(x_2-\mu_2)^2 & \cdots & E(x_2-\mu_2)(x_k-\mu_k) \\ \vdots & \vdots & \ddots & \vdots \\ E(x_1-\mu_1)(x_k-\mu_k) & E(x_2-\mu_2)(x_k-\mu_k) & \cdots & E(x_k-\mu_k)^2 \end{bmatrix} \end{aligned}$$

それぞれの成分には分散と共分散が入っていますから，この共分散行列 $\mathbf{\Sigma}$ と書くと以下が成り立っていることになります．

$$\mathbf{\Sigma} = E[\mathbf{v}\mathbf{v}'] = E\left[(\mathbf{x}-\mu)(\mathbf{x}-\mu)'\right] \tag{5.4.13}$$

ここまでの表現を使うと，標本平均と標本分散はそれぞれ母平均と母分散の一致推定量だということは，

$$\bar{\mathbf{X}} \xrightarrow{p} \mu, \quad \mathbf{S} \xrightarrow{p} \mathbf{\Sigma} \tag{5.4.14}$$

と表すことができます．それぞれ，k 次の縦ベクトル，$k \times k$ の行列ですから，表現が簡素になったことが分かるでしょう．

6　最小 2 乗推定量

6.1　推定量の導出

説明変数が k 個，サンプルの大きさが n のモデルを考えてみましょう．さしあたって説明変数の選択の問題は解決されているとして，定数項を明示的に取り扱うと次の線形連立方程式で表現されます．

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \\ Y_2 = \beta_0 + \beta_1 X_{22} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n \end{cases} \tag{6.1.1}$$

これを行列で表現したいので，式をじーっと見ると，右辺の第 $k+1$ 項目までは係数が共通であることに気がつくでしょう．そこで，

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{U} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{6.1.2}$$

とおいて，

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U} \tag{6.1.3}$$

とすれば，もとの式になります．ここで，縦ベクトル $\mathbf{y}$ は $n \times 1$，$\mathbf{X}$ は $n \times k$，縦ベクトル β は $k+1 \times 1$，縦ベクトル $\mathbf{U}$ は $n \times 1$，ですから，行列の積である $\mathbf{X}\beta$ の次元は $n \times 1$ となります．行列のふつうの積は「ヨコ×タテ」で定義されますから，1 行目がちゃんと

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \tag{6.1.4}$$

となっていることを確認しましょう．

さてここで，(6.1.3) についていくつか確認しておきましょう．$\mathbf{X}$ は確率変数，$\mathbf{U}$ のそれぞれの要素は実現値を観測できない確率変数です．観測できない β は定数ですが，$\mathbf{U}$ が確率変数ベクトルなので $\mathbf{y}$ のそれぞれの要素も確率変数となります．ここで，(6.1.3) という式の形と，$\mathbf{u}$ についていくつかの仮定を置けば，それっぽい β を決めることができる，というのが推定の手続きに他なりません．最小 **2** 乗法とは，$\mathbf{u}$ のそれぞれの要素はあまり大きくないと想定して，その 2 乗和を最小にするような β をもって推定値にする，という方法です．さてそこで，(6.1.3) を用いて最小にされるべき 2 乗和を計算してましょう．残差ベクトル $\hat{\mathbf{U}}$ を同様に定義して移項すると

$$\hat{\mathbf{U}} = \mathbf{y} - \mathbf{X}\mathbf{b} \tag{6.1.5}$$

となります．2 乗和は

$$\hat{u}_1^2+\hat{u}_2^2+\cdots+\hat{u}_n^2=\hat{\mathbf{U}}'\hat{\mathbf{U}}=\begin{bmatrix}\hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_n\end{bmatrix}\begin{bmatrix}\hat{u}_1\\ \hat{u}_2\\ \vdots\\ \hat{u}_n\end{bmatrix} \tag{6.1.6}$$

と書けますから，右辺も同じように書いてみると，

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=(\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b}) \tag{6.1.7}$$

和をとったものの転置行列は転置行列の和になるから，

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=(\mathbf{y}'-(\mathbf{X}\mathbf{b})')(\mathbf{y}-\mathbf{X}\mathbf{b}) \tag{6.1.8}$$

積の転置行列は転置したものの積になるから，

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=(\mathbf{y}'-\mathbf{b}'\mathbf{X}')(\mathbf{y}-\mathbf{X}\mathbf{b}) \tag{6.1.9}$$

カッコを外す分配則はふつうに成り立つから[24]，

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\mathbf{b}-\mathbf{b}'\mathbf{X}'\mathbf{y}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \tag{6.1.10}$$

3 項目を $\mathbf{b}'\mathbf{X}'$ と $\mathbf{y}$ の積とみると，積の転置行列は転置したものの積だから，

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\mathbf{b}-\mathbf{y}'\mathbf{X}\mathbf{b}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \tag{6.1.11}$$

2 項目と 3 項目をまとめると

$$\hat{\mathbf{U}}'\hat{\mathbf{U}}=\mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\mathbf{b}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \tag{6.1.12}$$

となります．適当に β の推定値 $\mathbf{b}$ を決めてやれば，それに応じた $\hat{\mathbf{u}}=\mathbf{y}-\mathbf{X}\mathbf{b}$ を計算することができるので，そこから計算された $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ のことを残差平方和とよびます．つまり，残差平方和を最小にする $\mathbf{b}$ を見つけるのが最小 2 乗法ということになります．残差平方和は推定値 $\mathbf{b}$ の関数ですから，残差平方和を最小にする $\mathbf{b}$ を見つけるには，残差平方和を $\mathbf{b}$ で微分してゼロと置いて解けばよいわけです[25]．

「残差平方和を $\mathbf{b}$ で微分して」といっても，$\mathbf{b}$ はベクトルですからそのまま微分を考えることができるわけではありません．ベクトルで微分するというと難しそうですが，ベクトルを構成するそれぞれの要素でそれぞれ微分して，それを重ねて書き直したもののことを，

[24] $(a+b)(c+d)=ac+ad+bc+bd$ が成り立つということ．

[25] 最大値を見つけることになるかもしれないので，ほんとうはもうちょっとチェックが必要．

ベクトルで微分するといっているだけのことです．つまり，関数 $f(\mathbf{x})$ が $(x_1, x_2, \cdots, x_k)$ の関数であるとき[26]，ベクトル $\mathbf{x} = (x_1, x_2, \cdots, x_k)$ で微分したものは，

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_k} \end{bmatrix} \tag{6.1.13}$$

です．もし，関数 $f(\mathbf{x})$ が $(x_1, x_2, \cdots, x_k)$ の線形和であれば，

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x} = \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \tag{6.1.14}$$

と表現できて，その微分は

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \mathbf{a} \tag{6.1.15}$$

となるので，まとめて書けば，

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a} \tag{6.1.16}$$

となります．2 次関数のばあいも同じように計算すれば，行列 $\mathbf{A}$ が対称行列なら，

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \tag{6.1.17}$$

となることが分かっています．さて，これだけの準備をして

$$\hat{\mathbf{U}}'\hat{\mathbf{U}} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \tag{6.1.18}$$

を $\mathbf{b}$ で微分することを考えましょう．最小 2 乗法で求められる推定値 $\mathbf{b} = \hat{\beta}$ は微分をゼロにする[27]から，

$$\frac{\partial(\hat{\mathbf{U}}'\hat{\mathbf{U}})}{\partial \mathbf{b}} = -2\mathbf{y}'\mathbf{X} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0} \tag{6.1.19}$$

[26]関数の値がベクトルではなくてただの数になるケースのみを考えます．ベクトル値関数の場合は横にも並べればよいだけです．

[27]各要素がすべてゼロになるということ．

が成り立ちます．転置を書き直して移項して 2 で割れば，正規方程式を得ます．

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \tag{6.1.20}$$

さて，行列では割り算が定義されていませんから，正規方程式から $\mathbf{b}$ を求めるには，

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \tag{6.1.21}$$

の両辺に左から $\mathbf{X}'\mathbf{X}$ の逆行列を掛ければよいことになります．ところが，逆行列というのはどんな行列に対しても定義されているわけではないことに注意しましょう．行列を構成するベクトルが 1 次独立でなければ逆行列は存在せず，その場合にその行列はランク落ちしています．それゆえ，データを入れた行列 $\mathbf{X}$ が多重共線性（perfect multicollinearity）をもっていれば逆行列が定義できず，あるいは多重共線に近い状況であれば逆行列の値が大きくぶれます．つまり，完全な多重共線が発生していれば，**OLS** 推定量は定義できないことになります．

さて，多重共線が発生していなければ，両辺に左から $\mathbf{X}'\mathbf{X}$ の逆行列を掛けて，OLS 推定量を計算することができます．すなわち，

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{6.1.22}$$

であり，この $\mathbf{b}$ をとくに $\hat{\beta}$ で表します．

6.2 不偏性・一致性の証明

このように求められた OLS 推定量が不偏性・一致性を持っていることを確認してみましょう．そのための前提を確認しておくと次のようになります．

仮定 1 誤差項 u_i の説明変数ベクトル $\mathbf{x_i}$ で条件づけられた期待値はゼロ．$E(u_i|\mathbf{x_i}) = 0$.

仮定 2 サンプル $(\mathbf{X_i}, Y_i)$ は独立に同一の分布に従う（i.i.d.）．

仮定 3 4 次モーメントの条件が満たされる．

仮定 4 多重共線が発生していない．すなわち $\mathbf{\Sigma_{xx}} \equiv E(\mathbf{x_i}\mathbf{x_i}')$ が逆行列をもつ．

仮定 **1** については，もっと弱い条件でもよいということが知られています．その条件は，誤差項と説明変数の積の期待値がゼロ（$E(\mathbf{x_i}u_i) = 0$）というもので，直交条件（**orthogonality condition**）と呼ばれます．仮定 **2** についてももっと弱い条件（ergotic stationary）でよいことが分かっています．

さて，これだけの準備をしてから不偏性の証明をしてみましょう．OLS 推定量の定義から，

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{6.2.23}$$

ですが，一番後ろの $\mathbf{y}$ はモデルの仮定から $\mathbf{y} = \mathbf{X}\beta + \mathbf{U}$ が成り立っているので代入すると，

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{U}) \tag{6.2.24}$$

分配則が成り立っていること，行列とその逆行列との積は単位行列になることから，

$$\begin{aligned} \mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} \end{aligned} \tag{6.2.25}$$

両辺の条件付き期待値をとると，

$$E(\mathbf{b}|\mathbf{X}) = E\left(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}|\mathbf{X}\right) \tag{6.2.26}$$

期待値の線形性と条件付き期待値の性質によって

$$E(\mathbf{b}|\mathbf{X}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left(\mathbf{U}|\mathbf{X}\right) \tag{6.2.27}$$

ここで仮定 1 から $E\left(\mathbf{U}|\mathbf{X}\right) = 0$ だから

$$E(\mathbf{b}|\mathbf{X}) = \beta \tag{6.2.28}$$

これで，OLS 推定量の不偏性が証明できました．

次に一致性の証明をしてみましょう．OLS 推定量の定義からさきほどと同じく $\mathbf{y}$ を消して移項すると，

$$\mathbf{b} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} = \left(\frac{1}{n}(\mathbf{X}'\mathbf{X})\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\mathbf{U}\right) \tag{6.2.29}$$

カッコの中を見てみると，

$$
\begin{aligned}
\frac{1}{n}(\mathbf{X}'\mathbf{X}) &= \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \\
&= \frac{1}{n}\begin{bmatrix} 1^2+1^2+\cdots+1^2 & \cdots & X_{k1}+X_{k2}+\cdots+X_{kn} \\ X_{21}+X_{22}+\cdots+X_{2n} & \cdots & X_{21}X_{k1}+X_{22}X_{k2}+\cdots+X_{2n}X_{kn} \\ \vdots & \ddots & \vdots \\ X_{k1}+X_{k2}+\cdots+X_{kn} & \cdots & X_{k1}^2+X_{k2}^2+\cdots+X_{kn}^2 \end{bmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} 1^2 & \cdots & X_{k1} \\ X_{21} & \cdots & X_{21}X_{k1} \\ \vdots & \ddots & \vdots \\ X_{k1} & \cdots & X_{k1}^2 \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}\mathbf{x_i}'
\end{aligned}
$$

同様に

$$
\frac{1}{n}\mathbf{X}'\mathbf{U} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}u_i \tag{6.2.30}
$$

となるので，

$$
\mathbf{S_{xx}} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}\mathbf{x_i}', \quad \bar{\mathbf{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}u_i
$$

とおけば，

$$
\mathbf{b} - \beta = \mathbf{S_{xx}}^{-1}\bar{\mathbf{g}}
$$

と書くことができます．さて，仮定 2 からサンプル $(\mathbf{X_i}, Y_i)$ は i.i.d. ですから，標本分散は真の分散に確率収束します．さらに確率収束の性質から逆数は逆数に収束して，

$$
\mathbf{S_{xx}} \xrightarrow{p} \mathbf{\Sigma_{xx}}, \quad \mathbf{S_{xx}}^{-1} \xrightarrow{p} \mathbf{\Sigma_{xx}}^{-1}
$$

また同様に，

$$
\bar{\mathbf{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}u_i \xrightarrow{p} E(\mathbf{x_i}u_i)
$$

ここで仮定 3 から $E(\mathbf{x_i}u_i) = 0$ が成り立つので，確率収束の性質から，

$$
\mathbf{S_{xx}}^{-1}\bar{\mathbf{g}} \xrightarrow{p} \mathbf{\Sigma_{xx}}^{-1}E(\mathbf{x_i}u_i) = \mathbf{0} \tag{6.2.31}
$$

まとめると,

$$\mathbf{b} - \beta = \mathbf{S_{xx}}^{-1}\bar{\mathbf{g}} \xrightarrow{p} \mathbf{0}$$

これで一致性が証明されました.