多変量正規分布

分散分析第９回講義：資料１

土居正明

1　はじめに

1.1　前提とする知識など

本稿は、「直交行列と Helmert 変換」の内容を多分に使いますので、読まれていない方はまずそちらをご覧ください。また、確率変数ベクトルとして $\mathbf{X}$ を用いるときと $\mathbf{x}$ を用いるときが両方あります。これらを区別することにはあまり熱心ではありませんので「ここは大文字にするべきだ」などの突っ込みは、していただくのは結構ですが改訂されない可能性をご考慮ください。

1.2　記号について

一つだけ、本稿では使いませんが、講義中板書で使う記号をここで定義しておきます。$\underline{GL_n(R)}$ で $\underline{n \times n}$ 可逆行列全体を意味します。つまり A が $n \times n$ 可逆行列である、ということを $A \in GL_n(R)$ と書きます[*1]。

2　多変量正規分布の定義と基本的性質

では多変量正規分布の定義を考えていきます。最初に「最も基本となる多変量正規分布」を考えましょう。

2.1　標準多変量正規分布

「定義：標準多変量正規分布」

確率変数 $X_1, \cdots, X_n$ が 独立に 標準正規分布 $N(0,1)$ に従うとする。このとき、確率変数を並べたベクトル $(X_1, \cdots, X_n)$ の従う確率分布のことを、**標準多変量正規分布**といい、$N(\mathbf{0}, I_n)$ と書く。

ここで、ベクトル $\mathbf{X}$、$\mathbf{x}$ を

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

とおきます。今、$X_1, \cdots, X_n$ はそれぞれ独立に $N(0,1)$ に従いますので、それぞれの確率密度関数は

$$f(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right) \qquad (i = 1, \cdots, n)$$

とおけます。独立な 確率分布に従う確率変数の同時密度関数は、それぞれの確率密度関数の積で表現できますので、

[*1] ノートを作っているときに「$n \times n$ 可逆行列」と何回も書いていて、うんざりしたので記号を導入することにしました。本稿作成時にはそれほど感じず毎回言葉で書いたので、とりあえず本稿中ではそのまま残してあります。

$(X_1, \cdots, X_n)$ の従う同時密度関数は

$$\begin{aligned} f(x_1, \cdots, x_n) &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2} \right) \right) \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2} \right) \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2} \mathbf{x}'\mathbf{x} \right) \end{aligned}$$

と表せます。これが**標準**多変量正規分布 $N(\mathbf{0}, I_n)$ の密度関数です。

2.2　一般の多変量正規分布

2.2.1　一般の多変量正規分布の定義

では次に、一般の多変量正規分布を定義しましょう。非常に重要なポイントは、**「多変量正規分布は、標準多変量正規分布を一次変換して、定数ベクトルを足したものである」**ということです。

はじめに記号をいくつか定義しておきます。

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix},\ \mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix},\ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

とします。そして、行列 A を逆行列を持つ $n \times n$ 行列[*2]とします。

「定義：多変量正規分布」

$\mathbf{X}$ が n 次元の標準多変量正規分布に従う とする。このとき $n \times n$ 行列 A と n 次元ベクトル $\boldsymbol{\mu}$ を用いて

$$\mathbf{Z} = A\mathbf{X} + \boldsymbol{\mu} \tag{1}$$

とおくと、$\mathbf{Z}$ の従う分布 を**多変量正規分布**といい、$N(\boldsymbol{\mu}, AA')$ と書く[*3]。

さて、これが定義です。まず最初に具体例を見ていきましょう。$n = 2$ で

$$A = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix},\ \boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

とします。このとき、$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\mathbf{0}, I_2)$、つまり、$X_1,\ X_2 \sim N(0,1)$ で独立のとき、

$$\mathbf{Z} = A\mathbf{X} + \boldsymbol{\mu}$$

とおくと、**Z の従う分布が多変量正規分布**です。

では次に、成分計算をしてみます。

$$(\mathbf{Z} =) \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = A\mathbf{X} + \boldsymbol{\mu} = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2X_1 + 3X_2 + 2 \\ 4X_1 + 5X_2 + 1 \end{pmatrix}$$

[*2] 「可逆行列」や、「正則行列」とも言います。A が「正則行列」⇔「行列式が 0 でない」⇔「各列ベクトルが一次独立である」⇔「各行ベクトルが一次独立である」⇔「列ベクトル全体が R^n の基底となる」など、色々なとらえ方があります。。

[*3] これから詳しく見ていきますが、このとき $E[Z] = \boldsymbol{\mu}$, $V[Z] = AA'$ です。また、$AA' = \Sigma$ とおいて $N(\boldsymbol{\mu}, \Sigma)$ と書かれることが多いですが、議論の流れ上もうしばらくは $N(\boldsymbol{\mu}, AA')$ と表現しておきます。

より、

$$Z_1 = 2X_1 + 3X_2 + 2$$
$$Z_2 = 4X_1 + 5X_2 + 1$$

となります。１変数の正規分布の基本的性質より、$Z_1,\ Z_2$ は独立な正規分布２つと定数の和ですので、共に正規分布に従います。同じ理由で、一般に n 次元の場合でも Z の各成分は全て正規分布に従う、ということが分かります。

$\mathbf{Z}$ の平均・分散は、以下のようになります。

$$E[\mathbf{Z}] = AE[\mathbf{X}] + \boldsymbol{\mu} = \mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu} \tag{2}$$
$$V[\mathbf{Z}] = AV[\mathbf{X}]A' + \mathbf{0} = AI_nA' = AA' \tag{3}$$

となります。次に、Z_1, Z_2 の共分散を求めてみましょう。X_1, X_2 はそれぞれ独立であり、$E[X_1] = 0,\ E[X_2] = 0$ のため、$E[Z_1] = E[2X_1 + 3X_2 + 2] = 2, E[Z_2] = E[4X_1 + 5X_2 + 1] = 1$ なので

$$\begin{aligned}
Cov[Z_1, Z_2] &= E[(Z_1 - E[Z_1])(Z_2 - E[Z_2])] \\
&= E[(Z_1 - 2)(Z_2 - 1)] \\
&= E[(2X_1 + 3X_2 + 2 - 2)(4X_1 + 5X_2 + 1 - 1] \\
&= E[(2X_1 + 3X_2)(4X_1 + 5X_2)] \\
&= E[2 \cdot 4X_1^2 + 2 \cdot 5X_1X_2 + 3 \cdot 4X_2X_1 + 3 \cdot 5X_2^2] \\
&= 2 \cdot 4E[X_1^2] + 2 \cdot 5E[X_1]E[X_2] + 3 \cdot 4E[X_2]E[X_1] + 3 \cdot 5E[X_2^2] \\
&\qquad (\because X_1,\ X_2 \text{ 独立より } E[X_1X_2] = E[X_1]E[X_2]) \\
&= 2 \cdot 4E[X_1^2] + 3 \cdot 5E[X_2^2] \qquad (\because E[X_1] = E[X_2] = 0) \\
&= 2 \cdot 4V[X_1] + 3 \cdot 5V[X_2] = 23 \\
&\qquad (\because V[X_1] = E[(X_1 - 0)^2) = E[X_1^2],\ V[X_2] = E[(X_2 - 0)^2] = E[X_2^2])
\end{aligned}$$

となります。ここで $V[X_1]$ の係数の 2 と 4 はそれぞれ Z_1, Z_2 の X_1 の係数、$V[X_2]$ の係数の 3 と 5 は Z_1, Z_2 の X_2 の係数となっています。つまり、Z_1 と Z_2 の共分散 は Z_1, Z_2 の中に**「X_1 という共通の因子がどのくらいずつ入っているか」「X_2 という共通の因子がどのくらいずつ入っているか」**$\cdots$ **「X_n という共通の因子がどのくらいずつ入っているか」**[*4]だけに依存することになります。

n 次元の場合に拡張して、重要な点をまとめておきます。n 次元多変量正規分布に従う確率変数ベクトル $\mathbf{Z}$ は、

(i) **n 次元標準多変量正規分布に従う確率変数ベクトル $\mathbf{X}$ に可逆行列をかけて、定数ベクトルを足したものである。**

(ii) **各成分は、独立な標準正規分布に従う確率変数の一次結合 (＋定数) である。従って、正規分布に従う。**

(iii) **成分同士の相関係数の計算には、各成分における「X_1 の係数同士の積」「X_2 の係数同士の積」…だけを用いる。**

2.2.2 多変量正規分布の確率密度関数

では次に $\mathbf{Z}$ の従う多変量正規分布 $N(\boldsymbol{\mu}, AA')$ の確率密度関数を導いていきましょう。

先ほど、$\mathbf{X}$ の従う確率密度関数が

$$f(x_1, \cdots, x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right)$$

ということはみました。これを (1) を用いて変形します。

$$\mathbf{Z} = A\mathbf{X} + \boldsymbol{\mu} \Longleftrightarrow \mathbf{X} = A^{-1}(\mathbf{Z} - \boldsymbol{\mu})$$

[*4] X_n の n は n 次元正規分布の n と同じ数です。つまりたとえば、2 次元正規分布には 2 個の独立な正規分布の和ですので、この因子 2 つがどの程度ずつ入っているか、が問題となります。

より、

$$dx_1 dx_2 \cdots dx_n = ||A^{-1}|| dz_1 dz_2 \cdots dz_n$$

が成り立ちますので[*5]、

$$\begin{aligned}
f(x_1, \cdots, x_n) dx_1 \cdots dx_n &= \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right) dx_1 dx_2 \cdots dx_n && (4)\\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\left\{A^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}'\left\{A^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}\right) ||A^{-1}|| dz_1 dz_2 \cdots dz_n && (5)\\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n ||A||} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'(A')^{-1}A^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n && (6)\\
&\quad \left(\because ||A^{-1}|| = \frac{1}{||A||}\right) && (7)\\
& && (8)\\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n ||A||} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'(AA')^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n && (9)\\
& && (10)\\
&\quad (\because (BC)^{-1} = C^{-1}B^{-1}) && (11)
\end{aligned}$$

となります。もう少し整理するために、式 (3) を利用しましょう。今

$$V[X] = AA'$$

より、これを Σ とおく と、$\Sigma = AA'$ であり、Σ は正値対称行列となるので $||\Sigma|| = |\Sigma| > 0$ となり[*6]、以下の性質が成り立ちます。

$$\begin{aligned}
&\Sigma^{-1} = (AA')^{-1}\\
&||\Sigma|| = ||A|| \cdot ||A'|| = ||A||^2 \Longleftrightarrow ||A|| = \sqrt{|\Sigma|}
\end{aligned}$$

となります。これを、(11) に代入すると

$$f(x_1, \cdots, x_n) dx_1 dx_2 \cdots dx_n = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n$$

となります。これより、$\mathbf{Z}$ の従う多変量正規分布の密度関数 $g(z_1, \cdots, z_n)$ は

$$\begin{aligned}
g(z_1, \cdots, z_n) dz_1 dz_2 \cdots dz_n &= f(x_1, \cdots, x_n) dx_1 dx_2 \cdots dx_n\\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n
\end{aligned}$$

つまり、

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right)$$

です。これは定理の形でまとめておきましょう。

[*5] ここで、$||A||$ は「A の行列式の絶対値」を指します。

[*6] 正値対称行列の定義は**「定理２」**の直前に回します。このあたりは結構面倒なので、申し訳ないですが基本的な性質は証明なしに用いることとします。

「定理１：多変量正規分布の確率密度関数」

$\mathbf{X}$ を n 次元標準多変量正規分布 $N(\mathbf{0}, I_n)$ に従う確率密度関数とする。このとき、$n \times n$ 可逆行列 A と定数ベクトル $\boldsymbol{\mu}$ を用いて $\mathbf{Z} = A\mathbf{X} + \boldsymbol{\mu}$ と表される確率変数ベクトル $\mathbf{Z}$ の従う多変量正規分布の密度関数は

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right)$$

と表される。このとき、この多変量正規分布を $N(\boldsymbol{\mu}, \Sigma)$ と書き、平均・分散は

$$E[X] = \boldsymbol{\mu},\ \ V[X] = \Sigma\ \ (= AA')$$

である。

2.2.3 系１

次に、定義と上の定理からほとんど明らかな系を示します。

「系１」

$\mathbf{Z}$ が n 次元の多変量正規分布 $N(\boldsymbol{\mu}, \Sigma)$ に従うとき、$n \times n$ 可逆行列 A と n 次元ベクトル $\boldsymbol{\nu}$ を用いて

$$\mathbf{W} = A\mathbf{Z} + \boldsymbol{\nu}$$

とおくと $\mathbf{W}$ は n 次元多変量正規分布 $N(A\boldsymbol{\mu} + \boldsymbol{\nu}, A\Sigma A')$ に従う[*7]。

（証明）

多変量正規分布の定義と上の定理より、確率変数ベクトル $\mathbf{Z} \sim N(\boldsymbol{\mu}, \Sigma)$ に対して、n 次元標準多変量正規分布に従う確率変数ベクトル $X \sim N(\mathbf{0}, I_n)$ と $\underline{BB' = \Sigma \text{ を満たす } n \times n \text{ 可逆行列 } B \text{ が存在して}}$[*8]

$$\mathbf{Z} = B\mathbf{X} + \boldsymbol{\mu}$$

とおけます。これより、$\mathbf{W}$ を $\mathbf{X}$ で表現すると

$$\begin{aligned}\mathbf{W} &= A\mathbf{Z} + \boldsymbol{\nu} = A\left(B\mathbf{X} + \boldsymbol{\mu}\right) + \boldsymbol{\nu} \\ &= \underbrace{(AB)}_{\text{可逆行列}}\mathbf{X} + \underbrace{A\boldsymbol{\mu} + \boldsymbol{\nu}}_{\text{定数ベクトル}}\end{aligned}$$

となります。多変量正規分布の定義より、このとき $\mathbf{W}$ が多変量正規分布に従います。平均・分散は上の定理より $\mathbf{Z}$ を基準に計算して

$$E[W] = AE[\mathbf{Z}] + \boldsymbol{\nu} = A\boldsymbol{\mu} + \boldsymbol{\nu},\ \ V[W] = AV[\mathbf{Z}]A' + \mathbf{0} = A\Sigma A'$$

となります。したがって、$\mathbf{W} \sim N(A\boldsymbol{\mu} + \boldsymbol{\nu}, A\Sigma A')$ となります。　（証明終わり）

---

*7 つまり「（標準でなくてもよい）多変量正規分布に従う確率変数ベクトルに行列をかけて、定数ベクトルを足しても、やっぱり多変量正規分布に従う」ということです。

*8 この部分は仮定します。「補足」の証明のところに少し説明を書いておきました。

3　定義と同値な表現

次に、上の「定理」と「定義」の関係について考えてみましょう。多変量正規分布について少しご存じの方に「多変量正規分布の定義は何ですか?」という風なご質問をすると、本稿の「定義」ではなくて、「定理」の方（確率密度関数の方）を「定義です」と言われる方が結構いらっしゃいます。実は**それが本稿の定義と同値である**ということをこれから示します[*9]。

最初に 1 点だけ注意しましょう。先ほど Σ を、可逆行列 A を用いて $\Sigma = AA'$ と定義しました。これより、Σ は対称行列であり、行列式は 0 ではありません。さらに、任意のベクトル $\mathbf{y}(\neq \mathbf{0})$ に対して、

$$\begin{aligned}\mathbf{y}'\Sigma\mathbf{y} &= \mathbf{y}'(AA')\mathbf{y} \\ &= (A'\mathbf{y})'(A'\mathbf{y}) \\ &= ||A'\mathbf{y}||^2 \\ &= |A'|^2||\mathbf{y}||^2 > 0\end{aligned}$$

となります。このようなとき、Σ は**正値対称行列**といいます。Σ が任意の正値対称行列のとき、$P\Sigma P' = Q$ （P：直交行列、Q：対角成分が全て正の対角行列）が成り立ちます[*10]。これより、以下の定理が成り立ちます。

「定理２：多変量正規分布の定義と同値な表現」

確率変数ベクトル $\mathbf{Z}$ の密度関数が、<u>正値対称行列</u> Σ と定数ベクトル $\boldsymbol{\mu}$ を用いて

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right)$$

と表現できることと、$\mathbf{Z}$ が n 次元多変量正規分布に従うことは同値である。

この定理より、**上の確率密度関数を見た時に、「各確率変数 $z_1, \cdots, z_n$ を n 個の独立な標準正規分布の一次結合 + 定数」と見てもよい**ことになりました。証明は形式的で結構面倒なので補足に回しますが、本稿の後の議論において重要なことは

<u>(i)　Σ が正値対称行列でありさえすれば、それに対して可逆な行列 A が存在して $\Sigma = AA'$ と表現できること</u>

<u>(ii)　$\mathbf{x} = A^{-1}(\mathbf{z}-\boldsymbol{\mu})$ とおくと、$\mathbf{x} \sim N(\mathbf{0}, I_n)$ となること</u>

の 2 つです。この性質は次章で使いますので覚えておいてください[*11]。

4　χ^2 分布との関わり

基本的なことは全て終わりましたので、以下は応用です。まずは χ^2 を見ていきましょう。

自由度 n の χ^2 分布とは、$x_1, \cdots, x_n \sim N(0,1)$ が独立なとき、

$$\chi^2 = x_1^2 + \cdots x_n^2$$

の従う分布のことでした。さて、この式をベクトル $\mathbf{x}$ を用いて表記してやると

$$\chi^2 = \mathbf{x}'\mathbf{x}$$

という風に書けます。この形の式は既に本稿に出てきましたね。そうです。標準多変量正規分布の密度関数

$$f(x_1, \cdots, x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right)$$

[*9] 本稿であえてこのような定義をしたのは、式で定義してしまうと上で述べたような多変量正規分布の「意味」が大変伝わりにくくなるからです。もう一度強調しておきますと、n 次元多変量正規分布の各成分は n **個の独立な標準正規分布と定数の和**になっています。これが本稿で最も強調したい点です。

[*10] この証明は結構面倒なので、線形代数の教科書に回します。

[*11] 「定義」→「定理 1」の流れでは、「最初に可逆行列 A があって、それから Σ を作る」という順なので、Σ に対して A が存在するのは当たり前です。一方で、今回の定理は最初に Σ を「任意の正値対称行列」ととりますので、これに対応する A が存在するのは全く自明ではありません。

の exp ( ) の中の主要部分です。$\mathbf{x}$ が n 次元の**標準**多変量正規分布に従う場合、ここの部分もまた自由度 n の χ^2 分布に従うのです。では、**一般の多変量正規分布**で考えてみましょう。密度関数は

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right)$$

となります。先の $\mathbf{x}'\mathbf{x}$ に対応する部分は

$$(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})$$

です。**これは何分布に従うのでしょうか？**

実は これもまた自由度 n の χ^2 分布に従う のです。これを示す重要なポイントは Σ **に対して、可逆行列** A **が存在して** $\Sigma = AA'$ **と書ける**ことです。$\mathbf{z}$ に対して、$\mathbf{x} = A^{-1}(\mathbf{z}-\boldsymbol{\mu})$ と置いてやります。このとき、

$$\begin{aligned}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu}) &= (\mathbf{z}-\boldsymbol{\mu})'(AA')^{-1}(\mathbf{z}-\boldsymbol{\mu}) && (12)\\ &= (\mathbf{z}-\boldsymbol{\mu})'A'^{-1}A^{-1}(\mathbf{z}-\boldsymbol{\mu}) && (13)\\ &= \{A^{-1}(\mathbf{z}-\boldsymbol{\mu})\}'\{A^{-1}(\mathbf{z}-\boldsymbol{\mu})\} && (14)\\ &= \mathbf{x}'\mathbf{x} && (15)\end{aligned}$$

となるので、

$$(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu}) = \mathbf{x}'\mathbf{x} = x_1^2 + \cdots + x_n^2 \ \sim\ \chi^2(n)$$

です。

さて、前章で $\mathbf{x}$ は $N(\mathbf{0}, I_n)$ に従うことはみていますので、最後の式 (15) が自由度 n の χ^2 分布に従うことはお分かりいただけたと思います。これも定理の形でまとめておきましょう。

「定理３：χ^2 分布」

$\mathbf{z}$ が多変量正規分布 $N(\boldsymbol{\mu}, \Sigma)$ に従うとき

$$(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})$$

は自由度 n の χ^2 分布に従う。

（注）この式はよく見る

$$\frac{(O-E)^2}{V}$$

の形の式を多次元に拡張したものと考えることができます[*12]。

4.1 定理３に対する注釈

いくつか誤解されそうな点を注意しておきましょう。

4.1.1 注意１

先ほど $\mathbf{x} \sim N(\mathbf{0}, I_n)$ に対して、$\mathbf{z} = A\mathbf{x} + \boldsymbol{\mu}$ と確率変数の変換をすると、$\mathbf{z}$ の従う分布は $N(\mathbf{0}, I_n)$ ではなく、$N(\boldsymbol{\mu}, AA')$ となる、ということをみました。

ところが、「今回は $\mathbf{z}$ を $\mathbf{x}$ に変換しても、そのまま同じ分布に従っているのではないか？」という風に思われる方がいらっしゃるかもしれません。

[*12] この定理と中心極限定理の２つが理解できれば、尤度比検定、Wald 検定、Score 検定が漸近的に χ^2 分布に従うことが統一的に理解できるようになります。

ですが、この点は問題がないのです。なぜなら、確かに $\mathbf{x}$ と $\mathbf{z}$ の従う分布はそれぞれ異なるのですが、

$$(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu}) = \mathbf{x}'\mathbf{x}$$

という風に、$\mathbf{z}$ と $\mathbf{x}$ でまとめ方が変わっているからです[*13]。そして、左辺と右辺は「全く同じ」ものなので、同じ分布に従うのです。

4.1.2 注意２

そんなこと言っても、分散が分かっていることなど基本的にはあり得ないからこの定理は使えないのではないか？ と思う方がいらっしゃるかも知れません。それに対する一つの回答は先の注釈で述べました「中心極限定理」で、「例数が多いときは漸近的には分散既知とほとんど同じ」とみなせることがあるから問題ない、というものです。

一方、「では Σ を推定量で置き換えたら？」というという疑問も当然出てくるでしょう。実は、Σ を適当な[*14]推定量で置き換えたときに

$$(\mathbf{z}-\boldsymbol{\mu})'\widehat{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})$$

の定数倍は *Hotelling の T 統計量* と言われ、F 分布に従うことが分かっています。「正規分布・t 分布・χ^2 分布・F 分布とは何か」の資料で「F 分布は、χ^2 分布の分散未知ヴァージョン」というご説明をしましたが、今回も「分散既知：χ^2 分布」「分散未知：F 分布」という、大変自然な対応がついています。$\widehat{\Sigma}$ の定義や、この事実の証明は簡単そうに見えて意外と面倒なので省略します。

5 定理の証明

では、以前結果のみ述べて証明を後回しにしておいた定理を１つ証明しましょう。準備として１つ、多変量正規分布と直交行列の両方を用いて補題を証明しておきます。

5.1 準備の補題

「補題１」

$\mathbf{x} \sim N(\mathbf{0}, I_n)$ とする。このとき、任意の $P \in O(n)$ に対して、$\mathbf{y} = P'\mathbf{x}$ とおくと、$\mathbf{y} \sim N(\mathbf{0}, I_n)$ である。

(証明)

密度関数を変形していけば終わりです。まず、$\mathbf{y} = P'\mathbf{x}$ より、$\underline{\mathbf{x} = P\mathbf{y}}$ です[*15]。

$\mathbf{x}$ の従う密度関数は

$$f(x_1, \cdots, x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right)$$

でした。さらに、$|P| = \pm 1$ より、

$$dx_1 \cdots dx_n = dy_1 \cdots dy_n$$

[*13] $\mathbf{x}'\mathbf{x}$ は、$\mathbf{x}$ をそのまま $\mathbf{z}$ に置き換えた $\mathbf{z}'\mathbf{z}$ とは違う分布になります。一方、$\mathbf{x}'\mathbf{x}$ と $(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})$ では、$\mathbf{x}$ のまとめかたと $\mathbf{z}$ のまとめ方が異なりますので、分布が一致していても問題ないのです。

[*14] 数学の文章中で**「適当な」**という表現が出てきたら、「いい加減な」ではなく「ふさわしい」という意味です。意味が大きく変わってしまいますので、この点にもご注意ください。

[*15] $P \in O(n)$ より、$P' = P^{-1}$ です。

が成り立ちます。

$$\begin{aligned}
f(x_1,\cdots,x_n)dx_1\cdots dx_n &= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right)dx_1\cdots dx_n \\
&= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}(P\mathbf{y})'(P\mathbf{y})\right)dy_1\cdots dy_n \quad (\because \mathbf{x}=P\mathbf{y}) \\
&= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\mathbf{y}'P'P\mathbf{y}\right)dy_1\cdots dy_n \\
&= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right)dy_1\cdots dy_n \quad (\because P\in O(n) \text{ より } P'P=I_n)
\end{aligned}$$

これより、$\mathbf{y}$ の従う密度関数を $g(y_1,\cdots,y_n)$ とおくと

$$\begin{aligned}
g(y_1,\cdots,y_n) &= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) \\
&= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}(y_1^2+\cdots+y_n^2)\right) \\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y_i^2}{2}\right)\right)
\end{aligned}$$

となり、$y_1,\cdots,y_n$ はそれぞれ独立に $N(0,1)$ に従う ことが分かります。これより定義から $\mathbf{y}$ は $N(\mathbf{0},I_n)$ に従います。

(証明終わり)

この補題で何が言いたいかと言いますと、**各成分が 独立な標準正規分布 に従う確率変数ベクトルを 直交行列で変換 しても、やはり各成分が 独立な標準正規分布 に従う**ということです。

5.2 定理と証明

示したいのは一般の正規分布を対象とした**「系２」**なのですが、最初に**「定理４」**で標準正規分布の場合に対して示しておいて、そこから拡張することにします[*16]。

「定理４：標準正規分布の標本平均と標本分散の独立性」

$x_1,\cdots,x_n \sim N(0,1)$ が独立とする。このとき、

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i \quad (=\bar{x}\text{とも書く})$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

は 互いに独立 であり、

$$\widehat{\mu} \sim N\left(0,\frac{1}{n}\right)$$

$$n\widehat{\sigma^2} \sim \chi^2(n-1)$$

となる[*17]。

---

*16 「定理」と「系」と聞くと、「定理」の方が凄そうな感じがしますが、「定理」の方が「準備」で、そこで頑張っておいて、言いたいことを「系」でさらっと述べる、ということはよく行われます。

*17 今回は、標準正規分布 $N(0,1)$ を扱っているので、μ,σ^2 は平均・分散という意味付けはありません。この後の「系２」で $N(\mu,\sigma^2)$ に拡張します。

(証明)

重要なポイントは Helmert 行列を用いて、新しい確率変数に変換する ことです。まず、$x_1, \cdots, x_n$ をまとめてベクトル $\mathbf{x}$ とします。そして n 次の Helmert 変換

$$H_n = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \cdots & 0 & 0 \\ \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \cdots & \frac{-(n-2)}{\sqrt{(n-2)+(n-2)^2}} & 0 \\ \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \cdots & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{-(n-1)}{\sqrt{(n-1)+(n-1)^2}} \end{pmatrix}$$

を用いて、

$$\mathbf{y} = H_n \mathbf{x}$$

を考えます。今、$\mathbf{x}$ は $N(\mathbf{0}, I_n)$ に従いますので、Helmert 変換が直交行列であることから、先の**「補題１」**より $\mathbf{y}$ も $N(\mathbf{0}, I_n)$ に従います。したがって、変換後の $y_1, \cdots, y_n$ もまた独立に標準正規分布に従う ということになります。密度関数を見ますと、先の補題の証明より

$$\frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right) dx_1 \cdots dx_n = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) dy_1 \cdots dy_n$$

が成り立ちます。つまり

$$\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} \iff x_1^2 + \cdots + x_n^2 = y_1^2 + \cdots + y_n^2 \tag{16}$$

が言えることが分かります。

次に具体的な成分に注目しますと、

$$\mathbf{y} = H_n \mathbf{x}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-2} \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \cdots & 0 & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \frac{1}{\sqrt{(n-3)+(n-3)^2}} & \cdots & 0 & 0 \\ \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \frac{1}{\sqrt{(n-2)+(n-2)^2}} & \cdots & \frac{-(n-2)}{\sqrt{(n-2)+(n-2)^2}} & 0 \\ \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \cdots & \frac{1}{\sqrt{(n-1)+(n-1)^2}} & \frac{-(n-1)}{\sqrt{(n-1)+(n-1)^2}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-2} \\ x_{n-1} \\ x_n \end{pmatrix}$$

となります。実は 重要なのは y_1 のみ なので、ここだけを計算してやると

$$
\begin{aligned}
y_1 &= \frac{1}{\sqrt{n}} \cdot x_1 + \cdots + \frac{1}{\sqrt{n}} \cdot x_n \\
&= \sqrt{n} \cdot \frac{x_1 + \cdots + x_n}{n} \\
&= \sqrt{n}\bar{x}
\end{aligned}
$$

ですので、

$$
\bar{x} = \frac{1}{\sqrt{n}} y_1 \tag{17}
$$

となります。これより

$$
\begin{aligned}
\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2 - n\left(\frac{1}{\sqrt{n}} y_1\right)^2\right) \qquad (\because (16), (17)) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2 - y_1^2\right) \\
&= \frac{1}{n}\sum_{i=2}^{n} y_i^2
\end{aligned}
$$

これより、$\widehat{\mu} = \bar{x}$ は y_1 のみからなり、$\widehat{\sigma^2}$ は $y_2, \cdots, y_n$ のみからなる ことが分かりました。今、**先の補題より** $y_1, \cdots, y_n$ **は 全て独立に 標準正規分布に従う**ことより、$\widehat{\mu}$ と $\widehat{\sigma^2}$ は独立である ことが分かります。

次に、$\widehat{\mu}, \widehat{\sigma^2}$ の従う分布ですが、$\widehat{\mu}$ の従う分布が $N\left(0, \frac{1}{n}\right)$ であることはよいでしょう。次に、$\widehat{\sigma^2}$ について

$$
n\widehat{\sigma^2} \sim \chi^2(n-1)
$$

を示しましょう。しかし、これも既に分かっているようなもので、$y_1, \cdots, y_n \sim N(0,1)$ でそれぞれ独立より、先の計算より

$$
\widehat{\sigma^2} = \frac{1}{n}\sum_{i=2}^{n} y_i^2 \implies n\widehat{\sigma^2} = \sum_{i=2}^{n} y_i^2 = \underbrace{y_2^2 + y_3^2 + \cdots + y_n^2}_{(n-1)\text{ 個}}
$$

となることから、$\sum$ で足されているのが 2 から n までの $(n-1)$ 個ですので、これは $\chi^2(n-1)$ に従います。(証明終わり)

ではこれを一般の正規分布に拡張しましょう。

「系 2：一般の正規分布の標本平均と標本分散の独立性」

$x_1, \cdots, x_n \sim N(\mu, \sigma^2)$ が独立とする。このとき、

$$
\begin{aligned}
\widehat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i \quad (= \bar{x}\text{とも書く}) \\
\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2
\end{aligned}
$$

は 互いに独立 であり、

$$
\widehat{\mu} \sim N\left(\mu, \frac{1}{n}\sigma^2\right), \qquad n\frac{\widehat{\sigma^2}}{\sigma^2} \sim \chi^2(n-1)
$$

となる。

(証明)

方針としては $\mathbf{x}$ を標準正規分布に結びつけて、「定理４」を使います。確率変数 $z_1, \cdots, z_n$ を

$$z_i = \frac{x_i - \mu}{\sqrt{\sigma^2}} \qquad (i = 1, \cdots n)$$

とおきますと、$E[z_i] = 0,\ V[z_i] = 1$ であり、$x_1, \cdots, x_n$ が全て独立に正規分布に従うことから、$z_1, \cdots, z_n$ **も全て独立で、こちらは標準正規分布に従います**[*18]。$\widehat{\mu}, \widehat{\sigma^2}$ を $z_1, \cdots, z_n$ を用いて表現しますと、

$$\begin{aligned}
\widehat{\mu} &= \frac{x_1 + \cdots + x_n}{n} \\
&= \frac{1}{n}\left(\frac{x_1 - \mu}{\sqrt{\sigma^2}} + \cdots \frac{x_1 - \mu}{\sqrt{\sigma^2}}\right)\sqrt{\sigma^2} + \mu \\
&= \frac{1}{n}(z_1 + \cdots + z_n)\sqrt{\sigma^2} + \mu \\
&= \sqrt{\sigma^2}\bar{z} + \mu \qquad \left(\bar{z} = \frac{1}{n}(z_1 + \cdots + z_n) \text{ とおく}\right)
\end{aligned}$$

$$\begin{aligned}
\widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\left(\frac{x_i - \mu}{\sqrt{\sigma^2}}\sqrt{\sigma^2} + \mu\right) - \left(\sqrt{\sigma^2}\bar{z} + \mu\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sqrt{\sigma^2}}\sqrt{\sigma^2} - \sqrt{\sigma^2}\bar{z}\right)^2 \\
&= \sigma^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2
\end{aligned}$$

となります。ここで、

$$\begin{aligned}
\widehat{\mu}_z &= \frac{1}{n}\sum_{i=1}^{n} z_i \\
\widehat{\sigma^2}_z &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2
\end{aligned}$$

とおくと、$z_1, \cdots, z_n$ は独立に $N(0, 1)$ に従いますので、**「定理４」**より

(i) $\widehat{\mu}_z$ と $\widehat{\sigma^2}_z$ は独立

(ii) $\widehat{\mu}_z \sim N\left(0, \frac{1}{n}\right)$

(iii) $n\widehat{\sigma^2}_z \sim \chi^2(n-1)$

が成り立ちます。今、

$$\begin{aligned}
\widehat{\mu} &= \sqrt{\sigma^2}\bar{z} + \mu = \sqrt{\sigma^2}\widehat{\mu}_z + \mu \\
\widehat{\sigma^2} &= \sigma^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2 = \sigma^2 \cdot \widehat{\sigma^2}_z \iff \widehat{\sigma^2}_z = \frac{\widehat{\sigma^2}}{\sigma^2}
\end{aligned}$$

より、上の (i)〜(iii) を用いると

(i)' $\widehat{\mu}$ と $\widehat{\sigma^2}$ は独立

(ii)' $\widehat{\mu} \sim N\left(\mu, \frac{1}{n}\sigma^2\right)$

(iii)' $n\frac{\widehat{\sigma^2}}{\sigma^2} \sim \chi^2(n-1)$

がそれぞれ成り立ちます。 (証明終わり)

[*18] つまり、$z_1, \cdots, z_n$ に対しては**「定理４」**が使える、ということです。

6　補足：「定理２」の証明

では、本文では述べませんでした**「定理２」**の証明をします。まずは**「定理２」**をもう一度書きましょう。

「定理２：多変量正規分布の定義と同値な表現」

確率変数 $\mathbf{Z}$ の密度関数が、正値対称行列 Σ を用いて

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right)$$

と表現できることと、$\mathbf{Z}$ が n 次元多変量正規分布に従うことは同値である。

(証明)

示すべきことは、$\mathbf{Z}$ の 密度関数が上の式で与えられるとき、$\mathbf{Z}$ が多変量正規分布に従うこと だけです。いま、「多変量正規分布の定義」は「標準多変量正規分布に従う確率変数に行列をかけて、定数ベクトルを足してやったものの従う分布」でしたので、逆に**「Z から定数ベクトルをひいて、一次変換してやると標準多変量正規分布に従う」**ことを示してやれば十分です。そのための方針としては、Σ をいじってうまく $\Sigma = AA'$ を満たす A に対応する行列が作れる ことを示せばよいことになります[*19]。

まず、Σ は正値対称行列ですので、ある直交行列 P と**対角成分が全て正の対角行列** Q を用いて、

$$P\Sigma P' = Q \tag{18}$$

と表せます[*20]。ここで、

$$Q = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \lambda_{n-2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \lambda_n \end{pmatrix} \quad (\lambda_1, \cdots \lambda_n > 0)$$

とおけるので、各成分の平方根も正に取ることができます。ここで R という行列を

$$R = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\lambda_3} & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_{n-2}} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sqrt{\lambda_{n-1}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \sqrt{\lambda_n} \end{pmatrix}$$

という風に定義します。このとき、R は対称行列 $(R' = R)$ より、

$$Q = RR = R'R$$

と書けます。これより、式 (18) は

$$\begin{aligned} P\Sigma P' = Q &\iff P\Sigma P' = R'R \\ &\iff \Sigma = P'R'RP \quad (\because P \in O(n) \text{ より、} P^{-1} = P') \\ &\iff \Sigma = (RP)'(RP) \end{aligned}$$

[*19] もう一度書いておきますが、A があるときに $\Sigma = AA'$ を作るのが簡単ですが、Σ **しかないときに** $\Sigma = AA'$ **を満たす** A **を見つけるのはそれほど簡単ではありません**。このような A が存在することすら、直感的には明らかではありません。そこで、それがきちんと存在することを見ていきましょう、ということです。

[*20] この部分の証明は面倒ですので省略します。線形代数の大抵の（きちんとした）教科書には書いてあるはずです。

と書けます。ここで、行列 A を

$$A = (RP)'$$

と定義しますと、

$$\Sigma = AA'$$

が成り立ちます。

いま、R：対角成分が全て正の対角行列、P：直交行列より、ともに可逆行列です。したがって、その積 $A' = RP$ もまた可逆行列となります。以上より、**Σ が正値対称行列でありさえすれば、可逆行列 A が存在して $\Sigma = AA'$ と表現できることが分かりました**。これを用いて、$\mathbf{Z}$ が多変量正規分布に従うことを見ていきます。

最初の議論とは逆に、確率変数ベクトル $\mathbf{X}$ を

$$\mathbf{X} = A^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \tag{19}$$

で定義します[*21]。このとき、$\mathbf{X}$ の従う密度関数が標準多変量正規分布になることを示します。先ほどとは逆に

$$dz_1 \cdots dz_n = ||A|| dx_1 \cdots dx_n \quad (\because ||A'|| = ||A||)$$

となり、さらに $\Sigma = A'A$ より $||A|| = \sqrt{|\Sigma|}$ が成り立ちますので、

$$dz_1 \cdots dz_n = \sqrt{|\Sigma|} dx_1 \cdots dx_n$$

です。これより密度関数は

$$\begin{aligned}
g(z_1, \cdots, z_n) dz_1 \cdots dz_n &= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{z} - \boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' (AA')^{-1} (\mathbf{z} - \boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' A'^{-1} A^{-1} (\mathbf{z} - \boldsymbol{\mu})\right) dz_1 dz_2 \cdots dz_n \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \left\{(A^{-1}(\mathbf{z} - \boldsymbol{\mu})\right\}' \left\{(A^{-1}(\mathbf{z} - \boldsymbol{\mu})\right\}\right) dz_1 dz_2 \cdots dz_n \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right) \sqrt{|\Sigma|} dx_1 dx_2 \cdots dx_n \\
&\quad \left(\because \mathbf{x} = A^{-1}(\mathbf{z} - \boldsymbol{\mu})\right),\ dz_1 \cdots dz_n = \sqrt{|\Sigma|} dx_1 \cdots dx_n) \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{x}\right) dx_1 dx_2 \cdots dx_n
\end{aligned}$$

となり、最後の式は $\mathbf{X} \sim N(\mathbf{0}, I_n)$ を示しています。ここで (19) を用いて $\mathbf{Z}$ を書きなおすと

$$\mathbf{Z} = A\mathbf{X} + \boldsymbol{\mu}$$

であることから、多変量正規分布の定義より $\mathbf{Z}$ が多変量正規分布に従うことが分かりました。

まとめますと、

$$g(z_1, \cdots, z_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{z} - \boldsymbol{\mu})\right)$$

は、Σ が正値対称行列でありさえすれば 多変量正規分布に従うことが示されました。　(証明終わり)

[*21] この証明の文脈では、この $\mathbf{X}$ は定義されたばかりなので、まだ何物かは全く分かっていません。$N(\mathbf{0}, I_n)$ に従うのですが、それはこれから確認していくことになります。