主成分分析の漸近理論

序文

主成分分析は，10 次元や 50 次元，100 次元といった多くの次元を持つ多次元データを 2 次元や 3 次元といった次元に圧縮し，データ構造を視覚的に把握する手法であり，元のデータの線形結合として合成された新たな変数の持つ意味を見いだすことによって，有益な情報抽出が可能となる．理論上は元々ある次元の数だけ主成分といういくつかの条件を持ったそれぞれのデータの線形結合を構成できるが，その線形結合が圧縮された小次元データ，つまりは計測されたデータの意味のある組み替えとなり，元の多次元データの代わりに主成分を分析することができる．具体的な応用としては，文字画像の圧縮と復元（LeCun *et al.*[16]，Bishop[12]）が挙げられる．

多次元データに内在する情報はデータのまとまりと散らばり，つまり分散で捉えられ，主成分の構成では分散をできる限り保存するように新しい変数を定義していくことになる．主成分の構成方法の背景として，データの共分散行列の固有値や固有ベクトルが鍵になる．具体的には，以下のように主成分は構成される．p 次元のデータを $\boldsymbol{X}$ とする．さらに，このデータに基づいた共分散行列の固有値を $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ として，各固有値に対応する正規化した固有ベクトルを $\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^p$ とすると，この $\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^p$ が主成分となる．データに対して主成分分析を適用することは，主成分 $\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^p$ から主成分 $\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^q$（$q \leq p$）を選択することであり，元の観測された p 次元データ $\boldsymbol{X}$ を q 次元データ $(\boldsymbol{\beta}^{1'}\boldsymbol{X},\dots,\boldsymbol{\beta}^{q'}\boldsymbol{X})'$ へと縮小できる．この主成分の構成より，得られた標本から主成分の構成を行うためには母集団の共分散行列またはその固有値と固有ベクトルの推定が必要となることがわかる．

本レポートは，母共分散行列の固有値・固有ベクトルの推定と検定を標本共分散行列の固有値・固有ベクトルを用いて行うことを目的としている．そこで，Anderson[11] を参考にし，確率変数の中心極限定理から確率変数ベクトルの中心極限定理と確率変数行列の中心極限定理を示し，標本の共分散行列の固有値・固有ベクトルを並べた確率変数行列の漸近分布から母共分散行列の固有値・固有ベクトルの推測を行う．また本レポートでは，主成分の構成としてデータの線形結合の分散の最大化を共分散行列の固有値・固有ベクトルの問題へ帰着できる様子を確認している．本レポートの構成は以下のとおりである．第 1 章では，確率論の基本概念を導入し，確率変数の中心極限定理から確率変数ベクトルと確率変数行列の中心極限定理を示している．第 2 章では，主成分の構成を共分散行列の固有値・固有ベクトル問題へ帰着できることを示す定理について述べている．第 3 章では，確率変数行列の中心極限定理を用いて，標本共分散行列の固有値・固有ベクトルの漸近分布を示している．第 4 章では，標本共分散行列の固有値・固有ベクトルの漸近理論に基づいて母固有値の推定と検定を解説している．第 5 章では，実データ解析を行っている．第 6 章では 第 1 章 から 第 4 章 に登場した定理を証明している．第 7 章では 第 6 章 の証明で用いた定義と補題を載せている．

本レポートは確率論・数理統計学の基礎を既に学んだ学生を読者として想定している．確率論・数理統計学の基礎ではなじみのない確率変数ベクトルと確率変数行列を扱っているが，確率変数から連続して読み進められるように心がけた．本レポートが読者に主成分分析への興味を促し，行列代数による確率論と数理統計学の導入となれば幸いである．

目次

第 1 章　中心極限定理 **1**
- 1.1 基本概念 1
- 1.2 確率変数ベクトルの中心極限定理 6
- 1.3 確率変数行列の中心極限定理 6

第 2 章　主成分分析 **8**

第 3 章　主成分分析の漸近理論 **9**
- 3.1 標本共分散行列の固有値と固有ベクトルの漸近分布 9

第 4 章　母固有値の漸近的な検定統計量 **11**
- 4.1 母固有値の漸近的な検定統計量 11
- 4.2 検定統計量を用いた母固有値の推定 11
- 4.3 検定統計量を用いた母固有値の検定 12

第 5 章　実データ解析 **14**
- 5.1 解析の概要と注意 14
- 5.2 両側検定 15
- 5.3 片側検定 16

第 6 章　証明 **18**

第 7 章　補填 **42**

参考文献 **48**

第 1 章　中心極限定理

基本概念として，確率変数や確率変数ベクトル，確率変数行列とこれらに関連する一般的な定義を行う．そして，確率変数と同様に確率変数ベクトルや確率変数行列で成り立つ性質を示し，確率変数の中心極限定理から確率変数ベクトルと確率変数行列の中心極限定理を示す．本レポートでは，$\mathbb{R}$ は実数の集合を表す．また，X_{ij} を (i,j) 成分とする $m \times n$ 行列を $\boldsymbol{X} = (X_{ij})_{m \times n}$ と略記し，値やベクトル $a_1, a_2, \ldots$ の列を $\{a_N\}$（$N = 1, 2, \ldots$）と略記する．

1.1　基本概念

まず，確率変数とそれに関連する一般的な定義を示す．

Definition 1.1 (確率と確率空間)**.** 組 $(\Omega, \mathscr{F}, \mathrm{P})$ が測度空間であり，$\mathrm{P}(\Omega) = 1$ であるとき，$(\Omega, \mathscr{F}, \mathrm{P})$ を確率空間という．すなわち，$(\Omega, \mathscr{F}, \mathrm{P})$ が以下を満たすとき確率空間という．

1. Ω は集合である．
2. $\mathscr{F}$ は Ω の $\sigma-$ 加法族である．すなわち，$\mathscr{F}$ が以下を満たす．
   (a) $\Omega \in \mathscr{F}$ である．
   (b) $A \in \mathscr{F}$ となるとき，$A^c \in \mathscr{F}$ である．ただし，A^c は A の補集合で $A^c = \Omega \backslash A$ である．
   (c) $A_n \in \mathscr{F}$（$n = 1, 2, \ldots$）となるとき，$\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$ である．
3. P は $\mathscr{F}$ で定義された関数で以下を満たす．
   (a) 任意の $A \in \mathscr{F}$ に対して $\mathrm{P}(A) \geq 0$ である．
   (b) （$\sigma-$ 加法性）$A_n \in \mathscr{F}$（$n = 1, 2, \ldots$）で $A_n \cap A_m = \emptyset$（$n \neq m$）となるとき，$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathrm{P}(A_n)$ となる．
   (c) $\mathrm{P}(\Omega) = 1$ である．

測度 P を確率という．また，上の条件 1，2 を満たす $(\Omega, \mathscr{F})$ を可測空間という．

Definition 1.2 (確率変数，吉田 [10])**.** $(\Omega, \mathscr{F}, \mathrm{P})$ を確率空間，$(\mathscr{X}, \mathscr{A})$ を可測空間とする．写像 $X : \Omega \to \mathscr{X}$ が可測写像である，すなわち，任意の $A \in \mathscr{A}$ に対して，$X^{-1}(A) \in \mathscr{F}$ であるとき，X は（$\mathscr{X}-$ 値）確率変数であるという．$\mathscr{X}$ が有限個または可算無限個の集合の場合，確率変数 X は離散型といい，実数の区間のような連続的な値を取る場合，確率変数 X は連続型という．

Definition 1.3 (確率分布，吉田 [10])**.** X を $(\Omega, \mathscr{F}, \mathrm{P})$ で定義された $(\mathscr{X}, \mathscr{A})-$ 値確率変数とする．X によって誘導される $(\mathscr{X}, \mathscr{A})$ 上の確率測度 P_X を X の確率分布という．P_X が $(\mathscr{X}, \mathscr{A})$ 上の確率測度 ν に等しいとき，X は確率分布 ν に従うという．

$\mathscr{X}=\mathbb{R}$, $\mathscr{A}=\boldsymbol{B}_1$（一次ボレル集合族[1]）のとき，$(\mathbb{R},\boldsymbol{B}_1)$ は可測空間となる．以降ではこの場合を考える．つまり，確率変数 X を連続型と考える．

Definition 1.4 (確率変数の分布関数と密度関数)**.** 確率変数 X に対し，X が $x\in\mathbb{R}$ 以下の値を取る確率

$$F_X(x)=\mathrm{P}(X\le x)=\mathrm{P}(\{\omega\in\Omega|X(\omega)\le x\})$$

を X の分布関数という．さらに，確率変数 X に対して，$F_X(x)$ が微分可能であるとき，

$$f_X(x)=\frac{d}{dx}F_X(x)$$

を満たす関数 $f_X(x)$ を X の密度関数という．この密度関数 $f_X(x)$ は以下を満たす関数としても定義できる．

1. 任意の $x\in\mathbb{R}$ で $f_X(x)\ge 0$ が成り立つ．
2. $\int_{-\infty}^{\infty}f_X(x)dx=1$ が成り立つ．
3. $F_X(x)=\int_{-\infty}^{x}f_X(t)dt$ が成り立つ．

Definition 1.5 (確率変数の期待値と分散)**.** 密度関数 f_X を持つ確率変数 X に対し，

$$\mathscr{E}X=\int_{-\infty}^{\infty}xf_X(x)dx$$

が存在すれば $\mathscr{E}X$ を X の期待値という．さらに，

$$\mathscr{V}X=\int_{-\infty}^{\infty}(x-\mathscr{E}X)^2f_X(x)dx$$

が存在すれば $\mathscr{V}X$ を X の分散という．

Definition 1.6 (確率変数の特性関数)**.** 密度関数 f_X を持つ確率変数 X と任意の $t\in\mathbb{R}$ に対し，

$$\phi_X(t)=\mathscr{E}e^{itX}=\int_{-\infty}^{\infty}e^{itx}f_X(x)dx$$

を X の特性関数という．ただし，$i^2=-1$ である．

本レポートでは以下の正規分布を主に仮定する．

Definition 1.7 (正規分布に従う確率変数の密度関数)**.** 密度関数

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\quad x\in\mathbb{R}$$

を持つ確率変数 X を（一変量）正規分布に従うといい，確率変数 X は $\mathscr{N}(\mu,\sigma^2)$ に従うとかく．ただし，$\mu\in\mathbb{R}$ で $\sigma>0$ であり，$\mathscr{E}X=\mu$, $\mathscr{V}X=\mathscr{E}(X-\mu)^2=\sigma^2$ である．

[1] $\mathbb{R}$ において，$-\infty\le a\le b\le\infty$ について $(a,b]$ なる形の集合の全体を $\mathscr{I}_1$ と記す．このとき，$\mathbb{R}$ において $\mathscr{I}_1$ を含む最小の $\sigma-$ 加法族が存在し，これを一次ボレル集合族という．

Theorem 1.1 (正規分布に従う確率変数の特性関数). $\mathcal{N}(\mu,\sigma^2)$ に従う確率変数 X の特性関数は，任意の $t\in\mathbb{R}$ に対して，

$$\phi_X(t)=e^{it\mu-\frac{t^2\sigma^2}{2}}$$

と求まる．ただし，$i^2=-1$ である．

確率変数ベクトルと確率変数行列は以下のように定義される．

Definition 1.8 (確率変数ベクトルと確率変数行列). 確率変数 X_{ij} を (i,j) 成分とする $m\times n$ 行列

$$\boldsymbol{X}=(X_{ij})_{m\times n}$$

を確率変数行列とよぶ．特に，確率変数行列が $p\times 1$ 行列または $1\times p$ 行列のとき p 次確率変数ベクトルとよぶ．p 次確率変数ベクトルついて，その確率分布は確率変数の可測空間を p 次元に拡張して考えられる．

確率変数ベクトルと確率変数行列を扱う上で転置行列が頻出するためその定義を導入する．

Definition 1.9 (転置行列). 行列 $\boldsymbol{X}=(X_{ij})_{m\times n}$ に対し，その列と行を入れ替えた $n\times m$ 行列

$$\boldsymbol{X}'=(X_{ji})_{n\times m}$$

を X の転置行列という．

Definition 1.10 (確率変数ベクトルの分布関数と密度関数). p 次確率変数ベクトル $\boldsymbol{X}=(X_1,\dots,X_p)'$ に対し，各 X_i（$i=1,\dots,p$）が $x_i\in\mathbb{R}$（$i=1,\dots,p$）以下の値を取る確率

$$F_{\boldsymbol{X}}(\boldsymbol{x})=\mathrm{P}(X_1\le x_1,\dots,X_p\le x_p)=\mathrm{P}(\{\omega\in\Omega|X_1(\omega)\le x_1,\dots,X_p(\omega)\le x_p\})$$

を $\boldsymbol{X}$ の（同時）分布関数という．さらに，確率変数ベクトル $\boldsymbol{X}$ に対して，$F_{\boldsymbol{X}}(\boldsymbol{x})$ が微分可能であるとき，

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{\partial^p}{\partial x_1\cdots\partial x_p}F_{\boldsymbol{X}}(\boldsymbol{x})$$

を満たす関数 $f_{\boldsymbol{X}}(\boldsymbol{x})$ を $\boldsymbol{X}$ の（同時）密度関数という．この密度関数 $f_{\boldsymbol{X}}(\boldsymbol{x})$ は以下を満たす関数としても定義できる．

1. 任意の $\boldsymbol{x}\in\mathbb{R}^p$ で $f_{\boldsymbol{X}}(\boldsymbol{x})\ge 0$ が成り立つ．
2. $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\dots,dx_p=1$ が成り立つ．
3. $F_{\boldsymbol{X}}(\boldsymbol{x})=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_p}f_{\boldsymbol{X}}(t_1,\dots,t_p)dt_1,\dots,dt_p$ が成り立つ．

また，各成分に注目した分布関数

$$F_{X_i}(x_i)=F_{\boldsymbol{X}}(X_1\le\infty,\dots,X_{i-1}\le\infty,X_i\le x_i,X_{i+1}\le\infty,\dots,X_p\le\infty)$$

を X_i の（周辺）分布関数という．その密度関数

$$f_{X_i}(x_i)=\frac{\partial}{\partial x_i}F_{X_i}(x_i)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\dots,dx_{i-1},dx_{i+1},\dots,dx_p$$

を X_i の（周辺）密度関数という．

Definition 1.11 (共分散)**.** 密度関数 $f_{\boldsymbol{X}}$ を持つ確率変数ベクトル $\boldsymbol{X} = (X_1, X_2)$ に対し，

$$\mathscr{C}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mathscr{E}X_1)(x_2 - \mathscr{E}X_2) f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_1, dx_2 = \mathscr{E}(X_1 - \mathscr{E}X_1)(X_2 - \mathscr{E}X_2)$$

が存在すれば $\mathscr{C}(X_1, X_2)$ を X_1 と X_2 の共分散という．

確率変数ベクトルと確率変数行列についても，確率変数と同様に期待値の定義がある．

Definition 1.12 (確率変数行列の期待値)**.** 確率変数行列 $\boldsymbol{X} = (X_{ij})_{m \times n}$ の期待値を

$$\mathscr{E}\boldsymbol{X} = (\mathscr{E}X_{ij})_{m \times n}$$

で定義する．

確率変数ベクトルにおいて，確率変数の分散と共分散にあたるものが以下の共分散行列である．

Definition 1.13 (共分散行列)**.** p 次確率変数ベクトル $\boldsymbol{X}$ からその期待値 $\mathscr{E}\boldsymbol{X}$ を引いた確率変数ベクトルと，その転置の積で表される確率変数行列 $(\boldsymbol{X} - \mathscr{E}\boldsymbol{X})(\boldsymbol{X} - \mathscr{E}\boldsymbol{X})'$ の期待値

$$\mathscr{E}(\boldsymbol{X} - \mathscr{E}\boldsymbol{X})(\boldsymbol{X} - \mathscr{E}\boldsymbol{X})' = \left(\mathscr{E}(X_i - \mathscr{E}X_i)(X_j - \mathscr{E}X_j)\right)_{p \times p} \tag{1.1}$$

を $\boldsymbol{X}$ の共分散行列とよぶ．(1.1) の i 番目の対角成分 $\mathscr{E}(X_i - \mathscr{E}X_i)^2$ は確率変数 X_i の分散であり，i 行 j 列目の非対角成分 $\mathscr{E}(X_i - \mathscr{E}X_i)(X_j - \mathscr{E}X_j)$ は確率変数 X_i と X_j（$i \neq j$）の共分散である．

Definition 1.14 (確率変数ベクトルの特性関数)**.** 密度関数 $f_{\boldsymbol{X}}$ を持つ確率変数ベクトル $\boldsymbol{X}$ に対し．任意の $\boldsymbol{t} \in \mathbb{R}^p$ について

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathscr{E}e^{i\boldsymbol{t}'\boldsymbol{X}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\boldsymbol{t}'\boldsymbol{X}} f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_1, \ldots, dx_p$$

を $\boldsymbol{X}$ の特性関数という．ただし，$i^2 = -1$ を満たす．

Definition 1.15 (独立性)**.** 密度関数 $f_{\boldsymbol{X}}$ を持つ確率変数ベクトルを $\boldsymbol{X} = (X_1, \ldots, X_p)'$ として，X_i の密度関数を $f_{X_i}(x_i)$ としたとき，任意の $x_1, \ldots, x_p \in \mathbb{R}$ に対して

$$f_{\boldsymbol{X}}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_{X_i}(x_i)$$

が成り立つならば確率変数 $X_1, \ldots, X_p$ は独立であるという．独立性と同値な性質の内いくつかは Lemma 7.8 に示す．また，独立な確率変数たちが同じ確率分布に従うとき，確率変数たちは独立同一であるという．確率変数ベクトルの独立性は，各変数間の独立性で定義される．

確率変数ベクトルを扱う上で，要素の期待値を調べる際は以下の結合モーメントがよく用いられる．

Definition 1.16 (結合モーメント)**.** 密度関数 $f_{\boldsymbol{X}}(\boldsymbol{x})$ を持つ p 次確率変数ベクトル $\boldsymbol{X} = (X_1, \ldots, X_p)'$ の結合モーメントを

$$\mathscr{E}X_1^{h_1} \cdots X_p^{h_p} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{h_1} \cdots x_p^{h_p} f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_1, \ldots, dx_p$$

で定義する．ただし，$h_1, \ldots, h_p$ は 0 以上の整数である．

確率変数ベクトルの特性関数と結合モーメントには，以下のような関係式が示される．

Theorem 1.2 (特性関数と結合モーメントの関係式)**.** 任意の定数ベクトル $\boldsymbol{t} = (t_1, ..., t_p)'$ に対する p次確率変数ベクトル $\boldsymbol{X} = (X_1, ..., X_p)'$ の特性関数 $\phi(\boldsymbol{t})$ と $\boldsymbol{X}$ の結合モーメントの間で

$$\begin{aligned}\mathscr{E} X_h X_j &= \frac{1}{i^2} \frac{\partial^2 \phi(\boldsymbol{t})}{\partial t_h \partial t_j}\bigg|_{\boldsymbol{t}=\boldsymbol{0}}, \\ \mathscr{E} X_h X_j X_k X_l &= \frac{1}{i^4} \frac{\partial^4 \phi(\boldsymbol{t})}{\partial t_h \partial t_j \partial t_k \partial t_l}\bigg|_{\boldsymbol{t}=\boldsymbol{0}}\end{aligned}$$

が成り立つ．ただし，$X_1, ..., X_p$ は独立で，$i^2 = -1$ を満たし，$1 \le h, j, k, l \le p$ である．

確率変数ベクトルの性質を証明する上で，以下の多次元の密度関数の変数変換が用いられる．

Theorem 1.3 (確率変数ベクトルに対する変数変換の公式，清水 [4])**.** p 次確率変数ベクトル $\boldsymbol{X}$ の密度関数を $f_{\boldsymbol{X}}$ とする．$\boldsymbol{\phi}$ を一階の偏導関数を持つ $\mathbb{R}^p$ から $\mathbb{R}^p$ への全単射な写像とする．このとき，$\boldsymbol{Y} = \boldsymbol{\phi}(\boldsymbol{X})$ の密度関数 $f_{\boldsymbol{Y}}$ は,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}\left(\boldsymbol{\phi}^{-1}(\boldsymbol{y})\right)\left|\det \boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y})\right| = \frac{1}{\left|\det \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})\right|} f_{\boldsymbol{X}}\left(\boldsymbol{\phi}^{-1}(\boldsymbol{y})\right)$$

と求まる．ここで，$\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})$ は関数 $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)$ に対するヤコビ行列であり，$\nabla = \left(\partial/\partial y_1, ..., \partial/\partial y_p\right)$ に対して,

$$\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y}) = \nabla \boldsymbol{\phi}(\boldsymbol{y}) = \begin{pmatrix} \frac{\partial \phi_1}{\partial y_1} & \cdots & \frac{\partial \phi_1}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_p}{\partial y_1} & \cdots & \frac{\partial \phi_p}{\partial y_p} \end{pmatrix}$$

で与えられる．特に，$\boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y}) = \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})^{-1}$ であるから，$\det \boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y}) = \left(\det \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})\right)^{-1}$ が成り立つ．

Definition 1.17 (正規分布に従う確率変数ベクトルの密度関数)**.** 密度関数

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{1}{2}p} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right], \quad \boldsymbol{x} \in \mathbb{R}^p$$

を持つ p 次確率変数ベクトル $\boldsymbol{X}$ を（多変量）正規分布に従うといい, 確率変数ベクトル $\boldsymbol{X}$ は $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとかく．ただし，$\boldsymbol{\mu} \in \mathbb{R}^p$ で $\boldsymbol{\Sigma}$ の要素は正であり，$|\boldsymbol{\Sigma}|$ は $\boldsymbol{\Sigma}$ の行列式を表し，$\boldsymbol{\Sigma}^{-1}$ は $\boldsymbol{\Sigma}$ の逆行列を表す．このとき，$\mathscr{E}\boldsymbol{X} = \boldsymbol{\mu}$，$\mathscr{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \boldsymbol{\Sigma}$ である．

Theorem 1.4 (多変量正規分布に従う確率変数ベクトルの特性関数)**.** $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従う p 次確率変数ベクトル $\boldsymbol{X}$ の特性関数は，任意の $\boldsymbol{t} \in \mathbb{R}^p$ に対して,

$$\phi(\boldsymbol{t}) = \mathscr{E} e^{i \boldsymbol{t}' \boldsymbol{X}} = e^{i \boldsymbol{t}' \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{t}' \boldsymbol{\Sigma} \boldsymbol{t}}$$

と求まる．

1.2　確率変数ベクトルの中心極限定理

標本数を大きくした際の挙動である中心極限定理を示すために，確率変数や確率変数ベクトル，確率変数行列の確率収束と分布収束を定義する．

Definition 1.18 (確率収束，野田，宮岡 [8]). 確率変数 $\{X_N\}$（$N = 1,2,\ldots$）が全ての $\varepsilon > 0$ に対して，$\lim_{N\to\infty} \mathrm{P}(|X_N - X| < \varepsilon) = 1$（または $\lim_{N\to\infty} \mathrm{P}(|X_N - X| \geq \varepsilon) = 0$）を満たすとき，$\{X_N\}$ は確率変数 X に確率収束するといい，$X_N \xrightarrow{p} X$ と表す．これは，解析的な意味での X_N の X への収束を表しているのではなく，実数列 $\{\mathrm{P}(|X_N - X| < \varepsilon)\}$ の 1 への収束を意味している．$\boldsymbol{X}_N$ が確率変数ベクトルならば，絶対値のかわりにノルム $||\boldsymbol{X}_N - \boldsymbol{X}||$ を用いて，$||\boldsymbol{X}_N - \boldsymbol{X}|| \xrightarrow{p} 0$ ならば $\boldsymbol{X}_N \xrightarrow{p} \boldsymbol{X}$ とする．確率変数行列の確率収束は，要素の二乗和やフロベニウスノルム（要素の二乗和の平方），または各要素の確率収束を考える．

Definition 1.19 (分布収束と漸近分布，野田，宮岡 [8]). $\boldsymbol{X}$，$\{\boldsymbol{X}_N\}$（$N = 1,2,\ldots$）を確率変数ベクトルとし，$F_{\boldsymbol{X}}$，$\{F_{\boldsymbol{X}_N}\}$ をそれぞれ対応した分布関数とする．$F_{\boldsymbol{X}}$ が連続な点 $\boldsymbol{x}$ 全てで

$$\lim_{N\to\infty} F_{\boldsymbol{X}_N}(\boldsymbol{x}) = F_{\boldsymbol{X}}(\boldsymbol{x})$$

ならば，$\{\boldsymbol{X}_N\}$ は確率変数ベクトル $\boldsymbol{X}$ に分布収束するといい，$\boldsymbol{X}_N \xrightarrow{d} \boldsymbol{X}$ で表す．$\boldsymbol{X}$ の分布を漸近分布，または極限分布といい，$\boldsymbol{X}_N \xrightarrow{d} \boldsymbol{X}$ のとき $\boldsymbol{X}_N$ は $\boldsymbol{X}$ の分布に漸近的に従うという．確率変数の分布収束は確率変数ベクトルと同様に一次元において定義され，確率変数行列の分布収束は行列の要素を並べたベクトルの分布収束で定義される．

ここで，確率変数の中心極限定理を導入する．

Theorem 1.5 (確率変数の中心極限定理, Cramér[14]). 確率変数 $X_1,\ldots,X_N$ を独立同一に平均 $\mathscr{E}X_i = \mu$，分散 $\mathscr{V}X_i = \sigma^2$ を持つ確率分布に従うとする．すると，その標本平均 $\overline{X}_N = (1/N)\sum_{i=1}^{N} X_i$ を標準化した確率変数

$$Z_N = \frac{\sqrt{N}(\overline{X}_N - \mu)}{\sigma}$$

は漸近的に $\mathscr{N}(0,1)$ に従う．

これまでに示された定理から，以下のような確率変数ベクトルの中心極限定理が示される．

Theorem 1.6 (確率変数ベクトルの中心極限定理). p 次確率変数ベクトル $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_N$ を独立同一に平均 $\mathscr{E}\boldsymbol{Y}_i = \boldsymbol{\nu}$，共分散行列 $\mathscr{E}(\boldsymbol{Y}_i - \boldsymbol{\nu})(\boldsymbol{Y}_i - \boldsymbol{\nu})' = \boldsymbol{T}$ を持つ確率分布に従うとする．すると，確率変数ベクトル $(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{\nu})$ の $N \to \infty$ とした極限分布は $\mathscr{N}(\boldsymbol{0},\boldsymbol{T})$ となる．

1.3　確率変数行列の中心極限定理

確率変数行列の漸近分布は，その要素をベクトルに並べた確率変数ベクトルの漸近分布として考えられる．従って，確率変数ベクトルの中心極限定理から確率変数行列の中心極限定理が以下のように示される．

Theorem 1.7 (確率変数行列の中心極限定理)**.** p 次確率変数ベクトル $\boldsymbol{X}_1,\dots,\boldsymbol{X}_N$ を独立同一に $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ に従うとする．$\boldsymbol{A}=\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\overline{\boldsymbol{X}}_N)(\boldsymbol{X}_\alpha-\overline{\boldsymbol{X}}_N)'$ とする．ただし，$\overline{\boldsymbol{X}}=(1/N)\sum_{i=1}^{N}\boldsymbol{X}_i$ である．すると，$\boldsymbol{B}=(1/\sqrt{n})[\boldsymbol{A}-n\boldsymbol{\Sigma}]$ の極限分布は要素それぞれの期待値が 0 で，行列の (i,j) 成分と (k,l) 成分の共分散が

$$\mathscr{E}b_{ij}b_{kl}=\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$$

である多変量正規分布となる．ただし，b_{ij} は $\boldsymbol{B}$ の (i,j) 成分で，σ_{ij} は $\boldsymbol{\Sigma}$ の (i,j) 成分，$n=N-1$ とする．

第2章　主成分分析

あらゆる科学や産業における統計解析においては，多くの次元を持つ多次元データの獲得が可能となり，多次元データから情報を抜き出す必要性が高まっている．多次元データをそのまま解析すると計算量が多く解析に時間が掛かり，また変数間の相関を分析することが難しい．そこで，主成分分析を行うと，情報を保存したまま多次元データの次元を縮小した新しい小次元データが得られ，さらには次元間の相関を分析したりデータを性質に分けて視覚化することができる．新しいデータは元の多次元データの線形結合で表され，その係数ベクトルが主成分と呼ばれる．主成分分析の実用的な説明は内田 [2] が詳しい．序文で示したような，主成分の構成が多次元データの共分散行列の固有値・固有ベクトルにより行われることは以下の定理として証明ができる．

主成分の構成

Theorem 2.1 (主成分の構成)**.** $\boldsymbol{X}$ を p 次確率変数ベクトルとする．ただし，$\mathscr{E}\boldsymbol{X}=\boldsymbol{0}$，$\mathscr{E}\boldsymbol{X}\boldsymbol{X}'=\boldsymbol{\Sigma}$ とする．すると，ある直交変換

$$\boldsymbol{U}=\boldsymbol{B}'\boldsymbol{X}$$

が存在する．ただし，$\boldsymbol{U}$ の共分散行列は $\mathscr{E}\boldsymbol{U}\boldsymbol{U}'=\boldsymbol{\Lambda}$ で，

$$\boldsymbol{\Lambda}=\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

を満たす．また，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ は $\boldsymbol{\Sigma}$ の固有方程式 $|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|=0$ の根である．また，$\boldsymbol{B}$ の r 列目 $\boldsymbol{\beta}^{(r)}$ は $(\boldsymbol{\Sigma}-\lambda_r\boldsymbol{I})\boldsymbol{\beta}^{(r)}=\boldsymbol{0}$ を満たし，この $\boldsymbol{\beta}^{(r)}$ を第 r 主成分と呼ぶ．そして，$\boldsymbol{U}$ の r 番目の要素は $U_r=\boldsymbol{\beta}^{(r)\prime}\boldsymbol{X}$ で，$U_1,\ldots,U_{r-1}$ と無相関な正規化線形結合の中で最も大きな分散を持つ．

Theorem 2.1 において，$\boldsymbol{U}$ は $\boldsymbol{X}$ から $U_1,\ldots,U_p$ への直交変換をまとめていることがわかる．また，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ が $\boldsymbol{\Sigma}$ の固有方程式 $|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|=0$ の根であること，$\boldsymbol{\beta}^{(r)}$ が方程式 $(\boldsymbol{\Sigma}-\lambda_r\boldsymbol{I})\boldsymbol{\beta}^{(r)}=\boldsymbol{0}$ を満たすことから，λ_r と $\boldsymbol{\beta}^{(r)}$ は $\boldsymbol{\Sigma}$ の固有値と固有ベクトルであることがわかる．従って，主成分の構成が共分散行列の固有値・固有ベクトルの問題として帰着されることがわかった．

第3章　主成分分析の漸近理論

標本共分散行列の固有値と固有ベクトルの漸近分布を調べる．

3.1　標本共分散行列の固有値と固有ベクトルの漸近分布

母集団に正規性を仮定する場合，以下のようなウィッシャート分布を導入すると確率変数行列の漸近分布を考えやすくなる．

Definition 3.1 (ウィッシャート分布)**.** p 次確率変数ベクトル $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$（$N>p$）が独立同一に $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ に従うとする．すると，$\boldsymbol{A}=\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'$ の密度関数は，$\boldsymbol{A}$ が正定値（Definition 7.1）のとき，

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}}}{2^{\frac{1}{2}pn}\pi^{p(p-1)/4}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]},\quad (n=N-1) \tag{3.1}$$

である．ただし，$\Gamma(\alpha)$ はガンマ関数であり，

$$\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}\,dx,\quad (\alpha>0)$$

と表される．また，$\boldsymbol{A}$ が正定値でないなら 0 で，$n<p$ のとき，$\boldsymbol{A}$ は密度を持たない．(3.1) を $w(\boldsymbol{A}|\boldsymbol{\Sigma},n)$ と書き，関連する分布，つまり上記の密度関数を持つ分布をウィッシャート分布といって $\mathcal{W}(\boldsymbol{\Sigma},n)$ と書く．このように $\mathcal{W}(\boldsymbol{\Sigma},n)$ に従う行列 $\boldsymbol{A}$ をウィッシャート行列とよぶ．

ここで，得られた標本の共分散行列である標本共分散行列を改めて定義する．本レポートでは用いないため省略するが，この標本共分散行列が母共分散行列に対して最尤推定量になることも調べられる．Anderson[11] を参照されたい．

Definition 3.2 (標本共分散行列)**.** p 次確率変数ベクトルの標本 $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$ に対して，

$$\boldsymbol{S}=\frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'$$

を標本共分散行列と呼ぶ．

後の定理や証明において，確率変数の漸近的な分散や共分散，確率変数ベクトルの漸近的な共分散行列を求める．そのための記号を用意する．

Definition 3.3 (漸近的な分散と共分散)**.** 確率変数 $\{X_N\}$（$N=1,2,\ldots$）の $N\to\infty$ とした漸近的な分散を

$$\mathcal{AV}X_N=\lim_{N\to\infty}\mathcal{E}(X_N-\mathcal{E}X_N)^2$$

と表し，確率変数 $\{X_N\}$，$\{Y_N\}$（$N = 1, 2, \ldots$）の $N \to \infty$ とした漸近的な共分散を

$$\mathscr{A}\mathscr{C}(X_N, Y_N) = \lim_{N\to\infty} \mathscr{E}(X_N - \mathscr{E}X_N)(Y_N - \mathscr{E}Y_N)$$

と表す．また，確率変数ベクトル $\{\boldsymbol{X}_N\}$（$N = 1, 2, \ldots$）の $N \to \infty$ とした漸近的な共分散行列を

$$\mathscr{A}\mathscr{C}\boldsymbol{X}_N = \lim_{N\to\infty} \mathscr{E}(\boldsymbol{X}_N - \mathscr{E}\boldsymbol{X}_N)(\boldsymbol{X}_N - \mathscr{E}\boldsymbol{X}_N)'$$

と表し，確率変数ベクトル $\{\boldsymbol{X}_N\}$，$\{\boldsymbol{Y}_N\}$（$N = 1, 2, \ldots$）の $N \to \infty$ とした漸近的な共分散行列を

$$\mathscr{A}\mathscr{C}(\boldsymbol{X}_N, \boldsymbol{Y}_N) = \lim_{N\to\infty} \mathscr{E}(\boldsymbol{X}_N - \mathscr{E}\boldsymbol{X}_N)(\boldsymbol{Y}_N - \mathscr{E}\boldsymbol{Y}_N)'$$

と表す．

以下，母集団の共分散行列 $\boldsymbol{\Sigma}$ の固有値は全て異なるとして，標本共分散行列の固有値と固有ベクトルの漸近分布を調べる．

p 次確率変数ベクトルの標本 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ に対して, 標本共分散行列を $\boldsymbol{S}$ とする．$n\boldsymbol{S}$ が $\mathscr{W}(\boldsymbol{\Sigma}, n)$ に従うとする．対角行列 $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $\boldsymbol{L} = \mathrm{diag}(l_1, \ldots, l_p)$ と直交行列 $B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$, $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p)$ を

$$\boldsymbol{\Sigma} = B\boldsymbol{\Lambda}B', \quad \boldsymbol{S} = \boldsymbol{BLB}'$$

で定義する．これは $\boldsymbol{\Sigma}$ と $\boldsymbol{S}$ の対角化であり，それぞれの直交行列を一意に定めるため $\lambda_1 > \lambda_2 > \cdots > \lambda_p$，$l_1 \geq l_2 \geq \cdots \geq l_p$，$\beta_{1i} \geq 0$，$b_{1i} \geq 0$，$i = 1, \ldots, p$ とする．行列 $\boldsymbol{G} = \sqrt{n}(\boldsymbol{B} - B)$ と対角行列 $\boldsymbol{D} = \sqrt{n}(\boldsymbol{L} - \boldsymbol{\Lambda})$ を定義する．すると，母共分散行列と標本共分散行列の固有値と固有ベクトルの間に以下のような定理が成り立つ．

Theorem 3.1 (標本共分散行列の固有値と固有ベクトルの漸近分布)**.** $\boldsymbol{D}$ と $\boldsymbol{G}$ の極限分布は独立に正規分布に従い, $\boldsymbol{D}$ の対角成分たちは独立となる．さらに, $\boldsymbol{D}$ の対角成分 d_i は極限分布 $\mathscr{N}(0, 2\lambda_i^2)$ に従う．また，$\boldsymbol{G} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_p)$ の極限分布において，$\boldsymbol{g}_i$ の共分散行列は,

$$\mathscr{A}\mathscr{C}\boldsymbol{g}_i = \sum_{k=1, k\neq i}^{p} \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \boldsymbol{\beta}_k \boldsymbol{\beta}_k'$$

と表せる．また，$\boldsymbol{G} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_p)$ の極限分布において，$\boldsymbol{g}_i$ と $\boldsymbol{g}_j$ の共分散行列は,

$$\mathscr{A}\mathscr{C}(\boldsymbol{g}_i, \boldsymbol{g}_j) = -\frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \boldsymbol{\beta}_j \boldsymbol{\beta}_i', \quad (i \neq j)$$

と表せる．

第4章　母固有値の漸近的な検定統計量

主成分の構成が共分散行列の固有値と固有ベクトルの問題に帰着されたことから，得られた標本によりその固有値・固有ベクトルを推定・検定したいと考えることは自然である．標本共分散行列の固有値と固有ベクトルの漸近分布を用いて，母共分散行列の固有値と固有ベクトルの漸近的な検定統計量を構成する．以下，第 3 章 と同じ仮定の下，記号も同じものを用いる．

4.1　母固有値の漸近的な検定統計量

母固有値の漸近的な検定統計量の導出に用いるため，標本固有値が母固有値の一致推定量であることを示す．

Definition 4.1 (一致推定量)**.** d 次定数ベクトル $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)'$ の一致推定量 $\boldsymbol{t}_N$ とは，$t_{iN} \xrightarrow{p} \theta_i$（$i = 1, \dots, d$）を満たすような確率変数ベクトル $\boldsymbol{t}_N = (t_{1N}, \dots, t_{dN})'$（$N = 1, 2, \dots$）のことである．確率変数の一致推定量は $d = 1$ の場合を考え，確率変数行列についてはその要素の確率収束で考える．

Theorem 4.1 (標本固有値の一致性)**.** 確率変数 l_i は λ_i の一致推定量である．

標本固有値が母固有値の一致推定量であることと第 3 章から，以下のような母固有値の漸近的な検定統計量が導出される．

Theorem 4.2 (母共分散行列の固有値の漸近的な検定統計量)**.**

$$\sqrt{n}\frac{l_i - \lambda_i}{\sqrt{2}l_i}$$

の極限分布は $\mathcal{N}(0,1)$ となる．

4.2　検定統計量を用いた母固有値の推定

Theorem 4.2 で示した母固有値の漸近的な検定統計量を用いて，母固有値の推定を行っていく．検定統計量を

$$\sqrt{\frac{n}{2}}\frac{l_i - \lambda_i}{\lambda_i} \tag{4.1}$$

とする．ただし，ここで l_i は標本共分散行列 $\boldsymbol{S}$ の固有値で，標本数 N に対して $n = N - 1$ である．すると，(4.1) は漸近的に $\mathcal{N}(0,1)$ に従うため，区間

$$-z(\alpha) \leq \sqrt{\frac{n}{2}}\frac{l_i - \lambda_i}{\lambda_i} \leq z(\alpha) \tag{4.2}$$

は信頼係数 $1 - \alpha$ を持つ λ_i の信頼区間を与える[1]．ただし，$z(\alpha)$ は $\mathcal{N}(0,1)$ の上側 $100(\alpha/2)\%$ 点，すなわち $\mathcal{N}(0,1)$ に従う確率変数 Z について $P_Z^H\{Z > z(\alpha)\} = \alpha/2$ を満たす点である[2]．λ_i の不等式

[1]漸近分布により区間が求められるから，厳密には近似的な信頼区間である．

[2]P_Z^H は帰無仮説 H の下で確率変数 Z により誘導される確率を表す．

になるように (4.2) に簡単な式変形を行うと

$$\frac{l_i}{1+\sqrt{2/n}z(\alpha)} \le \lambda_i \le \frac{l_i}{1-\sqrt{2/n}z(\alpha)} \tag{4.3}$$

となり，信頼係数 $1-\alpha$ の λ_i の信頼区間を与える．

信頼係数 $1-\alpha$ は $\sqrt{2/n}z(\alpha) < 1$ を満たすように十分大きな値を取るべきであることに注意しなければならない．信頼係数 $1-\alpha$ を大きく取れば $100(\alpha/2)\%$ 点 $z(\alpha)$ は大きくなる．このような信頼係数 $1-\alpha$ の設定を $\sqrt{2/n}z(\alpha) < 1$ を満たしつつ行えということである．この条件は，(4.3) の右辺の分母が $\sqrt{2/n}z(\alpha) = 1$ のとき 0 になってしまうため，それを避けつつ信頼係数 $1-\alpha$ の設定を行いたいという条件である．ただし，今は漸近分布に基づいて信頼区間を考えているため，$\sqrt{2/n}$ が非常に小さく，それに対して $z(\alpha)$ が大きすぎなければ条件 $\sqrt{2/n}z(\alpha) < 1$ を満たすため，十分大きな $1-\alpha$ を取ることは出来る．

4.3 検定統計量を用いた母固有値の検定

Theorem 4.2 で示した母固有値の漸近的な検定統計量を用いて，母固有値の仮説検定を行う．

まず，i 番目の母固有値についての帰無仮説 $H : \lambda_i = \lambda_i^0$，対立仮説 $K : \lambda_i \neq \lambda_i^0$ の有意水準 α の両側検定を考える．いま，検定統計量を (4.1) とする．すると，(4.1) は漸近的に $\mathcal{N}(0,1)$ に従うため，帰無仮説 $H : \lambda_i = \lambda_i^0$ における両側検定の（漸近的な）受容域は

$$-z(\alpha) \le \sqrt{\frac{n}{2}}\frac{l_i - \lambda_i^0}{\lambda_i^0} \le z(\alpha) \tag{4.4}$$

である．$\lambda_i = \lambda_i^0$ という仮定の元で，実現値 l_i に対する $\sqrt{n/2}(l_i - \lambda_i^0)/\lambda_i^0$ の値が (4.4) の区間に含まれなければ帰無仮説 $H : \lambda_i = \lambda_i^0$ は棄却され，対立仮説 $K : \lambda_i \neq \lambda_i^0$ が受容される．また，得られた実現値 l_i に対する $\sqrt{n/2}(l_i - \lambda_i^0)/\lambda_i^0$ の値が (4.4) の区間に含まれれば帰無仮説 $H : \lambda_i = \lambda_i^0$ は受容される．(4.4) を λ_i^0 について解くと受容域

$$\left[\frac{l_i}{1+\sqrt{2/n}z(\alpha)}, \frac{l_i}{1-\sqrt{2/n}z(\alpha)}\right] \tag{4.5}$$

を得る．従って，実現値 l_i より得られる区間 (4.5) に λ_i^0 が含まれなければ帰無仮説 $H : \lambda_i = \lambda_i^0$ が棄却され，対立仮説 $K : \lambda_i \neq \lambda_i^0$ が受容される．また，λ_i^0 が区間 (4.5) に含まれれば帰無仮説 $H : \lambda_i = \lambda_i^0$ は受容される．

有意水準 α は，4.2 の信頼係数についての注意と同様に，十分小さな値を取るべきであることに注意しなければならない．

さらに，$\mathcal{N}(0,1)$ に従う確率変数 Z に対して，実現値 l_i についての確率

$$P_Z^H\left\{Z < -\sqrt{\frac{n}{2}}\frac{l_i - \lambda_i^0}{\lambda_i^0}\right\} + P_Z^H\left\{Z > \sqrt{\frac{n}{2}}\frac{l_i - \lambda_i^0}{\lambda_i^0}\right\} \tag{4.6}$$

はこの両側検定の p 値[3]または確率値とよばれる．i 番目の母固有値について帰無仮説 $H' : \lambda_i \le \lambda_i^0$，

[3] 一般に，有意水準 α の仮説検定問題において，棄却域（例えば $\mathbb{R}$ から (4.5) を除いた区間）を $C_T(\alpha)$ とするとき，$T = t$ について $p(t) = \inf\{\alpha \mid t \in C_T(\alpha)\}$，つまり与えられた標本に対して，帰無仮説を棄却できる最小の有意水準を（t による）p 値という．（赤平 [1]）

対立仮説 K' : $\lambda_i > \lambda_i^0$ を立てた有意水準 α の片側検定の p 値は

$$P_Z^{H'}\left\{Z > \sqrt{\frac{n}{2}}\frac{l_i - \lambda_i^0}{\lambda_i^0}\right\} \tag{4.7}$$

である．

また，検定統計量 (4.1) の代わりに，$\sqrt{n}(\log l_i - \log \lambda_i)$ を使うことも出来る．

Theorem 4.3. $\sqrt{n}(\log l_i - \log \lambda_i)$ の極限分布は $\mathcal{N}(0,2)$ である．

Theorem 4.3 より，$\mathcal{N}(0,2)$ における $100(\alpha/2)\%$ 点を $\mathcal{N}(0,1)$ のときと同様に求めれば，帰無仮説 H : $\lambda_i = \lambda_i^0$ に対する有意水準 α の検定を，実現値 l_i に基づいてこれまでと同様に行うことが出来る．また，受容域を λ_i について指数を取るなどして解けば，信頼係数 $1-\alpha$ の λ_i の信頼区間が得られる．

ここでは紹介にとどめるが，母固有ベクトル $\boldsymbol{\beta}^{(i)}$ の要素の統計的推測は，$\boldsymbol{b}^{(i)}$ を平均 $\boldsymbol{\beta}^{(i)}$ で，共分散行列が $\mathscr{A}\mathscr{C}\boldsymbol{g}^{(i)}$ に $1/n$ を掛けた（非正則）共分散行列の正規分布に漸近的に従うということに基づいて進めることが出来る．

第5章　実データ解析

5.1　解析の概要と注意

R 言語の scar パッケージ（Chen, Samworth[13]）にある decathlon データセット（データ数 614, 次元数 10）を解析する．第 i 主成分に対応する未知な母固有値 λ_i に対して，帰無仮説と対立仮説を設定し，標本であるデータセットから標本固有値 l_i を導出して有意水準 α の仮説検定を行う．

ここで，第 4 章で与えられた検定統計量 (4.1) は母集団に正規性を仮定していたことに注意しなければならない．実際，以下の fig. 5.1 より decathlon データがおおむね正規分布に近似できそうなことが確認出来る．

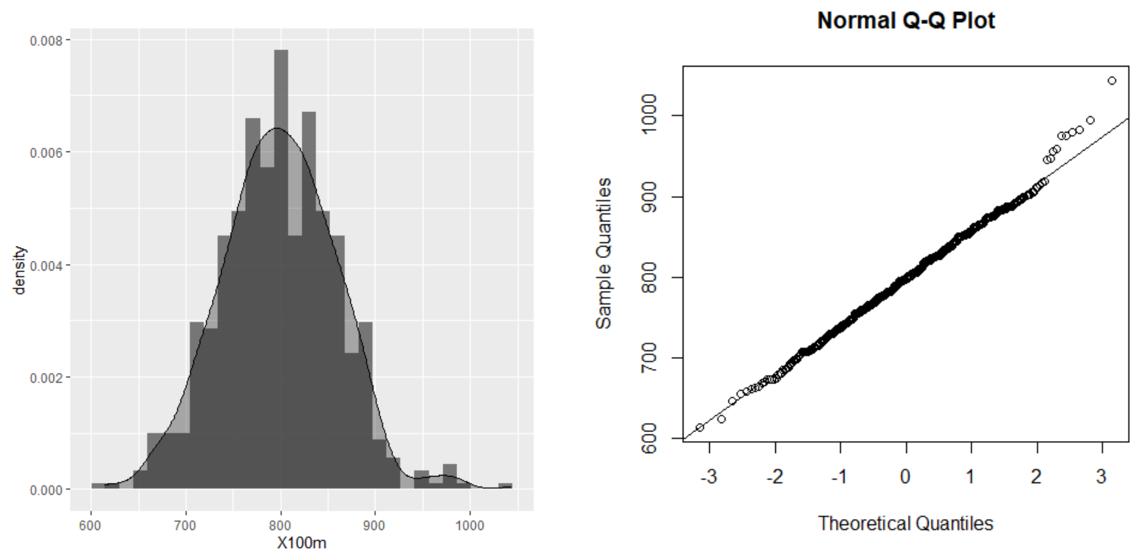

fig 5.1: 左は decathlon データの X100m 変数のヒストグラムと密度曲線で右は decathlon データの X100m 変数の qq プロット

また，fig. 5.2 の左の図を見ると，検定統計量 (4.1) によるヒストグラムが標準正規分布のグラフに近いことが確認出来る．よって，検定統計量 (4.1) は標準正規分布におおむね近似できそうである．fig. 5.2 の右の図を見ると，データセットから抽出する標本数が増えれば 95% 信頼区間が得られていることが確認出来る．よって，正規分布に従う母集団に対して検定統計量 (4.1) を用いた漸近的な信頼区間 (4.2) の構成が行えることが確認出来た．

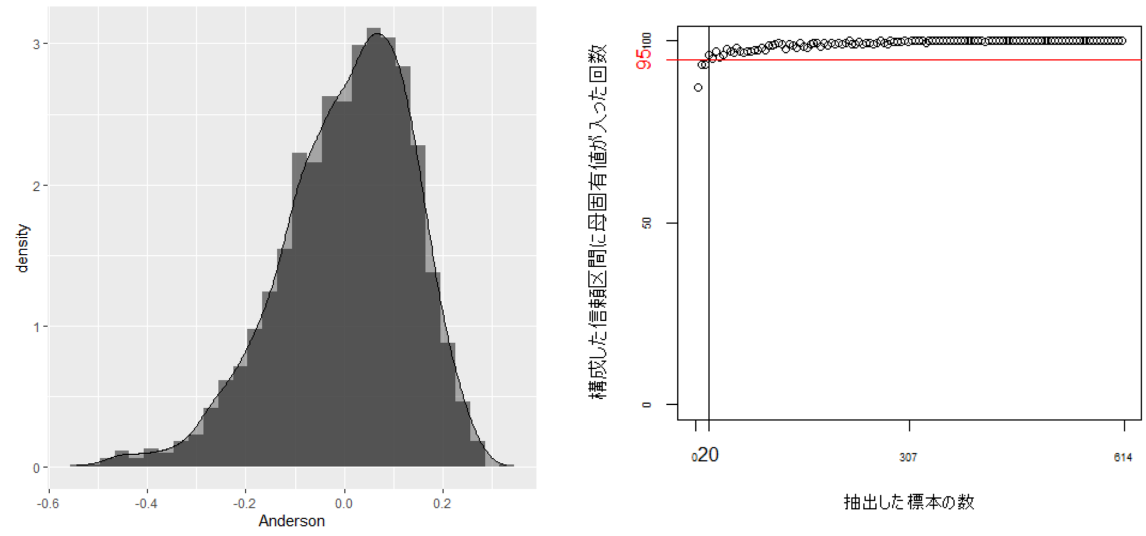

fig 5.2: 左は decathlon データを検定統計量 (4.1) に変換したヒストグラムと密度曲線で右は decathlon データの母固有値の 95% 信頼区間の構成に必要な標本数のプロット

5.2 両側検定

データセットを標本として考え，母集団の未知な母固有値 λ_i について有意水準 α の両側仮説検定を行う．ただし，帰無仮説 $H:\lambda_i=\lambda_i^0$，対立仮説 $K:\lambda_i\neq\lambda_i^0$ である．データセットから標本固有値 l_i を導出し，第 4 章で与えられる信頼係数 $1-\alpha$ を持つ λ_i の以下の信頼区間

$$\frac{l_i}{1+\sqrt{2/n}z(\alpha)}\leq\lambda_i\leq\frac{l_i}{1-\sqrt{2/n}z(\alpha)} \tag{5.1}$$

の下限（左辺）と上限（右辺）を求める．これを各主成分で求めて信頼区間を図示し，区間が帰無仮説 $H:\lambda_i=\lambda_i^0$ を含むのかを確認する．

λ_i^0 を固有値として非常に小さい値に設定したとき，$i=1,2,\ldots$ に対してこの仮説検定を帰無仮説が棄却されなくなるまで徐々に大きくする．すると，帰無仮説が棄却されるような固有値に対応する主成分が，有意水準 α の下では有意に λ_i^0 でない固有値を持つ．つまり元のデータの分散を説明するに有用な主成分であると主張できる．例えば，decathlon データにおいて 10 個の主成分を構成出来るが，この仮説検定によって帰無仮説 $\lambda_5=\lambda_5^0$ が棄却出来ないのであれば，設定している有意水準 α の下では第 1，第 2，第 3，第 4 主成分が主成分分析として利用するのに有効で，第 5 主成分以降は削減できると判断できる[1]．

fig. 5.3 は，仮説検定の有意水準を 0.01 として，帰無仮説 $H:\lambda_i=\lambda_i^0$ を $\lambda_i^0=5000$ とした場合の各標本固有値における信頼区間 (5.1) の様子である．

[1]ただし，「棄却されない」ことが意味することは「（材料が不十分で）棄却されない」ということであり，そのことを主張として使うことは本来注意が必要である．ここではその結果を一つの指針として適用してみる程度で考えれば良い．

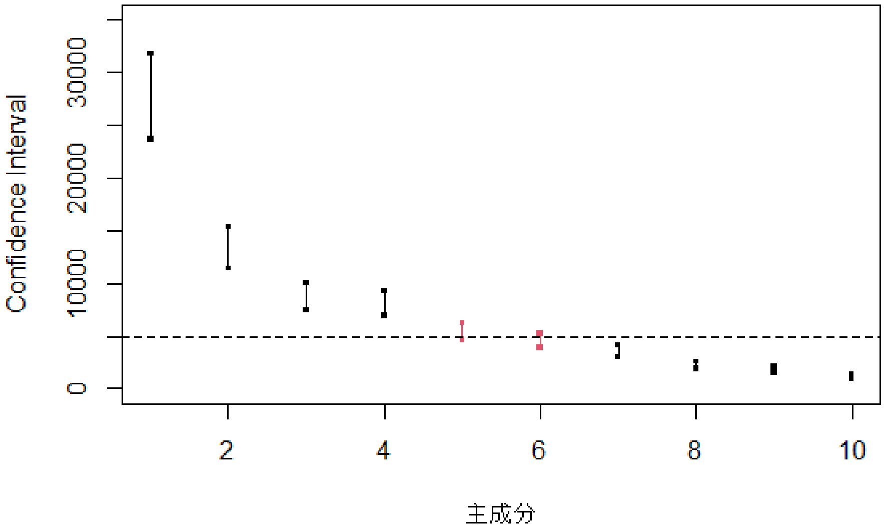

fig 5.3: 有意水準 0.01 の信頼区間 (5.1) の様子

fig. 5.3 を見ると，帰無仮説 $H : \lambda_5 = 5000$ が棄却出来ていないため，設定している有意水準 0.01 の下では第 1，第 2，第 3，第 4 主成分が主成分分析として利用するのに有効で，第 5 主成分以降は削減できると判断できる．

また，帰無仮説 $H : \lambda_5 = 5000$ の下での各標本固有値 $l_1 \geq \cdots \geq l_{10}$ における p 値 (4.6) を table 5.1 にまとめた．標本固有値が小さくなるにつれて p 値が小さくなっている．つまり，有意水準が大きければ小さい標本固有値について帰無仮説 $H : \lambda_i = 5000$ を棄却出来なくなっていくことが確認出来る．

table 5.1: 各標本固有値における p 値 (4.6)

	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
p 値	1.0e+00	1.0e+00	1.0e+00	1.0e+00	8.4e-01	2.9e-02	6.4e-08	2.5e-23	3.0e-29	5.6e-43

5.3 片側検定

データセットを標本として考え，母集団の未知な母固有値 λ_i について有意水準 α の片側仮説検定を行う．ただし，帰無仮説 $H' : \lambda_i \leq \lambda_i^0$，対立仮説 $K' : \lambda_i > \lambda_i^0$ である．データセットから標本固有値を導出し，第 4 章で与えられる信頼係数 $1-\alpha$ を持つ λ_i の以下の信頼区間

$$\frac{l_i}{1+\sqrt{2/n}z(\alpha)} \leq \lambda_i \tag{5.2}$$

の下限（左辺）を求める．これを各主成分で求めて信頼区間を図示し，区間が帰無仮説 $H' : \lambda_i \leq \lambda_i^0$ と重なるかどうか確認する．

fig. 5.4 は，仮説検定の有意水準を 0.01 として，帰無仮説 $H' : \lambda_i \leq \lambda_i^0$ を $\lambda_i^0 = 5000$ とした場合の各標本固有値における信頼区間 (5.2) の様子である．

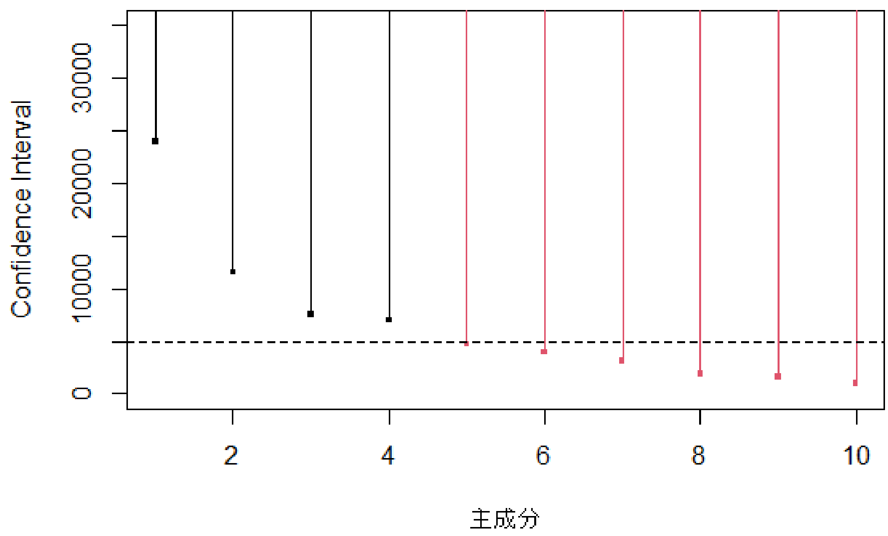

fig 5.4: 有意水準 0.01 の信頼区間 (5.2) の様子

fig. 5.4 を見ると，帰無仮説 $H' : \lambda_5 \leq 5000$ が棄却出来ていないため，設定している有意水準 0.01 の下では第 1，第 2，第 3，第 4 主成分が主成分分析として利用するのに有効で，第 5 主成分以降は削減できると判断できる．

また，帰無仮説 $H' : \lambda_5 \leq 5000$ の下での各標本固有値 $l_1 \geq \cdots \geq l_{10}$ における p 値 (4.7) を table 5.2 にまとめた．table 5.2 を見ると，標本固有値が小さくなるにつれて p 値が小さくなっている．つまり，有意水準が大きければ小さい標本固有値について帰無仮説 $H' : \lambda_i \leq 5000$ を棄却出来なくなっていくことが確認出来る．

table 5.2: 各標本固有値における p 値 (4.7)

	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
p 値	1.0e+00	1.0e+00	1.0e+00	1.0e+00	8.7e-01	2.9e-02	6.9e-08	2.3e-23	3.0e-29	5.8e-43

第6章　証明

期待値や分散について，注意が必要な場合を除いてその範囲を示さない．注意が必要な場合は [] でその範囲を示す．

Proof. 第 1 章 Theorem 1.1(p.3)（野田，宮岡 [8]）

X の標準化を $Y=(X-\mu)/\sigma$ とすると，$\mathscr{E}Y=(\mathscr{E}X-\mu)/\sigma=0$，$\mathscr{V}Y=\mathscr{E}Y^2=\mathscr{E}(X-\mu)^2/\sigma^2=1$ より，Y は $\mathscr{N}(0,1)$ に従う．また，$Y=(X-\mu)/\sigma$ を変形すると，$X=\sigma Y+\mu$ であるから，任意の $t\in\mathbb{R}$ に対し X の特性関数を $\phi_X(t)$，Y の特性関数を $\phi_Y(t)$ とすると，Lemma 7.5 に注意して

$$\phi_X(t)=\mathscr{E}e^{itX}=\mathscr{E}e^{it(\sigma Y+\mu)}=e^{it\mu}\mathscr{E}e^{i(t\sigma)Y}=e^{it\mu}\phi_Y(t\sigma) \tag{6.1}$$

と求まる．ここで，

$$\begin{aligned}\phi_Y(t)&=\int_{-\infty}^{\infty}e^{ity}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy\\&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(y^2-2ity)\right]dy\\&=e^{-t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(y-it)^2\right]dy\\&=e^{-t^2/2}\end{aligned}$$

である．三番目の等号では平方完成 $-(1/2)(y^2-2ity)=(1/2)\left\{(y-it)^2-(it)^2\right\}=-(1/2)(y-it)^2-(1/2)t^2$ を用いて，最後の等号では積分の被積分関数が $\mathscr{N}(it,1)$ に従う確率変数の密度関数であることから，その全積分は 1 となることを用いた．従って，(6.1) より

$$\phi_X(t)=e^{it\mu}e^{-(t\sigma)^2/2}=e^{it\mu-\frac{t^2\sigma^2}{2}}$$

を得る． □

Proof. 第 1 章 Theorem 1.5(p.6)（藤澤 [9]）

$\overline{X}_N$ の平均は $\mathscr{E}\overline{X}_N=(1/N)\sum_{i=1}^N\mathscr{E}X_i=\mu$ と求まり，分散は $\mathscr{V}\overline{X}_N=\mathscr{V}(1/N)(X_1+\cdots+X_N)=(1/N^2)(\mathscr{V}X_1+\cdots+\mathscr{V}X_N)=\sigma^2/N$ と求まる．X_i の標準化を $Y_i=(X_i-\mu)/\sigma$ とおき，その密度関数を f_{Y_i}，特性関数を任意の $t\in\mathbb{R}$ に対して $\phi_{Y_i}(t)$ とおくと，$t=0$ のとき，

$$\phi_{Y_i}(0)=\mathscr{E}e^{itY_i}\Big|_{t=0}=\mathscr{E}1=1$$

を得る．さらに $\phi_{Y_i}(t)$ を t で一回偏微分して $t=0$ を代入すると，

$$\begin{aligned}
\frac{\partial \phi_{Y_i}(0)}{\partial t} &= \frac{\partial}{\partial t}\int_{-\infty}^{\infty} e^{ity} f_{Y_i}(y)dy\bigg|_{t=0} \\
&= \int_{-\infty}^{\infty} (iy)e^{ity} f_{Y_i}(y)dy\bigg|_{t=0} \\
&= i\int_{-\infty}^{\infty} y f_{Y_i}(y)dy \\
&= i\mathscr{E}Y_i \\
&= i\cdot 0 \\
&= 0
\end{aligned}$$

を得る．さらに $\phi_{Y_i}(t)$ を t で二回偏微分して $t=0$ を代入すると，

$$\begin{aligned}
\frac{\partial^2 \phi_{Y_i}(0)}{\partial t^2} &= \frac{\partial^2}{\partial t^2}\int_{-\infty}^{\infty} e^{ity} f_{Y_i}(y)dy\bigg|_{t=0} \\
&= \int_{-\infty}^{\infty} (iy)^2 e^{ity} f_{Y_i}(y)dy\bigg|_{t=0} \\
&= -\int_{-\infty}^{\infty} y^2 f_{Y_i}(y)dy \\
&= -\mathscr{E}Y_i^2 \\
&= -1
\end{aligned}$$

を得る．ここで，$\overline{X}_N$ の標準化

$$Z_N = \frac{\sqrt{N}(\overline{X}_N - \mu)}{\sigma}$$

は

$$Z_N = \frac{1/N\sum_{i=1}^N X_i - \mu}{\sigma/\sqrt{N}} = \frac{1/N\sum_{i=1}^N(\sigma Y_i + \mu) - \mu}{\sigma/\sqrt{N}} = \frac{\sigma/N\sum_{i=1}^N Y_i + \mu - \mu}{\sigma/\sqrt{N}} = \frac{\sum_{i=1}^N Y_i}{\sqrt{N}}$$

とかけることから，任意の $t\in\mathbb{R}$ に対して Z_N の特性関数 $\phi_{Z_N}(t)$ は

$$\begin{aligned}
\phi_{Z_N}(t) &= \mathscr{E}e^{itZ_N} \\
&= \mathscr{E}e^{it\frac{\sum_{i=1}^N Y_i}{\sqrt{N}}} \\
&= \mathscr{E}e^{i(t/\sqrt{N})Y_1}\cdots\mathscr{E}e^{i(t/\sqrt{N})Y_N} \\
&= \left\{\phi_{Y_i}(t/\sqrt{N})\right\}^N \\
&= \left\{\phi_{Y_i}(0) + \frac{\partial}{\partial t}\phi_{Y_i}(0)(t/\sqrt{N}) + \frac{\partial^2}{\partial t^2}\phi_{Y_i}(0)\frac{(t/\sqrt{N})^2}{2} + o\left(\frac{t^2}{N}\right)\right\}^N \\
&= \left\{1 + (-t^2/2N) + o\left(\frac{t^2}{N}\right)\right\}^N
\end{aligned}$$

と求まる．ただし，三番目の等号では X_i たちの独立性より Y_i たちが独立であることを用いており，四番目の等号では X_i たちの同一性より Y_i たちが同一であることを用いている．そして，五番目の等号では $\phi_{Y_i}(t/\sqrt{N})$ のマクローリン展開を用いた．すると，$x\in\mathbb{R}$ に対する自然対数の定義

$$e^x = \lim_{N\to\infty}\left(1+\frac{x}{N}\right)^N$$

より，$N\to\infty$ のとき，

$$\phi_{Z_N}(t)\to e^{-t^2/N}$$

を得る．これは, Theorem 1.1 より, $\mathcal{N}(0,1)$ に従う確率変数の特性関数である．従って, Lemma 7.11 より，$\overline{X}_N=(1/N)\sum_{i=1}^N X_i$ の標準化 Z_N は $\mathcal{N}(0,1)$ に分布収束する． □

Proof. 第 1 章 Theorem 1.3(p.5)

$\boldsymbol{Y}$ の分布関数を考える．いま，$A=\prod_{i=1}^p(-\infty,X_i]$ とすると，

$$\begin{aligned}F_{\boldsymbol{Y}}(\boldsymbol{X})&=\mathrm{P}\left(\boldsymbol{\phi}(\boldsymbol{X})\in A\right)\\&=\mathrm{P}\left(\boldsymbol{X}\in\boldsymbol{\phi}^{-1}(A)\right)\\&=\int_{\boldsymbol{\phi}^{-1}(A)}f_{\boldsymbol{X}}(\boldsymbol{z})d\boldsymbol{z}\\&=\int_A f_{\boldsymbol{X}}\left(\boldsymbol{\phi}^{-1}(\boldsymbol{y})\right)\left|\det\boldsymbol{J}_{\boldsymbol{\phi}^{-1}(\boldsymbol{y})}\right|d\boldsymbol{y}\end{aligned}$$

が成り立つ．最後の等号では，$\boldsymbol{y}=\boldsymbol{\phi}(\boldsymbol{z})$ と置換することで $\boldsymbol{\phi}\left(\boldsymbol{\phi}^{-1}(A)\right)=A$ が成り立つことを用いて，さらに $d\boldsymbol{z}=\left|\det\boldsymbol{J}_{\boldsymbol{\phi}^{-1}(\boldsymbol{y})}\right|d\boldsymbol{y}$ が成り立つことを用いた．従って，分布関数を得た．また，Lemma 7.7 より，

$$\boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y})=\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})^{-1}$$

であるから，$\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})\boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y})=\boldsymbol{I}_p$（$p$ 次単位行列）が成り立ち，両辺の行列式を取ると

$$\det\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{y})\cdot\det\boldsymbol{J}_{\boldsymbol{\phi}^{-1}}(\boldsymbol{y})=1$$

となる． □

Proof. 第 1 章 Theorem 1.4(p.5)

$\boldsymbol{\Sigma}$ が正定値行列かつ正則行列であること，$\boldsymbol{I}$ が正定値行列であることに注意すると，Lemma 7.6 から，

$$\boldsymbol{G\Sigma G}'=\boldsymbol{I} \tag{6.2}$$

を満たす正則行列 $\boldsymbol{G}$ が存在する．両辺の逆行列を考えると，左辺は $(\boldsymbol{G\Sigma G}')^{-1}=\boldsymbol{G}'^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^{-1}$，右辺は $\boldsymbol{I}$ と等しくなる．$\boldsymbol{G}^{-1}=\boldsymbol{C}$ とおくと，正則行列の逆行列は正則行列であるから，$\boldsymbol{C}$ は正則行列で，(6.2) は

$$\boldsymbol{C}'\boldsymbol{\Sigma}^{-1}\boldsymbol{C}=\boldsymbol{I} \tag{6.3}$$

と書き換えられる．(6.3) を式変形すると，

$$\boldsymbol{\Sigma}^{-1}=\boldsymbol{C}'^{-1}\boldsymbol{C}^{-1}=(\boldsymbol{CC}')^{-1} \tag{6.4}$$

が成り立つ．ここで，

$$\boldsymbol{Y}=\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

とおく．すると，$\boldsymbol{Y}$ の同時密度関数 $f_{\boldsymbol{Y}}(\boldsymbol{y})$ に Theorem 1.3 を適用すると，$\boldsymbol{X}=\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{\mu}$ と $|\det \boldsymbol{J}_{\Delta}(\boldsymbol{y})|=|\det \boldsymbol{\Sigma}|^{1/2}$ より，

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}|\det \boldsymbol{\Sigma}|^{\frac{1}{2}} \\
&= \frac{1}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\boldsymbol{C}\boldsymbol{y}+\boldsymbol{\mu}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}\boldsymbol{y}+\boldsymbol{\mu}-\boldsymbol{\mu})} \\
&= \frac{1}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\boldsymbol{y}'\boldsymbol{C}')\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}\boldsymbol{y})} \\
&= \frac{1}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\boldsymbol{y}'\boldsymbol{y})} \\
&= \prod_{i=1}^{p} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{y_i^2}{2}}
\end{aligned}$$

となる．従って，$\boldsymbol{Y}$ は $\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$ に従い，さらに同時密度関数が周辺密度関数の積で表せたことから，Lemma 7.8 より，$\boldsymbol{Y}$ の各要素 Y_i（$i=1,\ldots,p$）が独立であることがわかる．また，その密度関数の形から Y_i たちが $\mathcal{N}(0,1)$ に従うことがわかる．ここで，$\boldsymbol{Y}$ の要素が独立であることに注意すると，確率変数ベクトル $\boldsymbol{Y}$ の特性関数 ψ は，任意の $\boldsymbol{u}\in\mathbb{R}^p$ に対して，

$$\psi(\boldsymbol{u})=\mathscr{E}e^{i\boldsymbol{u}'\boldsymbol{Y}}=\mathscr{E}\exp\left[i\sum_{j=1}^{p}u_jY_j\right]=\mathscr{E}\prod_{j=1}^{p}\exp[iu_jY_j]=\prod_{j=1}^{p}\mathscr{E}e^{iu_jY_j}=\prod_{j=1}^{p}e^{-\frac{1}{2}u_j^2}=e^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}} \tag{6.5}$$

と求まる．従って，$\boldsymbol{X}$ の特性関数 ϕ は，任意の $\boldsymbol{t}\in\mathbb{R}^p$ に対して

$$\phi(\boldsymbol{t})=\mathscr{E}e^{i\boldsymbol{t}'\boldsymbol{X}}=\mathscr{E}e^{i\boldsymbol{t}'(\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{\mu})}=e^{i\boldsymbol{t}'\boldsymbol{\mu}}\mathscr{E}e^{i\boldsymbol{t}'\boldsymbol{C}\boldsymbol{Y}}=e^{i\boldsymbol{t}'\boldsymbol{\mu}}e^{-\frac{1}{2}(\boldsymbol{t}'\boldsymbol{C})(\boldsymbol{t}'\boldsymbol{C})'}=e^{i\boldsymbol{t}'\boldsymbol{\mu}}e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{C}\boldsymbol{C}'\boldsymbol{t}}=e^{i\boldsymbol{t}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}$$

となる．ただし，四番目の等号では (6.5) に対して任意のベクトル $\boldsymbol{u}$ を $\boldsymbol{u}'=\boldsymbol{t}'\boldsymbol{C}$ として $\boldsymbol{Y}$ の特性関数を考え，最後の等号では (6.4) の両辺の逆行列を取った等号 $\boldsymbol{C}\boldsymbol{C}'=\boldsymbol{\Sigma}$ を代入している． □

Proof. 第 1 章 Theorem 1.2(p.5)

$\boldsymbol{X}$ の密度関数を $f_{\boldsymbol{X}}(\boldsymbol{x})$ とすると，$\boldsymbol{X}$ の特性関数は任意の定数ベクトル $\boldsymbol{t}=(t_1,\ldots,t_p)'$ に対して

$$\phi(\boldsymbol{t})=\mathscr{E}e^{i\boldsymbol{t}'\boldsymbol{X}}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p \tag{6.6}$$

となる．(6.6) を偏微分と積分の交換が可能であるとして，t_h（$1\le h\le p$）で偏微分すると，

$$\begin{aligned}
\frac{\partial\phi(\boldsymbol{t})}{\partial t_h} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial t_h}e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial\{i(t_1x_1+\cdots+t_px_p)\}}{\partial t_h}e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(ix_h)e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p
\end{aligned}$$

と求まる．さらに t_j（$1\le j\le p$）で偏微分すると，

$$\begin{aligned}
\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_h\partial t_j} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(ix_h)\frac{\partial}{\partial t_j}e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(ix_h)\frac{\partial\{i(t_1x_1+\cdots+t_px_p)\}}{\partial t_j}e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(i^2x_hx_j)e^{i(t_1x_1+\cdots+t_px_p)}f_{\boldsymbol{X}}(\boldsymbol{x})dx_1,\ldots,dx_p
\end{aligned}$$

と求まる．結果に対し，$\boldsymbol{t}=\boldsymbol{0}$ を代入すると，

$$\begin{aligned}
\left.\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_h \partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(i^2 x_h x_j) f_{\boldsymbol{X}}(\boldsymbol{x})\,dx_1,\ldots,dx_p \\
&= i^2\int_{x_j=-\infty}^{\infty}\int_{x_h=-\infty}^{\infty}(x_h x_j)\int_{x_p=-\infty}^{\infty}\cdots\int_{x_1=-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x})\,dx_1,\ldots,dx_p\,dx_h\,dx_j \\
&= i^2\int_{x_j=-\infty}^{\infty}\int_{x_h=-\infty}^{\infty}(x_h x_j) f_{X_h,X_j}(x_h,x_j)\,dx_h\,dx_j \\
&= i^2\mathscr{E}X_h X_j
\end{aligned}$$

となる．ただし，三番目の等号では同時密度関数と周辺密度関数の関係を用いており，$f_{X_h,X_j}(x_h,x_j)$ は X_h と X_j の（同時）密度関数である．従って，

$$\mathscr{E}X_h X_j = \frac{1}{i^2}\left.\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_h \partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}}$$

が成り立つ．

$$\mathscr{E}X_h X_j X_k X_l = \frac{1}{i^4}\left.\frac{\partial^4\phi(\boldsymbol{t})}{\partial t_h \partial t_j \partial t_k \partial t_l}\right|_{\boldsymbol{t}=\boldsymbol{0}}$$

についても，同様の操作を繰り返すことで示せる． □

Proof. 第 1 章 Theorem 1.6(p.6)

確率変数ベクトル $(1/\sqrt{N})\sum_{i=1}^{N}(\mathbf{Y}_i-\boldsymbol{\nu})$ の特性関数を，

$$\phi_N(\boldsymbol{t},u)=\mathscr{E}\exp\left[iu\boldsymbol{t}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{Y}_i-\boldsymbol{\nu})\right] \tag{6.7}$$

とする．ただし，u は任意の定数で，$\boldsymbol{t}$ は p 次元の任意の定数ベクトルである．通常の特性関数ではこの定数 u は表れないが，この特性関数を確率変数に対する特性関数として考えるときに任意の定数とするために必要であるから，$\boldsymbol{t}$ とともに表れている．固定した $\boldsymbol{t}$ に対して，(6.7) は

$$\phi_N(\boldsymbol{t},u)=\mathscr{E}\exp\left[iu\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{t}'\mathbf{Y}_i-\boldsymbol{t}'\boldsymbol{\nu})\right]=\mathscr{E}\exp\left[iu\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)\right]$$

と書き換えられるから，$\phi_N(\boldsymbol{t},u)$ は確率変数 $(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)$ の特性関数と考えられる．

ここで，確率変数ベクトルに対する期待値の線形性を用いて，$\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i$ の期待値と分散を求める．期待値は

$$\mathscr{E}\left[\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i\right]=\boldsymbol{t}'\mathscr{E}\mathbf{Y}_i-\mathscr{E}[\boldsymbol{t}'\mathscr{E}\mathbf{Y}_i]=\boldsymbol{t}'\boldsymbol{\nu}-\boldsymbol{t}'\boldsymbol{\nu}=0$$

となり，分散は

$$\begin{aligned}
\mathscr{V}\left[\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i\right] &= \mathscr{E}\left[(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)^2\right] \\
&= \mathscr{E}\left[(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)'\right] \\
&= \mathscr{E}\left[(\boldsymbol{t}'\mathbf{Y}_i-\mathscr{E}\boldsymbol{t}'\mathbf{Y}_i)(\mathbf{Y}_i'\boldsymbol{t}-\mathscr{E}\mathbf{Y}_i'\boldsymbol{t})\right] \\
&= \mathscr{E}\left[\boldsymbol{t}'(\mathbf{Y}_i-\mathscr{E}\mathbf{Y}_i)(\mathbf{Y}_i'-\mathscr{E}\mathbf{Y}_i')\boldsymbol{t}\right] \\
&= \boldsymbol{t}'\mathscr{E}\left[(\mathbf{Y}_i-\mathscr{E}\mathbf{Y}_i)(\mathbf{Y}_i'-\mathscr{E}\mathbf{Y}_i')\right]\boldsymbol{t} \\
&= \boldsymbol{t}'\mathscr{E}\left[(\mathbf{Y}_i-\mathscr{E}\mathbf{Y}_i)(\mathbf{Y}_i-\mathscr{E}\mathbf{Y}_i)'\right]\boldsymbol{t} \\
&= \boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}
\end{aligned}$$

となる．ここで，$\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i$ の標本平均 $(1/N)\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)$ の標準化を考えると，$\mathscr{E}\left[\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i\right] = 0$，$\mathscr{V}\left[\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i\right] = \boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}$，Theorem 1.5 より，$(1/N)\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)$ の標準化

$$\frac{(1/N)\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)}{\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}/N}} \tag{6.8}$$

は漸近的に $\mathscr{N}(0,1)$ に従う．$\mathscr{N}(0,1)$ に従うような確率変数を Z とすると，Z に分布収束するような確率変数 Z_N に対して $\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$ を掛けた確率変数 $\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}Z_N$ は，Lemma 7.10 より $\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}Z$ に分布収束するため，(6.8) に $\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$ を掛けた

$$\frac{(1/N)\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)}{(1/\sqrt{N})} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)$$

は $\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}Z$ に分布収束する．さらに，正規分布の線形変換の性質より，$\sqrt{\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}Z$ は $\mathscr{N}(0,\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t})$ に従う．以上より，$(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)$ は漸近的に $\mathscr{N}(0,\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t})$ に従う．

Theorem 1.1 より，$\mathscr{N}(0,\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t})$ に従う確率変数の特性関数は任意の u と固定した $\boldsymbol{t}$ に対し

$$e^{-\frac{1}{2}u^2\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$$

であるから，Lemma 7.11 より，確率変数 $(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{t}'\boldsymbol{Y}_i - \mathscr{E}\boldsymbol{t}'\boldsymbol{Y}_i)$ の特性関数 $\phi_N(\boldsymbol{t},u)$ は

$$\lim_{N\to\infty}\phi_N(\boldsymbol{t},u) = e^{-\frac{1}{2}u^2\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}} \tag{6.9}$$

を満たす．ただし，Lemma 7.11 の条件の確認として，$\boldsymbol{t} = \boldsymbol{0}$ の場合に $e^{-(1/2)u^2\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$ が連続か確認が必要だが，直感的にも

$$\lim_{\boldsymbol{t}\to\boldsymbol{0}} e^{-\frac{1}{2}u^2\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}} = e^{-\frac{1}{2}u^2\boldsymbol{0}'\boldsymbol{T}\boldsymbol{0}} = 1$$

であり問題ない．$\phi_N(\boldsymbol{t},u)$ が確率変数ベクトル $(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{\nu})$ の特性関数であったことを思い出すと，全ての $\boldsymbol{t}$ に対して，$u = 1$ とすると

$$\lim_{N\to\infty}\phi_N(\boldsymbol{t},1) = \lim_{N\to\infty}\mathscr{E}\exp\left[i\boldsymbol{t}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{\nu})\right] = e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}} \tag{6.10}$$

が得られる．ただし，二番目の等号は (6.9) より導かれる．特性関数 (6.10) は，Theorem 1.4 より，$\mathscr{N}(\boldsymbol{0},\boldsymbol{T})$ に従う確率変数ベクトルの特性関数であるから，Lemma 7.12 より，$(1/\sqrt{N})\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{\nu})$ の極限分布は $\mathscr{N}(\boldsymbol{0}.\boldsymbol{T})$ に従う．ただし，Lemma 7.11 の条件の確認として，$e^{-(1/2)\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$ は $\boldsymbol{t} = \boldsymbol{0}$ で連続であるから，$\boldsymbol{t} = \boldsymbol{0}$ のいくつかの近傍では収束は一様である． □

Proof. 第 1 章 Theorem 1.7(p.7)

まず，$\boldsymbol{Z}_1 = (Z_{11},\ldots,Z_{p1})',\ldots,\boldsymbol{Z}_N = (Z_{1N},\ldots,Z_{pN})'$ を独立に $\mathscr{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ に従う確率ベクトルとすると，Lemma 7.14 より，$\boldsymbol{A}$ は $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ と漸近的に同様の分布となる．$\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ の要素を

$$\boldsymbol{Y}_\alpha = \begin{pmatrix} Z_{1\alpha}^2 \\ Z_{1\alpha}Z_{2\alpha} \\ \vdots \\ Z_{2\alpha}^2 \\ \vdots \\ Z_{p\alpha}^2 \end{pmatrix} \tag{6.11}$$

のようにベクトルに整理する．確率変数ベクトル $\boldsymbol{Y}_\alpha$ に対し Theorem 1.6 を用いるならば，$\boldsymbol{Y}_\alpha$ の期待値 $\mathscr{E}\boldsymbol{Y}_\alpha$ と共分散行列 $\mathscr{E}(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)'$ を求めなければならない．確率変数ベクトルや確率変数行列に対する期待値の定義を考えると，$\boldsymbol{Y}_\alpha$ の要素が $\boldsymbol{Z}_\alpha$ の要素の積で構成されていることから，$\boldsymbol{Y}_\alpha$ の期待値や共分散行列は $\boldsymbol{Z}_\alpha$ の結合モーメントから導き出せる．実際，$1 \le i,j,k,l \le p$ に対して，

$$\boldsymbol{Y}_\alpha = \begin{pmatrix} Z_{1\alpha}^2 \\ Z_{1\alpha}Z_{2\alpha} \\ \vdots \\ Z_{i\alpha}^2 \\ \vdots \\ Z_{i\alpha}Z_{j\alpha} \\ \vdots \\ Z_{k\alpha}^2 \\ \vdots \\ Z_{k\alpha}Z_{l\alpha} \\ \vdots \\ Z_{p\alpha}^2 \end{pmatrix}$$

であるから，$(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)'$ は

$$(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)(\boldsymbol{Y}_\alpha - \mathscr{E}\boldsymbol{Y}_\alpha)' = \begin{pmatrix} Z_{1\alpha}^2 - \mathscr{E}Z_{1\alpha}^2 \\ Z_{1\alpha}Z_{2\alpha} - \mathscr{E}Z_{1\alpha}Z_{2\alpha} \\ \vdots \\ Z_{i\alpha}Z_{1\alpha} - \mathscr{E}Z_{i\alpha}Z_{1\alpha} \\ \vdots \\ Z_{i\alpha}^2 - \mathscr{E}Z_{i\alpha}^2 \\ \vdots \\ Z_{i\alpha}Z_{j\alpha} - \mathscr{E}Z_{i\alpha}Z_{j\alpha} \\ \vdots \\ Z_{p\alpha}^2 - \mathscr{E}Z_{p\alpha}^2 \end{pmatrix} \begin{pmatrix} Z_{1\alpha}^2 - \mathscr{E}Z_{1\alpha}^2 \\ Z_{1\alpha}Z_{2\alpha} - \mathscr{E}Z_{1\alpha}Z_{2\alpha} \\ \vdots \\ Z_{k\alpha}Z_{1\alpha} - \mathscr{E}Z_{k\alpha}Z_{1\alpha} \\ \vdots \\ Z_{k\alpha}^2 - \mathscr{E}Z_{k\alpha}^2 \\ \vdots \\ Z_{k\alpha}Z_{l\alpha} - \mathscr{E}Z_{k\alpha}Z_{l\alpha} \\ \vdots \\ Z_{p\alpha}^2 - \mathscr{E}Z_{p\alpha}^2 \end{pmatrix}', \quad (p^2 \times p^2 \text{行列}) \tag{6.12}$$

であり，この $(p(i-1)+j, p(k-1)+l)$ 成分は $(Z_{i\alpha}Z_{j\alpha} - \mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha} - \mathscr{E}Z_{k\alpha}Z_{l\alpha})$ とかける．従って，(6.12) の期待値の要素として，$\mathscr{E}(Z_{i\alpha}Z_{j\alpha} - \mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha} - \mathscr{E}Z_{k\alpha}Z_{l\alpha})$（$1 \le i,j,k,l \le p$）を計算すればよい．いま，$\boldsymbol{Z}_\alpha = (Z_{1\alpha},\dots,Z_{p\alpha})'$（$\alpha = 1,\dots,N$）が $\mathscr{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ に従っていたことから，$\mathscr{E}Z_{i\alpha} = 0$（$i = 1,\dots,p$，$\alpha = 1,\dots,N$）である．以下，$\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ とする．Theorem 1.2 を用いて $\mathscr{E}(Z_{i\alpha}Z_{j\alpha} - \mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha} - \mathscr{E}Z_{k\alpha}Z_{l\alpha})$（$1 \le i,j,k,l \le p$）を求めていく．$\boldsymbol{Z}_\alpha$ が $\mathscr{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ に従うことから，Theorem 1.4 より任意の p 次元定数ベクトル $\boldsymbol{t}$ に対して，$\boldsymbol{Z}_\alpha$ の特性関数は

$$\phi(\boldsymbol{t}) = \mathscr{E}[e^{i\boldsymbol{t}'\boldsymbol{Z}_\alpha}] = e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}} = e^{-\frac{1}{2}\sum_{a,b=1}^{p} t_a\sigma_{ab}t_b}$$

である．虚数単位と添え字が重ならないように注意すれば，

$$\frac{\partial \phi(\boldsymbol{t})}{\partial t_h} = \phi(\boldsymbol{t})\left\{-\frac{1}{2}\frac{\partial}{\partial t_h}\sum_{a,b=1}^{p} t_a\sigma_{ab}t_b\right\} = \phi(\boldsymbol{t})\left\{-\frac{1}{2}2\sum_{a=1}^{p}\sigma_{ah}t_a\right\} = \phi(\boldsymbol{t})\{-\boldsymbol{\sigma}_h\boldsymbol{t}\}$$

と

$$\begin{aligned}\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_h\partial t_j} &= \phi(\boldsymbol{t})\{\boldsymbol{\sigma}_j\boldsymbol{t}\}\{\boldsymbol{\sigma}_h\boldsymbol{t}\} + \phi(\boldsymbol{t})\left\{-\frac{\partial}{\partial t_j}(\boldsymbol{\sigma}_h\boldsymbol{t})\right\}\\ &= \phi(\boldsymbol{t})\{\boldsymbol{\sigma}_j\boldsymbol{t}\}\{\boldsymbol{\sigma}_h\boldsymbol{t}\} + \phi(\boldsymbol{t})(-\sigma_{jh})\\ &= \phi(\boldsymbol{t})\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) - \sigma_{jh}\right\}\end{aligned} \tag{6.13}$$

が求まる．ただし，$\boldsymbol{\sigma}_s = (\sigma_{1s},\ldots,\sigma_{ps})'$ とおいている．(6.13) は $\boldsymbol{t}=\boldsymbol{0}$ のとき，

$$\left.\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_h\partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}} = 1(0-\sigma_{jh}) = -\sigma_{jh}$$

で，$1/i^2 = -1$ を掛けると σ_{jh} となる．従って，

$$\mathscr{E}Z_{i\alpha}Z_{j\alpha} = \frac{1}{i^2}\left.\frac{\partial^2\phi(\boldsymbol{t})}{\partial t_i\partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}}$$

より $\mathscr{E}Z_{i\alpha}Z_{j\alpha} = \sigma_{ji}$ を得る．(6.13) をさらに t_k で偏微分すると，

$$\begin{aligned}\frac{\partial^3\phi(\boldsymbol{t})}{\partial t_h\partial t_j\partial t_k} &= \phi(\boldsymbol{t})\{-(\boldsymbol{\sigma}_k\boldsymbol{t})\}\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) - \sigma_{jh}\right\} + \phi(\boldsymbol{t})\left\{\frac{\partial}{\partial t_k}\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t})\right\}\\ &= \phi(\boldsymbol{t})\{-(\boldsymbol{\sigma}_k\boldsymbol{t})\}\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) - \sigma_{jh}\right\} + \phi(\boldsymbol{t})\left\{\sigma_{kj}(\boldsymbol{\sigma}_h\boldsymbol{t}) + \left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)\sigma_{kh}\right\}\end{aligned}$$

と求まり，さらに t_l で偏微分すると，

$$\begin{aligned}\frac{\partial^4\phi(\boldsymbol{t})}{\partial t_h\partial t_j\partial t_k\partial t_l} &= \phi(\boldsymbol{t})(-\boldsymbol{\sigma}_l\boldsymbol{t})\left\{(-\boldsymbol{\sigma}_k\boldsymbol{t})\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) + (\boldsymbol{\sigma}_k\boldsymbol{t})\sigma_{jh}\right\}\\ &\quad + \phi(\boldsymbol{t})\left\{\frac{\partial}{\partial t_l}(-\boldsymbol{\sigma}_k\boldsymbol{t})\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) + \frac{\partial}{\partial t_l}(\boldsymbol{\sigma}_k\boldsymbol{t})\sigma_{jh}\right\}\\ &\quad + \sigma_{kj}\sigma_{lh} + \sigma_{kh}\sigma_{lj}\end{aligned} \tag{6.14}$$

と求まる．(6.14) の一行目は，偏微分もなく，$\boldsymbol{t}=\boldsymbol{0}$ のとき 0 となることがわかる．二行目の第一項は $(-\boldsymbol{\sigma}_k\boldsymbol{t})\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t})$ を $(-\boldsymbol{\sigma}_k\boldsymbol{t})\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t})\right\}$ と見て関数の積の微分を考えると，

$$\frac{\partial}{\partial t_l}(-\boldsymbol{\sigma}_k\boldsymbol{t})\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t})\right\} = -\sigma_{lk}\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t}) + (-\boldsymbol{\sigma}_k\boldsymbol{t})\frac{\partial}{\partial t_l}\left\{\left(\boldsymbol{\sigma}_j\boldsymbol{t}\right)(\boldsymbol{\sigma}_h\boldsymbol{t})\right\} \tag{6.15}$$

となる．$\boldsymbol{t}=\boldsymbol{0}$ のとき，(6.15) は 0 になる．二行目の第二項は $\sigma_{lk}\sigma_{jh} + (\boldsymbol{\sigma}_k\boldsymbol{t})$ に等しく，$\boldsymbol{t}=\boldsymbol{0}$ のとき $\sigma_{lk}\sigma_{jh}$ となる．従って，$\boldsymbol{\Sigma}$ が対称行列であることに注意すると，

$$\left.\frac{\partial^4\phi(\boldsymbol{t})}{\partial t_h\partial t_j\partial t_k\partial t_l}\right|_{\boldsymbol{t}=\boldsymbol{0}} = \sigma_{lk}\sigma_{jh} + \sigma_{kj}\sigma_{lh} + \sigma_{kh}\sigma_{lj} = \sigma_{hj}\sigma_{kl} + \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk}$$

と求まる．$1/i^4 = 1$ をかけると $\sigma_{hj}\sigma_{kl} + \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk}$ になる．従って，

$$\mathscr{E}Z_{h\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha} = \frac{1}{i^4}\left.\frac{\partial^4\phi(\boldsymbol{t})}{\partial t_h\partial t_j\partial t_k\partial t_l}\right|_{\boldsymbol{t}=\boldsymbol{0}}$$

より，$\mathscr{E}Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha}=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$ を得る．以上の結果から，

$$\begin{aligned}
\mathscr{E}(Z_{i\alpha}Z_{j\alpha}-\mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha}-\mathscr{E}Z_{k\alpha}Z_{l\alpha}) &= \mathscr{E}(Z_{i\alpha}Z_{j\alpha}-\sigma_{ij})(Z_{k\alpha}Z_{l\alpha}-\sigma_{kl}) \\
&= \mathscr{E}\left[Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha}-\sigma_{kl}Z_{i\alpha}Z_{j\alpha}-\sigma_{ij}Z_{k\alpha}Z_{l\alpha}+\sigma_{ij}\sigma_{kl}\right] \\
&= \mathscr{E}Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha}-\sigma_{kl}\mathscr{E}Z_{i\alpha}Z_{j\alpha}-\sigma_{ij}\mathscr{E}Z_{k\alpha}Z_{l\alpha}+\sigma_{ij}\sigma_{kl} \\
&= \sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}-\sigma_{kl}\sigma_{ij}-\sigma_{ij}\sigma_{kl}+\sigma_{ij}\sigma_{kl} \\
&= \sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}
\end{aligned}$$

を得る．従って，$\mathscr{E}(Z_{i\alpha}Z_{j\alpha}-\mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha}-\mathscr{E}Z_{k\alpha}Z_{l\alpha})=\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$ を得る．

以上より，$\boldsymbol{Z}_\alpha$ の結合モーメントは，

$$\begin{aligned}
\mathscr{E}Z_{i\alpha}Z_{j\alpha} &= \sigma_{ij}, \\
\mathscr{E}Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha} &= \sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}, \\
\mathscr{E}(Z_{i\alpha}Z_{j\alpha}-\mathscr{E}Z_{i\alpha}Z_{j\alpha})(Z_{k\alpha}Z_{l\alpha}-\mathscr{E}Z_{k\alpha}Z_{l\alpha}) &= \sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}
\end{aligned}$$

と得られた．ただし，i,j,k,l に重複がある場合は偏微分をする際の添え字に注意して σ の添え字を対応させればよい．

$\boldsymbol{Y}_\alpha$ の期待値 $\mathscr{E}\boldsymbol{Y}_\alpha$ と共分散行列 $\mathscr{E}(\boldsymbol{Y}_\alpha-\mathscr{E}\boldsymbol{Y}_\alpha)(\boldsymbol{Y}_\alpha-\mathscr{E}\boldsymbol{Y}_\alpha)'$ にこの結果を代入すると，

$$\mathscr{E}\boldsymbol{Y}_\alpha=\begin{pmatrix}\sigma_{11}\\ \sigma_{12}\\ \vdots\\ \sigma_{ii}\\ \vdots\\ \sigma_{ij}\\ \vdots\\ \sigma_{kk}\\ \vdots\\ \sigma_{kl}\\ \vdots\\ \sigma_{pp}\end{pmatrix} \tag{6.16}$$

と

$$\mathscr{E}(\boldsymbol{Y}_\alpha-\mathscr{E}\boldsymbol{Y}_\alpha)(\boldsymbol{Y}_\alpha-\mathscr{E}\boldsymbol{Y}_\alpha)'=\begin{pmatrix}
2\sigma_{11}^2 & \cdots & 2\sigma_{1k}\sigma_{1l} & \cdots & 2\sigma_{1p}^2 \\
2\sigma_{11}\sigma_{21} & \cdots & \sigma_{1k}\sigma_{2l}+\sigma_{1l}\sigma_{2k} & \cdots & 2\sigma_{1p}\sigma_{2p} \\
\vdots & & \cdots & & \vdots \\
2\sigma_{ii}^2 & \cdots & 2\sigma_{ik}\sigma_{il} & \cdots & 2\sigma_{ip}^2 \\
\vdots & & \cdots & & \vdots \\
2\sigma_{i1}\sigma_{j1} & \cdots & \sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk} & \cdots & 2\sigma_{ip}\sigma_{jp} \\
\vdots & & \cdots & & \vdots \\
2\sigma_{p1}^2 & \cdots & 2\sigma_{pk}\sigma_{pl} & \cdots & 2\sigma_{pp}^2
\end{pmatrix}. \tag{6.17}$$

が求まる．

以下，$\boldsymbol{\nu} := \mathscr{E}\boldsymbol{Y}_\alpha$，$\boldsymbol{T} := \mathscr{E}(\boldsymbol{Y}_\alpha - \boldsymbol{\nu})(\boldsymbol{Y}_\alpha - \boldsymbol{\nu})'$ とおく．Theorem 1.6 との対応を見ると，(6.11) で定義されるベクトル $\boldsymbol{Y}_\alpha$ たちは，$\boldsymbol{\nu}$ の要素が (6.11) と同様の順序でベクトルに整理された $\boldsymbol{\Sigma}$ の要素で，$\boldsymbol{T}$ が (6.17) の場合の Theorem 1.6 の条件を満たす．つまり，Theorem 1.6 を適用すると，$(1/\sqrt{N})\sum_{\alpha=1}^N(\boldsymbol{Y}_\alpha - \boldsymbol{v})$ の $N \to \infty$ とした極限分布は $\mathscr{N}(\boldsymbol{0}, \boldsymbol{T})$ となる．以上より，$\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ をベクトルに並べた確率変数ベクトル $\boldsymbol{Y}_\alpha$ に対して，Theorem 1.6 を適用することで，$(1/\sqrt{N})\sum_{\alpha=1}^N(\boldsymbol{Y}_\alpha - \boldsymbol{v})$ の漸近分布が $\mathscr{N}(\boldsymbol{0}, \boldsymbol{T})$ であることがわかった．Theorem 1.7 の主張は，$\boldsymbol{B} = (1/\sqrt{n})[\boldsymbol{A} - n\boldsymbol{\Sigma}]$ の漸近分布であった．この漸近分布とは，$\boldsymbol{B}$ をベクトルに並べた確率変数ベクトルの漸近分布のことであり，この確率変数ベクトルとは $(1/\sqrt{n})[\boldsymbol{W} - n\boldsymbol{\nu}]$ にあたる．ただし，$\boldsymbol{W}$ は $\boldsymbol{A}$ を $\boldsymbol{Y}_\alpha$ と同様の順序で並べたベクトルであり，(6.16) より $n\boldsymbol{\nu}$ が $n\boldsymbol{\Sigma}$ の並べ替えであることはわかる．そして，Lemma 7.14 より $\boldsymbol{A}$ と $\sum_{\alpha=1}^n \boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ が漸近的に同等であったことから，$(1/\sqrt{n})[\boldsymbol{W} - n\boldsymbol{v}]$ の漸近分布は $(1/\sqrt{N})\sum_{\alpha=1}^N(\boldsymbol{Y}_\alpha - \boldsymbol{v}) = (1/\sqrt{N})\left(\sum_{\alpha=1}^N \boldsymbol{Y}_\alpha - N\boldsymbol{v}\right)$ の漸近分布と同等で $\mathscr{N}(\boldsymbol{0}, \boldsymbol{T})$ であり，この漸近分布はそれぞれの平均が 0 で，さらに共分散行列 $\boldsymbol{T}$ の要素が $\boldsymbol{B}$ の i,j 成分と k,l 成分に対応した $\boldsymbol{\Sigma}$ の要素を用いて $\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$ と表せた．従って，$\boldsymbol{B} = (1/\sqrt{n})[\boldsymbol{A} - n\boldsymbol{\Sigma}]$ の漸近分布は Theorem 1.7 の主張通りである． □

Proof. 第 2 章 Theorem 2.1(p.8)

この証明では，主成分の構成としてデータの線形結合の分散の最大化を行い，それがデータの共分散行列の固有値・固有ベクトルの問題に帰着されることを証明する．なお，主成分の構成として新しいデータ点と元のデータ点の射影誤差の最小化から同様の帰着も行える（Bishop[12]）．

p 次元確率変数ベクトル $\boldsymbol{X}$ の線形結合 $\boldsymbol{\beta}'\boldsymbol{X}$ を考え（ただし係数ベクトルは正規化 $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ を仮定する），その中でも分散が最大な線形結合が $\boldsymbol{\Sigma}$ の最大固有値 λ_1 に対応する固有ベクトル $\boldsymbol{\beta}^1$ による線形結合 $\boldsymbol{\beta}^{1'}\boldsymbol{X}$ であることを示す．さらに，線形結合の期待値 $\mathscr{E}\boldsymbol{\beta}^{1'}\boldsymbol{X}$ が固有値 λ_1 に等しいことも示す．そして，$\boldsymbol{\beta}^1$ の次に線形結合 $\boldsymbol{\beta}'\boldsymbol{X}$ の分散を最大化し，$\boldsymbol{\beta}^{1'}\boldsymbol{X}$ と無相関となるような線形結合が $\boldsymbol{\Sigma}$ の固有値で二番目に大きな λ_2 に対応する固有ベクトル $\boldsymbol{\beta}^2$ による $\boldsymbol{\beta}^{2'}\boldsymbol{X}$ であること，さらには線形結合の期待値 $\mathscr{E}\boldsymbol{\beta}^{2'}\boldsymbol{X}$ が固有値 λ_2 に等しいことも示す．この証明を帰納的に示していくことで，Theorem 2.1 の直交変換 $\boldsymbol{B}$ として上記の固有ベクトルを並べた行列 $(\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^p)$ があることを示す．

係数ベクトルとして扱うベクトル $\boldsymbol{\beta}$ を $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ とする．ただし，$\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ を満たす．すると，$\boldsymbol{\beta}'\boldsymbol{X}$ の分散は

$$\mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X} - \mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X}))^2 = \mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X})^2 = \mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X})(\boldsymbol{\beta}'\boldsymbol{X}) = \mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta} = \boldsymbol{\beta}'\mathscr{E}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} \tag{6.18}$$

と書き換えられる．(6.18) を最大化するような $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ を満たすベクトル $\boldsymbol{\beta}$ を見つけるため，等号制約のある最適化として Lemma 7.15 を適用する．ラグランジュ関数として

$$\phi = \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - 1) = \sum_{i,j}\beta_i\sigma_{ij}\beta_j - \lambda\left(\sum_i \beta_i^2 - 1\right)$$

とおく．ただし，λ はラグランジュ乗数である．Lemma 7.15 より，(6.18) を最大化するようなベクトル $\boldsymbol{\beta}$ を見つけるためには

$$\frac{\partial\phi}{\partial\boldsymbol{\beta}} = \boldsymbol{0}, \tag{6.19}$$

$$\frac{\partial\phi}{\partial\lambda} = 0 \tag{6.20}$$

を満たす $\boldsymbol{\beta}$, λ を求めれば良い．ベクトル微分の定義（Definition 7.2）に注意して，まず (6.19) の左辺を求めると

$$\frac{\partial \phi}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\left\{\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - 1)\right\} = \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}) - \lambda\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\beta})$$

と求まる．この第一項目は

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta})}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} \frac{\partial(\Sigma_{i,j}\beta_i\sigma_{ij}\beta_j)}{\partial \beta_1} \\ \vdots \\ \frac{\partial(\Sigma_{i,j}\beta_i\sigma_{ij}\beta_j)}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} \Sigma_j\sigma_{1j}\beta_j + \Sigma_i\sigma_{i1}\beta_i \\ \vdots \\ \Sigma_j\sigma_{pj}\beta_j + \Sigma_i\sigma_{ip}\beta_i \end{pmatrix} = \begin{pmatrix} 2\Sigma_j\sigma_{1j}\beta_j \\ \vdots \\ 2\Sigma_j\sigma_{pj}\beta_j \end{pmatrix} = 2\boldsymbol{\Sigma}\boldsymbol{\beta} \tag{6.21}$$

と求まり，第二項目は

$$-\lambda\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\beta}) = -\lambda\begin{pmatrix} \frac{\partial(\boldsymbol{\beta}'\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial(\boldsymbol{\beta}'\boldsymbol{\beta})}{\partial \beta_p} \end{pmatrix} = -\lambda\begin{pmatrix} \frac{\partial(\Sigma_i\beta_i^2)}{\partial \beta_1} \\ \vdots \\ \frac{\partial(\Sigma_i\beta_i^2)}{\partial \beta_p} \end{pmatrix} = -\lambda\begin{pmatrix} 2\beta_1 \\ \vdots \\ 2\beta_p \end{pmatrix} = -2\lambda\boldsymbol{\beta} \tag{6.22}$$

と求まる．(6.19) に (6.21) と (6.22) を代入すると，

$$2\boldsymbol{\Sigma}\boldsymbol{\beta} - 2\lambda\boldsymbol{\beta} = \boldsymbol{0}$$

を得る．(6.20) は等号制約と同じ式となるため，$\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ の下で $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ を最大化する $\boldsymbol{\beta}$ は $2\boldsymbol{\Sigma}\boldsymbol{\beta} - 2\lambda\boldsymbol{\beta} = \boldsymbol{0}$ を満たすことがわかった．この等式を式変形すると

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{I})\boldsymbol{\beta} = \boldsymbol{0} \tag{6.23}$$

を満たす $\boldsymbol{\beta}$ を考えれば良いことがわかる．ただし，$\boldsymbol{I}$ は p 次の単位行列を表す．$\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ の下で，(6.23) の解を得るため，$\boldsymbol{\Sigma} - \lambda\boldsymbol{I}$ を非正則とする．つまり，λ が方程式

$$|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| = 0 \tag{6.24}$$

を満たす．ただし，$|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}|$ は $\boldsymbol{\Sigma} - \lambda\boldsymbol{I}$ の行列式を表す．$\boldsymbol{\Sigma}$ や $\boldsymbol{I}$ は $p \times p$ 行列であるから，行列式の定義より $|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}|$ は λ に関する次数 p の多項式であり，(6.24) は p 個の根を持つ．これらの解を $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ とする[1]．(6.23) の両辺に左から $\boldsymbol{\beta}'$ を掛けると，

$$\begin{aligned} \boldsymbol{\beta}'(\boldsymbol{\Sigma} - \lambda\boldsymbol{I})\boldsymbol{\beta} = 0 &\Rightarrow \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\beta}'\lambda\boldsymbol{I}\boldsymbol{\beta} = 0 \\ &\Rightarrow \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} = \lambda\boldsymbol{\beta}'\boldsymbol{I}\boldsymbol{\beta} = \lambda\boldsymbol{\beta}'\boldsymbol{\beta} = \lambda \end{aligned} \tag{6.25}$$

を得る．ただし，始めの等号の右辺では，$\boldsymbol{\beta}'\boldsymbol{0} = 0$ であることを用いている．よって，$\boldsymbol{\beta}$ が (6.23) と $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ を満たすならば，(6.18) で与えられた $\boldsymbol{\beta}'\boldsymbol{X}$ の分散が λ に等しいことが示された．従って，最大の分散として最大の根 λ_1 を (6.23) で用いるべきである．$\boldsymbol{\beta}^{(1)}$ を $(\boldsymbol{\Sigma} - \lambda_1\boldsymbol{I})\boldsymbol{\beta} = \boldsymbol{0}$ の正規化された解とする．すると，$\boldsymbol{U}_1 = \boldsymbol{\beta}^{(1)'}\boldsymbol{X}$ は最大の分散を持つ正規化線形結合となる．

ここで，$\boldsymbol{U}_1 = \boldsymbol{\beta}^{(1)'}\boldsymbol{X}$ と無相関な線形結合の中で最も大きな分散を持つ正規化された線形結合 $\boldsymbol{\beta}'\boldsymbol{X}$ を見つける．$\boldsymbol{\beta}'\boldsymbol{X}$ と $\boldsymbol{U}_1$ が無相関とは，共分散が 0 であることだから，つまり

$$\mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X} - \mathscr{E}\boldsymbol{\beta}'\boldsymbol{X})(\boldsymbol{U}_1 - \mathscr{E}\boldsymbol{U}_1) = 0 \tag{6.26}$$

[1] ラグランジュ乗数であった λ は実数とは限らないため，このように大小関係が付けられるとは限らないはずである．しかし，実際は Lemma 7.16 より実数であることが証明でき，大小付けが出来る．

を満たす．左辺については，

$$\mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X}-\mathscr{E}\boldsymbol{\beta}'\boldsymbol{X})(\boldsymbol{U}_1-\mathscr{E}\boldsymbol{U}_1)=\mathscr{E}(\boldsymbol{\beta}'\boldsymbol{X}-\boldsymbol{\beta}'\mathscr{E}\boldsymbol{X})(\boldsymbol{U}_1-\boldsymbol{\beta}^{(1)'}\mathscr{E}\boldsymbol{X})=\mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{U}_1 \tag{6.27}$$

が成り立つ．よって，

$$0=\mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{U}_1=\mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(1)}=\boldsymbol{\beta}'\mathscr{E}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(1)}=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)} \tag{6.28}$$

が成り立つ．ただし，一番目の等号では (6.26) と (6.27) を用いて，二番目の等号では $\boldsymbol{U}_1$ が確率変数であることより $\boldsymbol{U}_1'=\boldsymbol{U}_1$ が成り立つことを用いた．五番目の等号は (6.23) において λ_1 とそれに対応する $\boldsymbol{\beta}^{(1)}$ を λ と $\boldsymbol{\beta}$ に代入した等号 $\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=\lambda_1\boldsymbol{\beta}^{(1)}$ から従う．この結果から，無相関である (6.26) という意味でも，ベクトル $\boldsymbol{\beta}$ と $\boldsymbol{\beta}^{(1)}$ の内積 $\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}$ が 0 になる (6.28) という意味でも，$\boldsymbol{\beta}'\boldsymbol{X}$ は $\boldsymbol{U}_1=\boldsymbol{\beta}^{(1)'}\boldsymbol{X}$ に直交することがわかった[2]．つまり，$\boldsymbol{\beta}'\boldsymbol{X}$ が $\boldsymbol{U}_1=\boldsymbol{\beta}^{(1)'}\boldsymbol{X}$ と無相関であるという条件は，$\boldsymbol{\beta}$ と $\boldsymbol{\beta}^{(1)}$ が直交するという条件と同値であることがわかった．ここで，上記の $\boldsymbol{\beta}$ について等号制約のある線形結合 $\boldsymbol{\beta}'\boldsymbol{X}$ の最適化を行うため，再び Lemma 7.15 を用いる．ラグランジュ関数として

$$\phi_2=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}-\lambda(\boldsymbol{\beta}'\boldsymbol{\beta}-1)-2\nu_1\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)} \tag{6.29}$$

とおく．ただし，λ と ν_1 はラグランジュ乗数である．(6.29) の第一項目が最大化したい $\boldsymbol{\beta}'\boldsymbol{X}$ の分散であり，第二項目と第三項目が等号制約の項になる．第二項目は $\boldsymbol{\beta}$ の正規化，第三項目が (6.28) の等号制約の項になる．(6.29) を $\boldsymbol{\beta}$ でベクトル微分すると

$$\frac{\partial\phi_2}{\partial\boldsymbol{\beta}}=\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}-\lambda(\boldsymbol{\beta}'\boldsymbol{\beta}-1)-2\nu_1\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}\right\}=\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta})-\lambda\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\beta})-2\nu_1\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}) \tag{6.30}$$

と求まる．第一項目，第二項目はラグランジュ関数 ϕ のベクトル微分と同様であり，第三項目は

$$-2\nu_1\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)})=-2\nu_1\begin{pmatrix}\frac{\partial(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)})}{\partial\beta_1}\\ \vdots\\ \frac{\partial(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)})}{\partial\beta_p}\end{pmatrix}=-2\nu_1\begin{pmatrix}\frac{\partial(\Sigma_{i,j}\beta_i\sigma_{ij}\beta_j^{(1)})}{\partial\beta_1}\\ \vdots\\ \frac{\partial(\Sigma_{i,j}\beta_i\sigma_{ij}\beta_j^{(1)})}{\partial\beta_p}\end{pmatrix}=-2\nu_1\begin{pmatrix}\Sigma_j\sigma_{1j}\beta_j^{(1)}\\ \vdots\\ \Sigma_j\sigma_{pj}\beta_j^{(1)}\end{pmatrix}=-2\nu_1\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}$$

と求まる．従って，結果を (6.30) に代入すると，

$$\frac{\partial\phi_2}{\partial\boldsymbol{\beta}}=2\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}-2\nu_1\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)} \tag{6.31}$$

を得る．ϕ に対しても行ったように，ここでも $\partial\phi_2/\partial\boldsymbol{\beta}=\boldsymbol{0}$ となる点 $\boldsymbol{\beta}$ を求める．(6.31) を $\boldsymbol{0}$ と結び，左から $\boldsymbol{\beta}'$ をかけると

$$0=2\boldsymbol{\beta}^{(1)'}\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}^{(1)'}\boldsymbol{\beta}-2\nu_1\boldsymbol{\beta}^{(1)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=-2\nu_1\lambda_1 \tag{6.32}$$

を得る．ただし，二つ目の等号では，第一項目が転置を取ると (6.28) より 0 に，第二項目も同様に転置を取ると (6.28) より 0 に，第三項目は λ_1 に対応する (6.23) の解を $\boldsymbol{\beta}^{(1)}$ としていたこと及び (6.25) より $\boldsymbol{\beta}^{(1)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=\lambda_1$ が従う．よって，$\lambda_1\neq 0$ より，$\nu_1=0$ が得られる．$\nu_1=0$ として，(6.31) を $\boldsymbol{0}$ に結ぶと (6.23) と同じ関係式が導かれる．つまり，$\boldsymbol{\beta}$ が $\boldsymbol{\beta}^{(1)}$ と同様に (6.23) を満たす必要が

[2] 注意として，この証明では $\boldsymbol{\Sigma}$ を正定値と仮定している．Lemma 7.1 より $\boldsymbol{\Sigma}$ が正定値のとき λ_1 は正であるから，$\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$ となるのは $\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$ の場合のみであり，$\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$ から $\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$ が導かれている．

あるとわかる．$\boldsymbol{\beta}$ の正規化から $\boldsymbol{\beta} \neq \mathbf{0}$ であり，$\boldsymbol{\beta}^{(1)}$ に対する議論と同じ議論から，λ は (6.23) を満たすことがわかる．ここで，$\lambda_1, \ldots, \lambda_p$ の内，(6.23) と $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$，$\lambda_1 \boldsymbol{\beta}'\boldsymbol{\beta}^{(1)} = 0$ を満たすようなものを $\lambda_{(2)}$ とする．ただし，これに対応する $\boldsymbol{\beta}$ を $\boldsymbol{\beta}^{(2)}$ とよび，$\boldsymbol{X}$ との線形結合を $\boldsymbol{U}_2 = \boldsymbol{\beta}^{(2)'}\boldsymbol{X}$ とする．

これまでのような λ と $\boldsymbol{\beta}$ の決定は同様に続けていくことが出来て，r 番目まで決定出来たとすると，$r+1$ 番目での決定では，まず線形結合 $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_r$ と無相関な正規化線形結合の内，$\boldsymbol{\beta}'\boldsymbol{X}$ が最大の分散を持つ $\boldsymbol{\beta}$ を見つける．$\boldsymbol{U}_i$（$i = 1, \ldots, r$）と $\boldsymbol{\beta}'\boldsymbol{X}$ の無相関性より，この $\boldsymbol{\beta}$ は，

$$0 = \mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{U}_i = \mathscr{E}\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(i)} = \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} = \lambda_{(i)}\boldsymbol{\beta}'\boldsymbol{\beta}^{(i)}, \quad (i = 1, \ldots, r) \tag{6.33}$$

を満たす．(6.28) の導出と同様に，$\boldsymbol{U}_i = \boldsymbol{\beta}^{(i)'}\boldsymbol{X}$ であることと，この $\boldsymbol{\beta}^{(i)}$ に対応する λ が $\lambda_{(i)}$ であることに注意すればよい．最適化したい目的関数が $\boldsymbol{\beta}'\boldsymbol{X}$ の分散と一致するラグランジュ関数を

$$\phi_{r+1} = \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - 1) - 2\sum_{i=1}^{r} \nu_i \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)}$$

とおく．ただし，$\lambda, \nu_1, \ldots, \nu_r$ はラグランジュ乗数である．ϕ_{r+1} は (6.33) から $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} = 0$ が等号制約として与えられた上で，(6.29) の構成と同様にしてラグランジュ関数を構成しているといえる．ここで，$\boldsymbol{\beta}$ による ϕ_{r+1} のベクトル微分は，(6.31) と同様にして，

$$\frac{\partial \phi_{r+1}}{\partial \boldsymbol{\beta}} = 2\boldsymbol{\Sigma}\boldsymbol{\beta} - 2\lambda\boldsymbol{\beta} - 2\sum_{i=1}^{r} \nu_i \boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} \tag{6.34}$$

と求まる．いままでと同様に $\partial\phi_{r+1}/\partial\boldsymbol{\beta} = \mathbf{0}$ を満たす点 $\boldsymbol{\beta}$ を求める．(6.34) を $\mathbf{0}$ と結び，左から $\boldsymbol{\beta}^{(j)'}$（$j = 1, \ldots, r$）をかけると，

$$0 = 2\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta} - 2\lambda\boldsymbol{\beta}^{(j)'}\boldsymbol{\beta} - 2\nu_j\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)} \tag{6.35}$$

と求まる．(6.32) とほぼ同様に求まるが，最後の項は

$$-2\sum_{i=1}^{r} \nu_i \boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} = -2(\nu_1\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)} + \cdots + \nu_j\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)} + \cdots + \nu_r\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(r)}) = -2\nu_j\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)}$$

となる．ここでは，$\boldsymbol{U}_j = \boldsymbol{\beta}^{(j)'}\boldsymbol{X}$ が各 $\boldsymbol{U}_i$（$i = 1, \ldots, r, i \neq j$）と直交していることから，(6.33) のようにして $\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} = 0$ がわかることを用いている．

ここでもし $\lambda_{(j)} \neq 0$ ならば，$0 = \lambda_{(j)}\boldsymbol{\beta}'\boldsymbol{\beta}^{(j)}$ より，$\boldsymbol{\beta}'\boldsymbol{\beta}^{(j)} = 0$ が導かれる．また，(6.28) とその後の議論と同様に，対応する $\boldsymbol{\beta}$ と λ の関係式を用いて $\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)} = \lambda_{(j)}$ と求まり，(6.33) より，$\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta} = 0$ であるから，(6.35) は $0 = -2\nu_j\lambda_{(j)}$ のみ残る．従って，$\lambda_{(j)} \neq 0$ より，$\nu_j = 0$ が与えられるから，これを (6.35) に代入して 0 と結ぶと $\boldsymbol{\beta}$ と λ について (6.23) が得られる．もし $\lambda_{(j)} = 0$ ならば，$\boldsymbol{\beta}^{(j)}$ と $\lambda_{(j)}$ について (6.23) に注意すると，$\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)} = \lambda_{(j)}\boldsymbol{\beta}^{(j)} = 0$ となるから，(6.34) における $\sum_{i=1}^{r}$ の j 番目の項がそもそも消えるため，(6.35) における最後の項が消える．従って，(6.35) を 0 と結ぶと $\boldsymbol{\beta}$ と λ について (6.23) が得られる．

ここで，$\boldsymbol{\beta}^{(2)}$ の決定時と同様に，$\lambda_1, \ldots, \lambda_p$ の内，(6.23) と $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$，$\lambda_{(i)}\boldsymbol{\beta}'\boldsymbol{\beta}^{(i)} = 0$（$i = 1, \ldots, r$）を満たすようなものを $\lambda_{(r+1)}$ とする．ただし，これに対応する $\boldsymbol{\beta}$ を $\boldsymbol{\beta}^{(r+1)}$ と置き，$\boldsymbol{X}$ との線形結合を $\boldsymbol{U}_{r+1} = \boldsymbol{\beta}^{(r+1)'}\boldsymbol{X}$ とする．もし $\lambda_{(r+1)} = 0$ かつ $\lambda_{(j)} = 0$（$j \neq r+1$）ならば，(6.28) 以降のようにして，$\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(r+1)}$ から $\boldsymbol{\beta}^{(j)'}\boldsymbol{\beta}^{(r+1)} = 0$ を導くことが出来ない．というのも，$\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(r+1)}$ は $\boldsymbol{\beta}^{(j)'}\boldsymbol{X}$ と $\boldsymbol{\beta}^{(r+1)'}\boldsymbol{X}$ が無相関であることから (6.33) より求まる式だが，(6.33) のようにして $\lambda_{(j)}\boldsymbol{\beta}^{(r+1)'}\boldsymbol{\beta}^{(j)} = 0$ と求まるものの，$\lambda_{(j)} = 0$ より $\boldsymbol{\beta}^{(r+1)'}\boldsymbol{\beta}^{(j)} = 0$ とは限らないためである．$r+1$ と j を入れ替えても同様のことが言える．始めに $\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)} = 0$ が導けたのは，λ_1 を $\boldsymbol{\Sigma}$ の最大固有値としていたため，λ_1 が

0 ならば $\boldsymbol{\Sigma}$ が 0 になってしまうという結論があったためである．しかし，$\boldsymbol{\beta}^{(r+1)}$ は，$\boldsymbol{\beta}^{(r+1)}$ や $\lambda_{(j)}$ が 0 となる $\boldsymbol{\beta}^{(j)}$ たちの線形結合に置き換えられるため，新しい $\boldsymbol{\beta}^{(r+1)}$ はすべての $\boldsymbol{\beta}^{(j)}$（$j=1,\ldots,r$）に直交する．つまり，$r+1$ 番目のベクトルと j（$j \neq r+1$）番目のベクトルに対応する固有値が 0 であっても，これらのベクトルは直交する．

このような手続きは $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ と (6.23)，$0=\lambda_{(i)}\boldsymbol{\beta}'\boldsymbol{\beta}^{(i)}$（$i=1,\ldots,m$）を満たすベクトル $\boldsymbol{\beta}$ が見つからなくなる $(m+1)$ 段階目まで続けることが出来る．$\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)}$ は，それぞれが直交するという条件から，線形独立でなければならないため，$m=p$ または $m<p$ のどちらかである．というのも，$m>p$ はつまり，固有多項式 (6.24) でその次数を超える p 番目以降の根 λ に番号付けをしていることになるため，すでに番号付けした同じ根に番号付けをすることになる．重解となる固有値に対応する固有ベクトルたちは線形独立となるが，このような重複の場合は同じ固有ベクトル $\boldsymbol{\beta}$ が表れてしまう．一致するベクトル同士は線形従属であるから，$\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)}$ が線形独立であることに反してしまう．

ここで，不等号 $m<p$ が矛盾を導くことを示す．もし $m<p$ ならば，$\boldsymbol{\beta}^{(i)'}\boldsymbol{e}_j=0$ で $\boldsymbol{e}_i'\boldsymbol{e}_j=\delta_{ij}$ を満たす $p-m$ 個のベクトル $\boldsymbol{e}_{m+1},\ldots,\boldsymbol{e}_p$ が存在する．ただし，$\boldsymbol{e}_j$ は $\boldsymbol{\beta}^{(i)}$（$i=1,\ldots,m$）に直交し，$\delta_{ij}=1$（$i=j$），$\delta_{ij}=0$（$i\neq j$）を満たす．これは，Lemma 7.4 から示せる．Lemma 7.4 を今の条件に適用すると，行列 $(\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)})$ は $p\times m\,(p>m)$ で，$(\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)})'(\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)})=\boldsymbol{I}_m$ を満たすため，$p\times(p-m)$ 行列 $\boldsymbol{E}$ で $(\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)}\ \boldsymbol{E})$ が直交行列となるものが存在する．$(\boldsymbol{e}_{m+1},\ldots,\boldsymbol{e}_p)=\boldsymbol{E}$ とする．今，(6.23) において $\lambda=\theta$ とした時，$\boldsymbol{EC}=\sum_i c_i\boldsymbol{e}_i$ が (6.23) の解，つまり $\boldsymbol{\beta}$ となるような $p-m$ 個の要素を持つベクトル $\boldsymbol{C}$ と数 θ が存在することを示す．つまり，Lemma 7.4 から存在が示せる $\boldsymbol{E}$ を使って，見つからないはずの新しい $\boldsymbol{\beta}^{(m+1)}$ とそれに対応する $\lambda_{(m+1)}$ を見つけてしまおうということである．$\boldsymbol{E}'\boldsymbol{E}=\boldsymbol{I}$ に注意すると，(6.24) と同じ等式 $|\boldsymbol{\Sigma}-\theta\boldsymbol{I}|=0$ から $|\boldsymbol{E}'\boldsymbol{\Sigma}\boldsymbol{E}-\theta\boldsymbol{I}|=0$ が導出出来る．この根 θ と，それに対応した (6.23) のような関係式 $\boldsymbol{E}'\boldsymbol{\Sigma}\boldsymbol{EC}=\theta\boldsymbol{c}$ を満たす係数ベクトル $\boldsymbol{C}$ を考える．等式

$$\boldsymbol{\beta}^{(i)'}\boldsymbol{\Sigma}\boldsymbol{EC}=\lambda_{(i)}\boldsymbol{\beta}^{(i)'}\sum_j c_j\boldsymbol{e}_j=\lambda_{(i)}\sum_j c_j\boldsymbol{\beta}^{(i)'}\boldsymbol{e}_j=0$$

より，ベクトル $\boldsymbol{\Sigma}\boldsymbol{EC}$ は $\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)}$ に直交することがわかる．さらに，

$$\boldsymbol{\Sigma}\boldsymbol{E}\boldsymbol{c}=\begin{pmatrix}\sigma_{11}&\cdots&\sigma_{1p}\\ \vdots&\ddots&\vdots\\ \sigma_{p1}&\cdots&\sigma_{pp}\end{pmatrix}\begin{pmatrix}e_{1m+1}&\cdots&e_{1p}\\ \vdots&\ddots&\vdots\\ e_{pm+1}&\cdots&e_{pp}\end{pmatrix}\begin{pmatrix}c_1\\ \vdots\\ c_{p-m}\end{pmatrix}=c_1\left(\sigma_{11},\ldots,\sigma_{1p}\right)\boldsymbol{e}_{m+1}+\cdots+c_{p-m}\left(\sigma_{p1},\ldots,\sigma_{pp}\right)\boldsymbol{e}_p$$

より，$\boldsymbol{\Sigma}\boldsymbol{E}\boldsymbol{c}$ は $\boldsymbol{e}_{m+1},\ldots,\boldsymbol{e}_p$ で張られた空間であることがわかる．また，

$$g_1=c_1(\sigma_{11},\ldots,\sigma_{1p}),\ldots,g_{p-m}=c_{p-m}(\sigma_{p1},\ldots,\sigma_{pp})$$

とおくことで，$\boldsymbol{g}=(g_1,\ldots,g_{p-m})'$ を用いて

$$\begin{aligned}\boldsymbol{\Sigma}\boldsymbol{E}\boldsymbol{c}&=c_1\left(\sigma_{11},\ldots,\sigma_{1p}\right)\boldsymbol{e}_{m+1}+\cdots+c_{p-m}\left(\sigma_{p1},\ldots,\sigma_{pp}\right)\boldsymbol{e}_p\\&=\boldsymbol{e}_{m+1}g_1+\cdots+\boldsymbol{e}_p g_{p-m}=\boldsymbol{E}\boldsymbol{g}\end{aligned}$$

と書けることがわかる．ここで，$\boldsymbol{\Sigma}\boldsymbol{EC}=\boldsymbol{E}\boldsymbol{g}$ の左から $\boldsymbol{E}'$ を掛けると，$\boldsymbol{E}'\boldsymbol{\Sigma}\boldsymbol{EC}=\boldsymbol{E}'\boldsymbol{E}\boldsymbol{g}=\boldsymbol{g}$ が成り立つ．従って，$\boldsymbol{g}=(\boldsymbol{E}'\boldsymbol{\Sigma}\boldsymbol{E})\boldsymbol{C}=\theta\boldsymbol{c}$ であるから，$\boldsymbol{\Sigma}\boldsymbol{EC}=\boldsymbol{E}\boldsymbol{g}=\theta\boldsymbol{E}\boldsymbol{c}$ を得る．これはつまり，これまで決定してきた $\boldsymbol{\beta}$ と λ の関係式 (6.23) をベクトル $\boldsymbol{E}\boldsymbol{c}$ と数 θ が満たすことを意味する．さらに，$\mathscr{E}\boldsymbol{\beta}^{(j)'}\boldsymbol{X}(\boldsymbol{E}\boldsymbol{c})'\boldsymbol{X}=\mathscr{E}\boldsymbol{\beta}^{(j)'}\boldsymbol{X}\boldsymbol{X}'(\boldsymbol{E}\boldsymbol{c})=\boldsymbol{\beta}^{(j)'}\mathscr{E}\boldsymbol{X}\boldsymbol{X}'(\boldsymbol{E}\boldsymbol{c})=\boldsymbol{\beta}^{(j)'}\boldsymbol{\Sigma}\boldsymbol{E}\boldsymbol{c}=0$ より，$(\boldsymbol{E}\boldsymbol{c})'\boldsymbol{X}$ は $\boldsymbol{\beta}^{(j)'}\boldsymbol{X}$ と無相関

であるから，$\boldsymbol{\beta}^{(m+1)}$ として $\boldsymbol{Ec}$ を，$\lambda_{(m+1)}$ として θ を導くことができた．よって，この結論は，$m < p$ という仮定を打ち消すため，$m = p$ が示された．つまりは，正規化や方程式 (6.23)，無相関性といったいくつかの条件を満たすベクトル $\boldsymbol{\beta}$ と λ として，$\boldsymbol{\beta}^{(p)}$ と $\lambda_{(p)}$ まで決定できることが示された．

これまで決定できた $\boldsymbol{\beta}$ と λ たちを行列でまとめる．$\boldsymbol{B} = (\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(p)})$ として，

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{(1)} & 0 & \cdots & 0 \\ 0 & \lambda_{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{(p)} \end{pmatrix}$$

とする．すると，$\boldsymbol{\Sigma\beta}^{(r)} = \lambda_{(r)}\boldsymbol{\beta}^{(r)}$ という等式は，

$$\boldsymbol{\Sigma B} = \boldsymbol{B\Lambda} \tag{6.36}$$

とかけて，$\boldsymbol{\beta}^{(r)'}\boldsymbol{\beta}^{(r)} = 1$ と $\boldsymbol{\beta}^{(r)'}\boldsymbol{\beta}^{(s)} = 0$（$r \neq s$）は，

$$\boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I} \tag{6.37}$$

とかける．ただし，$\boldsymbol{I}$ は p 次の単位行列である．$\boldsymbol{BB}' = \boldsymbol{I}$ も成り立つため，$\boldsymbol{B}$ は直交行列である．(6.36) に左から $\boldsymbol{B}'$ をかけると，(6.37) から，

$$\boldsymbol{B}'\boldsymbol{\Sigma B} = \boldsymbol{\Lambda} \tag{6.38}$$

を得る．さらに，

$$|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| = |\boldsymbol{B}'||\boldsymbol{\Sigma} - \lambda\boldsymbol{I}||\boldsymbol{B}| = |\boldsymbol{B}'\boldsymbol{\Sigma B} - \lambda\boldsymbol{B}'\boldsymbol{B}| = |\boldsymbol{\Lambda} - \lambda\boldsymbol{I}| = \prod_i (\lambda_{(i)} - \lambda) \tag{6.39}$$

が成り立つ．ただし，一つ目の等号は $\boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I}$ と $\boldsymbol{B}$：$p \times p$ 行列に注意して Lemma 7.2 より $|\boldsymbol{B}'\boldsymbol{B}| = 1 \Rightarrow |\boldsymbol{B}'||\boldsymbol{B}| = 1$ を利用した．二つ目の等号は一つ目の等号と同様に，Lemma 7.2 から行列 $\boldsymbol{B}$ と $\boldsymbol{B}'$ を行列式の中に入れた．三つ目の等号は (6.37) と (6.38) を代入した．四つ目の等号は $\boldsymbol{\Lambda} - \lambda\boldsymbol{I}$ の成分が対角成分以外 0 であるから，行列式は対角成分の積となることを用いた．よって，(6.39) を 0 に結ぶと，(6.39) の根 λ が $\boldsymbol{\Lambda}$ の対角成分 $\lambda_{(i)}$（$i = 1, \dots, p$）となることがわかった．そしてこの $\lambda_{(1)}, \dots, \lambda_{(p)}$ たちは $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)}$ となるように決定していたため，同様に大小づけられていた p 個の λ_i（$i = 1, \dots, p$）と対応し，$\lambda_{(1)} = \lambda_1, \lambda_{(2)} = \lambda_2, \dots, \lambda_{(p)} = \lambda_p$ が成り立つ．従って，Theorem 2.1 が帰納的に示された．

ただし，$\mathscr{E}\boldsymbol{UU}' = \boldsymbol{\Lambda}$ は (6.38) を用いて，$\mathscr{E}\boldsymbol{UU}' = \mathscr{E}\boldsymbol{B}'\boldsymbol{XX}'\boldsymbol{B} = \boldsymbol{B}'\boldsymbol{\Sigma B} = \boldsymbol{\Lambda}$ より成り立つ．また，$\boldsymbol{U}$ の r 番目の要素 $U_r = \boldsymbol{\beta}^{(r)'}\boldsymbol{X}$ が $U_1, \dots, U_{r-1}$ と無相関な正規化線形結合の中で最も大きな分散を持つことは以下のように帰納的に確認出来る．

$U^* = \sum_{i=1}^p c_i U_i$（$\sum_{i=1}^p c_i^2 = 1$）とする．このとき，

$$\begin{aligned} \mathscr{V}U^* &= \sum c_i^2 \mathscr{V}U_i = \sum c_i^2 \mathscr{V}\boldsymbol{\beta}^{(i)'}\boldsymbol{X} = \sum c_i^2\lambda_i \\ &= c_1^2\lambda_1 + \cdots + c_p^2\lambda_p = \left(1 - \sum_{i=2}^p c_i^2\right)\lambda_1 + \cdots + c_p^2\lambda_p \\ &= \left(1 - \sum_{i=2}^p c_i^2\right)\lambda_1 + \sum_{i=2}^p c_i^2\lambda_i = \lambda_1 + \sum_{i=2}^p c_i^2(\lambda_i - \lambda_1) \end{aligned} \tag{6.40}$$

が成り立つ．ただし，$\mathscr{V}\boldsymbol{U}^*$ は $\boldsymbol{U}^*$ の分散を表す．$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ より，(6.40) の第二項に注目すると，$c_i^2 = 0$（$i = 2, ..., p$）に対して明らかに最大となる．従って，$\boldsymbol{U}^*$ が最大の分散を持つとき，$\boldsymbol{U}^* = \pm\boldsymbol{U}_1$ であることがわかった．ただし，$\boldsymbol{U}^*$ の分散が最大時，$\mathscr{V}\boldsymbol{U}^* = \mathscr{V}\boldsymbol{U}_1 = \mathscr{V}(-\boldsymbol{U}_1)$ であるから，$\boldsymbol{U}^* = \pm\boldsymbol{U}_1$ である．つまり，$\boldsymbol{X}_i$ たちの任意の正規化線形結合 $\boldsymbol{U}^*$ が最大の分散を持つとき，その分散は $\boldsymbol{U}_1$ と一致することが言えた．同様に，$\boldsymbol{U}_2$ は $\boldsymbol{U}_1$ と無相関な正規化線形結合で最も大きな分散を持つことも言える．以下それを示すため，$\boldsymbol{U}^* = \sum_{i=1}^p c_i \boldsymbol{U}_i$ に $\boldsymbol{U}_1$ と無相関であるという条件を加えると，

$$0 = \mathscr{E}\sum c_i U_i U_1 = \sum \mathscr{E} c_i \boldsymbol{\beta}^{(i)\prime}\boldsymbol{X}\boldsymbol{\beta}^{(1)\prime}\boldsymbol{X} = \sum c_i \mathscr{E}\boldsymbol{\beta}^{(i)\prime}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(1)} = \sum c_i \boldsymbol{\beta}^{(i)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)} = c_1\lambda_1 \tag{6.41}$$

が導ける．ただし，最後の等号では，(6.33) の直交性と (6.25) の $\boldsymbol{\beta}$ と λ の関係式を使った．(6.41) を用いて，

$$\begin{aligned}
\mathscr{V}\boldsymbol{U}^* &= \sum c_i^2\lambda_i = \sum_{i=2}^p c_i^2\lambda_i = c_2^2\lambda_2 + \cdots + c_p^2\lambda_p \\
&= \left(1 - \sum_{i=3}^p c_i^2\right)\lambda_2 + \cdots + c_p^2\lambda_p = \left(1 - \sum_{i=3}^p c_i^2\right)\lambda_2 + \sum_{i=3}^p c_i^2\lambda_i \\
&= \lambda_2 + \sum_{i=3}^p c_i^2(\lambda_i - \lambda_2)
\end{aligned}$$

が成り立つ．従って，$\boldsymbol{U}_1$ の時と同様の議論から，$\boldsymbol{X}_i$ たちの任意の正規化線形結合 $\boldsymbol{U}^*$ で，$\boldsymbol{U}_1$ と無相関なものが最大の分散を持つとき，その分散は $\boldsymbol{U}_2$ と一致することが言える．順々に，$\boldsymbol{U}_3$ から $\boldsymbol{U}_p$ についても，このような分散の最大性の特徴が確認できる． □

Proof. 第 3 章 Theorem 3.1(p.10)

確率変数行列 $\boldsymbol{T} = B'\boldsymbol{S}B$ とおく．$\boldsymbol{Y}$ を

$$\boldsymbol{T} = \boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}' \tag{6.42}$$

を満たす直交行列とする．つまり，(6.42) は Lemma 7.3 による対称行列 $\boldsymbol{T}$ の対角化である．直交行列では各列ベクトルが正規化されているが，その条件だと符号を除いての一意となるため，符号を固定して一意に定める必要がある．従って，(6.42) で $\boldsymbol{Y}$ の対角成分 y_{ii}（$i = 1, ..., p$）を常に $y_{ii} \geq 0$ とする．$\sqrt{n}(\boldsymbol{T} - \boldsymbol{\Lambda}) = \boldsymbol{U}$，$\sqrt{n}(\boldsymbol{Y} - \boldsymbol{I}) = \boldsymbol{W}$ とする．すると，$\sqrt{n}(\boldsymbol{T} - \boldsymbol{\Lambda}) = \boldsymbol{U}$ と $\boldsymbol{T} = \boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}'$ より，$\boldsymbol{\Lambda} + (1/\sqrt{n})\boldsymbol{U} = \boldsymbol{T} = \boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}'$ が成り立つ．また，$\boldsymbol{D} = \sqrt{n}(\boldsymbol{L} - \boldsymbol{\Lambda})$ より $\boldsymbol{L} = \boldsymbol{\Lambda} + (1/\sqrt{n})\boldsymbol{D}$ が成り立ち，$\sqrt{n}(\boldsymbol{Y} - \boldsymbol{I}) = \boldsymbol{W}$ より $\boldsymbol{Y} = \boldsymbol{I} + (1/\sqrt{n})\boldsymbol{W}$ が成り立つから，

$$\boldsymbol{\Lambda} + \frac{1}{\sqrt{n}}\boldsymbol{U} = \left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{W}\right)\left(\boldsymbol{\Lambda} + \frac{1}{\sqrt{n}}\boldsymbol{D}\right)\left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{W}\right)'$$

と書ける．これは

$$\begin{aligned}
\boldsymbol{U} &= \sqrt{n}\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}\right)\left(\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{D}\right)\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}\right)' - \sqrt{n}\boldsymbol{\Lambda} \\
&= \sqrt{n}\left(\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{D}+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{\Lambda}+\frac{1}{n}\boldsymbol{W}\boldsymbol{D}\right)\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}'\right) - \sqrt{n}\boldsymbol{\Lambda} \\
&= \left(\sqrt{n}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{W}\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{D}\right)\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}'\right) - \sqrt{n}\boldsymbol{\Lambda} \\
&= \sqrt{n}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{W}'+\boldsymbol{D}+\frac{1}{\sqrt{n}}\boldsymbol{D}\boldsymbol{W}'+\boldsymbol{W}\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}'+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{D}+\frac{1}{n}\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}' - \sqrt{n}\boldsymbol{\Lambda} \\
&= \boldsymbol{W}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{\Lambda}\boldsymbol{W}'+\frac{1}{\sqrt{n}}(\boldsymbol{W}\boldsymbol{D}+\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}'+\boldsymbol{D}\boldsymbol{W}')+\frac{1}{n}\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}'
\end{aligned}$$

と式変形できるため，

$$\boldsymbol{U} = \boldsymbol{W}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{\Lambda}\boldsymbol{W}'+\frac{1}{\sqrt{n}}\left(\boldsymbol{W}\boldsymbol{D}+\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}'+\boldsymbol{D}\boldsymbol{W}'\right)+\frac{1}{n}\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}' \tag{6.43}$$

に等しい．$\boldsymbol{Y}$ が直交行列であることから $\boldsymbol{I}=\boldsymbol{Y}\boldsymbol{Y}'=\left\{\boldsymbol{I}+(1/\sqrt{n})\boldsymbol{W}\right\}\left\{\boldsymbol{I}+(1/\sqrt{n})\boldsymbol{W}'\right\}$ が成り立つ．この左辺と右辺から，

$$\boldsymbol{0} = \boldsymbol{W}+\boldsymbol{W}'+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{W}' \tag{6.44}$$

を得る．

ここから，$\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ と $\boldsymbol{G}=\sqrt{n}(\boldsymbol{B}-B)$ の漸近分布を求めていく．つまり，n（または N）について極限を取った分布を求めていくが，その分布が近似的に何かしらの分布に一致していることを仮定する．近似が正しいかどうかは最後に確認するとして，その方法を発見的に進める．(6.43) と (6.44) について $n\to\infty$ として，係数が $(1/\sqrt{n})$ と $(1/n)$ の項を無視すると，

$$\boldsymbol{U} \approx \boldsymbol{W}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{\Lambda}\boldsymbol{W}' \tag{6.45}$$

$$\boldsymbol{0} \approx \boldsymbol{W}+\boldsymbol{W}' \tag{6.46}$$

を得る．(6.46) から，$\boldsymbol{W}\approx-\boldsymbol{W}'$ を得る．これにより $w_{ij}\approx -w_{ji}$ と $w_{ii}\approx 0$ が自然にわかる．さらに，$\boldsymbol{W}\approx-\boldsymbol{W}'$ を (6.45) へ代入し，結果的な要素を書けば，行列 $\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$，$\boldsymbol{W}=\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$，$\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ について漸近的な近似

$$d_i \approx u_{ii}, \quad (i=1,\ldots,p), \tag{6.47}$$

$$w_{ij} \approx \frac{u_{ij}}{\lambda_j-\lambda_i}, \quad (i\neq j,\ i,j=1,\ldots,p) \tag{6.48}$$

がわかる．この二つの等号については以下のように行列の要素を書き出して確認出来る．まず，

$$\boldsymbol{U}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_p), \quad \boldsymbol{W}=(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_p), \quad \boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\ldots,\lambda_p), \quad \boldsymbol{D}=\mathrm{diag}(d_1,\ldots,d_p)$$

とする．ただし, 列番号と添え字が一致するように $\boldsymbol{u}_i=(u_{1i},\ldots,u_{pi})'$, $\boldsymbol{w}_i=(w_{1i},\ldots,w_{pi})'$（$i=1,\ldots,p$）である．すると，

$$\boldsymbol{W}\boldsymbol{\Lambda}=(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_p)\mathrm{diag}(\lambda_1,\ldots,\lambda_p)=(\lambda_1\boldsymbol{w}_1,\ldots,\lambda_p\boldsymbol{w}_p)=(\lambda_j w_{ij})_{p\times p},$$

$$
\begin{aligned}
\boldsymbol{\Lambda W} &= \mathrm{diag}(\lambda_1,\dots,\lambda_p)(\boldsymbol{w}_1,\dots,\boldsymbol{w}_p)\\
&= \mathrm{diag}(\lambda_1,\dots,\lambda_p)\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p}\\ \vdots & \vdots & \ddots & \vdots\\ w_{p1} & w_{p2} & \cdots & w_{pp}\end{pmatrix}\\
&= \begin{pmatrix} \lambda_1 w_{11} & \lambda_1 w_{12} & \cdots & \lambda_1 w_{1p}\\ \vdots & \vdots & \ddots & \vdots\\ \lambda_p w_{p1} & \lambda_p w_{p2} & \cdots & \lambda_p w_{pp}\end{pmatrix}\\
&= (\lambda_i w_{ij})_{p\times p}
\end{aligned}
$$

と書けるから，$\boldsymbol{W\Lambda}$ と $\boldsymbol{\Lambda W}$ の (i,j) 成分に注目すると，

$$
\boldsymbol{W\Lambda}-\boldsymbol{\Lambda W} = \lambda_j w_{ij} - \lambda_i w_{ij} = (\lambda_j-\lambda_i)w_{ij} = \begin{cases} 0 & (\text{対角成分})\\ (\lambda_j-\lambda_i)w_{ij} & (\text{非対角成分 } i\neq j,\ i,j=1,\dots,p)\end{cases} \tag{6.49}
$$

が成り立つ．また，

$$
\boldsymbol{U}-\boldsymbol{D} = (\boldsymbol{u}_1,\dots,\boldsymbol{u}_p) - \mathrm{diag}(d_1,\dots,d_p) = \begin{cases} u_{ii}-d_i & (\text{対角成分 } i=1,\dots,p)\\ u_{ij} & (\text{非対角成分 } i\neq j,\ i,j=1,\dots,p)\end{cases} \tag{6.50}
$$

が成り立つ．(6.45) より，$\boldsymbol{W\Lambda}-\boldsymbol{\Lambda W}=\boldsymbol{U}-\boldsymbol{D}$ が成り立つので，(6.49) と (6.50) の対角成分と非対角成分をそれぞれ比較すると，(6.47) と (6.48) が確認できる．

ここで，確率変数行列 $n\boldsymbol{T}=nB'\boldsymbol{S}B$ は $\mathscr{W}(\boldsymbol{\Lambda},n)$ に従うことを示す．$n\boldsymbol{S}$ が $\mathscr{W}(\boldsymbol{\Sigma},n)$ に従うことから，その密度は

$$
\frac{|n\boldsymbol{S}|^{\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}n\boldsymbol{S}}}{2^{\frac{1}{2}pn}\pi^{p(p-1)/4}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]} \tag{6.51}
$$

とかける．$n\boldsymbol{T}=nB'\boldsymbol{S}B$ より，$n\boldsymbol{S}=nB\boldsymbol{T}B'$ であるから，(6.51) は

$$
\frac{|nB\boldsymbol{T}B'|^{\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}nB\boldsymbol{T}B'}}{2^{\frac{1}{2}pn}\pi^{p(p-1)/4}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]} \tag{6.52}
$$

とかける．ただし，B が直交行列（$B'=B^{-1}$）であることから $n\boldsymbol{T}=nB'\boldsymbol{S}B$ に対するヤコビ行列は 1 になる．さらに，B が直交行列であることから，

$$
|nB\boldsymbol{T}B'|^{\frac{1}{2}(n-p-1)} = |B|^{\frac{1}{2}(n-p-1)}|n\boldsymbol{T}|^{\frac{1}{2}(n-p-1)}|B|^{-\frac{1}{2}(n-p-1)} = |n\boldsymbol{T}|^{\frac{1}{2}(n-p-1)}
$$

と

$$
e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}nB\boldsymbol{T}B'} = e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}Bn\boldsymbol{T}B'} = e^{-\frac{1}{2}\mathrm{tr}B'\boldsymbol{\Sigma}^{-1}Bn\boldsymbol{T}} = e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Lambda}^{-1}n\boldsymbol{T}}
$$

が成り立ち，また

$$
|\boldsymbol{\Sigma}|^{\frac{1}{2}n} = |B|^{-\frac{1}{2}n}|B|^{\frac{1}{2}n}|\boldsymbol{\Sigma}|^{\frac{1}{2}n} = |B'|^{\frac{1}{2}n}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}|B|^{\frac{1}{2}n} = |B'\boldsymbol{\Sigma}B|^{\frac{1}{2}n} = |\boldsymbol{\Lambda}|^{\frac{1}{2}n}
$$

が成り立つ．従って，(6.52) は

$$
\frac{|n\boldsymbol{T}|^{\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Lambda}^{-1}n\boldsymbol{T}}}{2^{\frac{1}{2}pn}\pi^{p(p-1)/4}|\boldsymbol{\Lambda}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}
$$

と表せる．この密度は $\mathscr{W}(\boldsymbol{\Lambda},n)$ に従う確率変数行列の密度であるから，$n\boldsymbol{T}$ が $\mathscr{W}(\boldsymbol{\Lambda},n)$ に従うことが確認できた．

ここで，$n\boldsymbol{T}=nB'\boldsymbol{S}B$ と Theorem 1.7 の対応を考える．$n\boldsymbol{T}$ が $\mathscr{W}(\boldsymbol{\Lambda},n)$ に従うということから，ここで観測されている N 個の p 次元標本を $\boldsymbol{X}_1^*,\ldots,\boldsymbol{X}_N^*$（$N>p$）とすると，これらは独立に $\mathscr{N}(\boldsymbol{\mu}^*,\boldsymbol{\Lambda})$ に従い，$n\boldsymbol{T}=\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha^*-\overline{\boldsymbol{X}}^*)(\boldsymbol{X}_\alpha^*-\overline{\boldsymbol{X}}^*)'$ であることがわかる．従って，$n\boldsymbol{T}$ は Theorem 1.7 における $\boldsymbol{A}$ と同様の条件を満たす行列である．よって，$\tilde{\boldsymbol{T}}=(1/\sqrt{n})[n\boldsymbol{T}-n\boldsymbol{\Lambda}]$ の極限分布は要素それぞれの期待値が 0 で，行列の (i,j) 成分と (k,l) 成分の共分散が

$$\mathscr{E}\tilde{t}_{ij}\tilde{t}_{kl}=\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}$$

である正規分布となる．ただし，$\tilde{t}_{ij}$ は $\tilde{\boldsymbol{T}}$ の i,j 成分で，λ_{ij} は $\boldsymbol{\Lambda}$ の i,j 成分である．

$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ という定義を思い出すと，$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})=(n/\sqrt{n})(\boldsymbol{T}-\boldsymbol{\Lambda})=(1/\sqrt{n})[n\boldsymbol{T}-n\boldsymbol{\Lambda}]$ であるから，$\boldsymbol{U}$ は上述の $\tilde{\boldsymbol{T}}=(1/\sqrt{n})[n\boldsymbol{T}-n\boldsymbol{\Lambda}]$ に等しく，その極限分布は要素それぞれの平均が 0 で，行列の (i,j) 成分と (k,l) 成分の共分散が

$$\mathscr{E}u_{ij}u_{kl}=\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}$$

である正規分布となる．ただし，u_{ij} は $\boldsymbol{U}$ の i,j 成分である．ただし，$\boldsymbol{\Lambda}$ が対角行列であったことを思い出すと，その要素 λ_{ij} は $i\neq j$ のときに 0 となるため，今後は定理の設定通り λ_{ii} を λ_i と表記する．すると，$\boldsymbol{U}$ の要素 u_{ij} の極限分布における分散 $\mathscr{AV}u_{ij}$ は，

$$\mathscr{AV}u_{ij}=\mathscr{E}u_{ij}^2=\mathscr{E}u_{ij}u_{ij}=\lambda_{ii}\lambda_{jj}+\lambda_{ij}\lambda_{ji}=\lambda_i\lambda_j+\lambda_{ij}\lambda_{ji}$$

と求まる．以上より，$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の要素は極限分布において，平均 0，分散 $\mathscr{AV}u_{ii}=\lambda_i\lambda_i+\lambda_{ii}\lambda_{ii}=\lambda_i\lambda_i+\lambda_i\lambda_i=2\lambda_i^2$，$\mathscr{AV}u_{ij}=\lambda_i\lambda_j+0=\lambda_i\lambda_j$（$i\neq j$）の正規分布に従う．

Theorem 1.7 は極限分布における要素の独立性が成り立つことは主張していないため，Theorem 3.1 の主張に含まれる独立性を導く必要がある．漸近分布における $\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の要素の共分散は

$$\mathscr{E}u_{ij}u_{kl}=\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}$$

となっていた．異なる要素同士かつ $i\neq l$ かつ $j\neq k$ のとき，つまり対角成分を軸に対称の位置にある要素同士以外では，共分散が 0 になる．一般には，共分散が 0 であることから独立性は導かれないが，Lemma 7.17 より多変量正規分布においては共分散が 0 であること（無相関であること）と独立性が必要十分条件になる．従って，$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の要素を列ベクトルに並べれば Theorem 1.7 よりこれは多変量正規分布に従い，要素のうち共分散が 0 になるような要素同士の独立性が主張できる．よって，漸近分布において $\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の共分散が 0 になるような要素たちが実際に独立であることが導かれた．

ここで，$\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ と $\boldsymbol{W}=\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$ の極限分布を考える．$d_i\approx u_{ii}$（$i=1,\ldots,p$）と $w_{ij}\approx u_{ij}/(\lambda_j-\lambda_i)$（$i\neq j,\ i,j=1,\ldots,p$）が成り立っていたことと，$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の要素 u_{ij} の極限分布の結果 $\mathscr{E}u_{ij}u_{kl}=\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}$ より，$\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ と $\boldsymbol{W}=\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$ の極限分布は正規分布で，$d_1,\ldots,d_p,w_{12},w_{13},\ldots,w_{1p},w_{23},w_{24},\ldots,w_{2p},\ldots,w_{p-1\,p}$ は平均 0，分散

$$\mathscr{AV}d_i\approx\mathscr{AV}u_{ii}=2\lambda_i^2,\quad(i=1,\ldots,p),$$

$$\mathscr{AV}w_{ij}\approx\mathscr{AV}\left[\frac{u_{ij}}{\lambda_j-\lambda_i}\right]=\frac{1}{(\lambda_j-\lambda_i)^2}\mathscr{AV}u_{ij}=\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2},\quad(i=1,\ldots,p-1,\ j=i+1,\ldots,p)$$

で独立である．独立性については，

$$\mathscr{AC}(d_i,d_j)=\mathscr{E}d_id_j\approx\mathscr{E}u_{ii}u_{jj}=\lambda_{ij}\lambda_{ij}+\lambda_{ij}\lambda_{ij}=0,\quad (i\neq j,i,j=1,\ldots,p),$$

$$\begin{aligned}\mathscr{AC}(w_{ij},w_{kl})=\mathscr{E}w_{ij}w_{kl}&\approx\mathscr{E}\left(\frac{u_{ij}}{\lambda_j-\lambda_i}\right)\left(\frac{u_{kl}}{\lambda_l-\lambda_k}\right)\\&=\frac{1}{(\lambda_j-\lambda_i)(\lambda_l-\lambda_k)}\mathscr{E}u_{ij}u_{kl}\\&=0,\\&(i\neq k,\ j\neq l,\ i=1,\ldots,p-1,\ j=i+1,\ldots,p,\ k=1,\ldots,p-1,\ l=k+1,\ldots,p),\end{aligned}$$

$$\begin{aligned}\mathscr{AC}(d_i,w_{kl})=\mathscr{E}d_iw_{kl}&\approx\mathscr{E}\left[u_{ii}\frac{u_{kl}}{\lambda_l-\lambda_k}\right]\\&=\frac{1}{\lambda_l-\lambda_k}\mathscr{E}u_{ii}u_{kl}\\&=\frac{1}{\lambda_l-\lambda_k}(\lambda_{ik}\lambda_{il}+\lambda_{il}\lambda_{ik})\\&=0,\\&(i\neq k\ \text{または}\ i\neq l,i=1,..,p,\ k=1,\ldots,p-1,\ l=k+1,\ldots,p)\end{aligned}$$

であり，挙げた要素同士の共分散は 0 となるから，$\boldsymbol{U}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})$ の極限分布で述べたように独立となる．注意として，w_{ii} は極限分布において常に 0 であるからここでは考えず，また極限分布においては $w_{ij}\approx -w_{ji}$ であったから w_{ij} と w_{ji} は独立でないため，$i<j$ となる部分で独立性とその分散を示した．従って，$\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ についてまとめると，その対角要素は漸近的に独立で $\mathscr{N}(0,2\lambda_i^2)$ に従うことがわかった．

ここで，$B\boldsymbol{W}=B\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$ と $\boldsymbol{G}=\sqrt{n}(\boldsymbol{B}-B)$ が同じ極限分布を持つことを示す．ただし，$\boldsymbol{\Sigma}=B\boldsymbol{\Lambda}B'$，$\boldsymbol{S}=\boldsymbol{BLB}'$ であった．なぜなら，極限分布において $\boldsymbol{D}$ と $\boldsymbol{W}$ が独立であることは示したから，$\boldsymbol{W}$ に左から B を掛けた $B\boldsymbol{W}$ と $\boldsymbol{D}$ の独立性は確かめられ，$B\boldsymbol{W}$ と $\boldsymbol{G}$ が漸近的に等しいことを示せば，$\boldsymbol{D}$ と $\boldsymbol{G}$ の独立性も言えるからである．そして，$\boldsymbol{W}$ の漸近的な分散と共分散から $B\boldsymbol{W}$ の漸近的な分散と共分散を求めることができるため，$B\boldsymbol{W}$ と漸近的に等しい行列 $\boldsymbol{G}$ の漸近的な分散と共分散も求まるからである．$B\boldsymbol{W}=B\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$ と $\boldsymbol{G}=\sqrt{n}(\boldsymbol{B}-B)$ が同じ極限分布を持つことを示す．$B\boldsymbol{W}=B\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})=\sqrt{n}(B\boldsymbol{Y}-B)$ より，$B\boldsymbol{W}$ が $\boldsymbol{G}=\sqrt{n}(\boldsymbol{B}-B)$ と漸近的に等しいことを示すには $B\boldsymbol{Y}$ と $\boldsymbol{B}$ が漸近的に等しいことを示せば良い．$\boldsymbol{W}=\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})$ より，$\boldsymbol{Y}=\boldsymbol{I}+(1/\sqrt{n})\boldsymbol{W}$ であるから，$n\to\infty$ とすると $\boldsymbol{Y}\xrightarrow{p}\boldsymbol{I}$ が示せ，これより $B\boldsymbol{Y}\xrightarrow{p}B$ が成り立ち，さらに $B\xrightarrow{p}\pm\boldsymbol{B}$ が成り立つから，$B\boldsymbol{Y}\xrightarrow{p}\pm\boldsymbol{B}$ が成り立つ．従って，任意に高い確率で $\boldsymbol{B}$ の各列が近似的に $B\boldsymbol{Y}$ の対応する列と同じであるから，$\boldsymbol{B}$ の各列は，漸近的に ±（$B\boldsymbol{Y}$ の対応する列）である．従って，$\boldsymbol{G}=\sqrt{n}(\boldsymbol{B}-B)$ は $B\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})=B\boldsymbol{W}$ の極限分布を持つ．

まとめると，漸近的に $\boldsymbol{G}$ は $\boldsymbol{D}$ と（共分散が 0 になるような要素同士に限って）独立で，正規分布に従い，期待値や分散，共分散は $B\boldsymbol{W}$ と同じものを持つ．実際に，$\boldsymbol{G}\approx B\boldsymbol{W}$ の列ベクトルの漸近的な期待値や共分散行列を求めていく．

$$B=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p),\quad \boldsymbol{W}=\begin{pmatrix}w_{11}&\cdots&w_{1p}\\\vdots&\ddots&\vdots\\w_{p1}&\cdots&w_{pp}\end{pmatrix},\quad \boldsymbol{G}=(\boldsymbol{g}_1,\ldots,\boldsymbol{g}_p)$$

とする．すると，$BW=(\boldsymbol{\beta}_1 w_{11}+\cdots+\boldsymbol{\beta}_p w_{p1},\ldots,\boldsymbol{\beta}_1 w_{1p}+\cdots+\boldsymbol{\beta}_p w_{pp})\approx(\boldsymbol{g}_1,\ldots,\boldsymbol{g}_p)=\boldsymbol{G}$ である．よって，まず期待値は

$$\mathscr{E}\boldsymbol{g}_i \approx \mathscr{E}\left[\boldsymbol{\beta}_1 w_{1i}+\cdots+\boldsymbol{\beta}_p w_{pi}\right]=\boldsymbol{0}, \quad (i=1,\ldots,p)$$

と求まる．また，共分散行列は

$$\begin{aligned}
\mathscr{A}\mathscr{C}\boldsymbol{g}_i &= \mathscr{E}\boldsymbol{g}_i\boldsymbol{g}_i' \approx \mathscr{E}(\boldsymbol{\beta}_1 w_{1i}+\cdots+\boldsymbol{\beta}_p w_{pi})(\boldsymbol{\beta}_1 w_{1i}+\cdots+\boldsymbol{\beta}_p w_{pi})' \\
&= \mathscr{E}(\boldsymbol{\beta}_1 w_{1i}+\cdots+\boldsymbol{\beta}_p w_{pi})(\boldsymbol{\beta}_1' w_{1i}+\cdots+\boldsymbol{\beta}_p' w_{pi}) \\
&= \mathscr{E}\left[\boldsymbol{\beta}_1 w_{1i}(\boldsymbol{\beta}_1' w_{1i}+\cdots+\boldsymbol{\beta}_p' w_{pi})+\cdots+\boldsymbol{\beta}_p w_{pi}(\boldsymbol{\beta}_1' w_{1i}+\cdots+\boldsymbol{\beta}_p' w_{pi})\right] \\
&= \mathscr{E}\left[(\boldsymbol{\beta}_1\boldsymbol{\beta}_1' w_{1i}^2+\cdots+\boldsymbol{\beta}_1\boldsymbol{\beta}_p' w_{1i}w_{pi})+\cdots+(\boldsymbol{\beta}_p\boldsymbol{\beta}_1' w_{pi}w_{1i}+\cdots+\boldsymbol{\beta}_p\boldsymbol{\beta}_p' w_{pi}^2)\right] \\
&= (\boldsymbol{\beta}_1\boldsymbol{\beta}_1'\mathscr{E}w_{1i}^2+\cdots+\boldsymbol{\beta}_1\boldsymbol{\beta}_p'\mathscr{E}w_{1i}w_{pi})+\cdots+(\boldsymbol{\beta}_p\boldsymbol{\beta}_1'\mathscr{E}w_{pi}w_{1i}+\cdots+\boldsymbol{\beta}_p\boldsymbol{\beta}_p'\mathscr{E}w_{pi}^2) \\
&= \left(\sum_{k=1,k\neq i}^{p}\boldsymbol{\beta}_1\boldsymbol{\beta}_k'\mathscr{E}w_{1i}w_{ki}\right)+\cdots+\left(\sum_{k=1,k\neq i}^{p}\boldsymbol{\beta}_p\boldsymbol{\beta}_k'\mathscr{E}w_{pi}w_{ki}\right) \\
&= \sum_{l=1,l\neq i}^{p}\left(\sum_{k=1,k\neq i}^{p}\boldsymbol{\beta}_l\boldsymbol{\beta}_k'\mathscr{E}w_{li}w_{ki}\right)
\end{aligned} \tag{6.53}$$

と求まる．ただし，$w_{ii}\approx 0$（$i=1,\ldots,p$）に注意して和を取っている．w_{ij} については，共分散が 0 になるもの同士でしか共分散の計算を行っていなかった．なので，$\mathscr{E}w_{li}w_{ki}$（$k\neq i$）を求める必要がある．$w_{ij}\approx u_{ij}/(\lambda_j-\lambda_i)$（$i\neq j$, $i,j=1,\ldots,p$）と $\mathscr{E}u_{ij}u_{kl}=\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}$ に注意すると，

$$\begin{aligned}
\mathscr{E}w_{li}w_{ki} &\approx \mathscr{E}\frac{u_{li}}{\lambda_i-\lambda_l}\frac{u_{ki}}{\lambda_i-\lambda_k} \\
&= \frac{1}{\lambda_i-\lambda_l}\frac{1}{\lambda_i-\lambda_k}\mathscr{E}u_{li}u_{ki} \\
&= \frac{1}{\lambda_i-\lambda_l}\frac{1}{\lambda_i-\lambda_k}(\lambda_{lk}\lambda_{ii}+\lambda_{li}\lambda_{ik}) \\
&= \frac{1}{\lambda_i-\lambda_l}\frac{1}{\lambda_i-\lambda_k}(\lambda_{lk}\lambda_i), \quad (k\neq i)
\end{aligned}$$

であるから，(6.53) は「$k=l$」以外で $\boldsymbol{0}$ になる．よって，

$$\mathscr{A}\mathscr{C}\boldsymbol{g}_i=\sum_{l=1,l\neq i}^{p}\boldsymbol{\beta}_l\boldsymbol{\beta}_l'\frac{1}{(\lambda_i-\lambda_l)^2}(\lambda_l\lambda_i)=\sum_{k=1,k\neq i}^{p}\frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'$$

と求まる．

最後に，$\boldsymbol{g}_i$ と $\boldsymbol{g}_j$（$i \neq j$）の共分散行列を求める．

$$
\begin{aligned}
\mathscr{AC}(\boldsymbol{g}_i, \boldsymbol{g}_j) &= \mathscr{E}\boldsymbol{g}_i\boldsymbol{g}_j' \approx \mathscr{E}(\boldsymbol{\beta}_1 w_{1i} + \cdots + \boldsymbol{\beta}_p w_{pi})(\boldsymbol{\beta}_1 w_{1j} + \cdots + \boldsymbol{\beta}_p w_{pj})' \\
&= \mathscr{E}(\boldsymbol{\beta}_1 w_{1i} + \cdots + \boldsymbol{\beta}_p w_{pi})(\boldsymbol{\beta}_1' w_{1j} + \cdots + \boldsymbol{\beta}_p' w_{pj}) \\
&= \mathscr{E}\left[\boldsymbol{\beta}_1 w_{1i}(\boldsymbol{\beta}_1' w_{1j} + \cdots + \boldsymbol{\beta}_p' w_{pj}) + \cdots + \boldsymbol{\beta}_p w_{pi}(\boldsymbol{\beta}_1' w_{1j} + \cdots + \boldsymbol{\beta}_p' w_{pj})\right] \\
&= \mathscr{E}\left[(\boldsymbol{\beta}_1\boldsymbol{\beta}_1' w_{1i}w_{1j} + \cdots + \boldsymbol{\beta}_1\boldsymbol{\beta}_p' w_{1i}w_{pj}) + \cdots + (\boldsymbol{\beta}_p\boldsymbol{\beta}_1' w_{pi}w_{1j} + \cdots + \boldsymbol{\beta}_p\boldsymbol{\beta}_p' w_{pi}w_{pj})\right] \\
&= (\boldsymbol{\beta}_1\boldsymbol{\beta}_1'\mathscr{E}w_{1i}w_{1j} + \cdots + \boldsymbol{\beta}_1\boldsymbol{\beta}_p'\mathscr{E}w_{1i}w_{pj}) + \cdots + (\boldsymbol{\beta}_p\boldsymbol{\beta}_1'\mathscr{E}w_{pi}w_{1j} + \cdots + \boldsymbol{\beta}_p\boldsymbol{\beta}_p'\mathscr{E}w_{pi}w_{pj}) \\
&= \left(\sum_{k=1,k\neq j}^{p} \boldsymbol{\beta}_1\boldsymbol{\beta}_k'\mathscr{E}w_{1i}w_{kj}\right) + \cdots + \left(\sum_{k=1,k\neq j}^{p} \boldsymbol{\beta}_p\boldsymbol{\beta}_k'\mathscr{E}w_{pi}w_{kj}\right) \\
&= \sum_{l=1,l\neq i}^{p}\left(\sum_{k=1,k\neq j}^{p} \boldsymbol{\beta}_l\boldsymbol{\beta}_k'\mathscr{E}w_{li}w_{kj}\right)
\end{aligned}
\tag{6.54}
$$

と求まる．ただし，$w_{ii} \approx 0$（$i = 1, \dots, p$）に注意して和を取っている．ここで，$\mathscr{E}w_{li}w_{kj}$（$k \neq j, l \neq i$）を求める必要がある．$w_{ij} \approx u_{ij}/(\lambda_j - \lambda_i)$（$i \neq j$, $i, j = 1, \dots, p$）と $\mathscr{E}u_{ij}u_{kl} = \lambda_{ik}\lambda_{jl} + \lambda_{il}\lambda_{jk}$ に注意すると，

$$
\begin{aligned}
\mathscr{E}w_{li}w_{kj} &\approx \mathscr{E}\frac{u_{li}}{\lambda_i - \lambda_l}\frac{u_{kj}}{\lambda_j - \lambda_k} \\
&= \frac{1}{\lambda_i - \lambda_l}\frac{1}{\lambda_j - \lambda_k}\mathscr{E}u_{li}u_{kj} \\
&= \frac{1}{\lambda_i - \lambda_l}\frac{1}{\lambda_j - \lambda_k}(\lambda_{lk}\lambda_{ij} + \lambda_{lj}\lambda_{ik}) \\
&= \frac{1}{\lambda_i - \lambda_l}\frac{1}{\lambda_j - \lambda_k}(\lambda_{lj}\lambda_{ik}), \quad (k \neq i, \quad l \neq j)
\end{aligned}
$$

であるから，(6.54) は「$k = i$ かつ $l = j$」以外で $\mathbf{0}$ になる．よって，

$$
\mathscr{AC}(\boldsymbol{g}_i, \boldsymbol{g}_j) = \boldsymbol{\beta}_j\boldsymbol{\beta}_i'\frac{1}{\lambda_i - \lambda_j}\frac{1}{\lambda_j - \lambda_i}(\lambda_{jj}\lambda_{ii}) = -\frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_i'
$$

と求まる． □

Proof. 第 4 章 Theorem 4.1(p.11)

Theorem 3.1 で考えた $\boldsymbol{\Sigma} = B\Lambda B'$ と $\boldsymbol{S} = \boldsymbol{BLB}'$ を再び考える．Theorem 3.1 では，$\boldsymbol{B} \xrightarrow{p} \pm B$ が確認できていた．従って，$\boldsymbol{S} \xrightarrow{p} \boldsymbol{\Sigma}$ を確認できれば $\boldsymbol{L} \xrightarrow{p} \boldsymbol{\Lambda}$ が示せる．以下，$\boldsymbol{X}_1, \dots, \boldsymbol{X}_N$ を平均 $\boldsymbol{\mu}$，共分散行列 $\boldsymbol{\Sigma}$ を持つ独立同分布に多変量正規分布に従う N 個の標本として，標本共分散行列 $\boldsymbol{S} = (1/n)\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})'$ が母共分散行列 $\boldsymbol{\Sigma}$ の一致推定量であることを示す．ただし，$n = N - 1$ である．

標本共分散行列 $\boldsymbol{S} = (1/n)\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})'$ の (i, j) 成分 $s_{ij} = (1/n)\sum_{\alpha=1}^{N}(X_{i\alpha} - \overline{X}_i)(X_{j\alpha} - \overline{X}_j)$ は，Lemma 7.19 の $\boldsymbol{b} = \boldsymbol{\mu}$ の場合を考えると，

$$
s_{ij} = \frac{1}{n}\sum_{\alpha=1}^{N}(X_{i\alpha} - \mu_i)(X_{j\alpha} - \mu_j) + \frac{N}{n}(\overline{X}_i - \mu_i)(\overline{X}_j - \mu_j)
$$

と書ける．第二項目 $(N/n)(\overline{X}_i - \mu_i)(\overline{X}_j - \mu_j)$ については，Lemma 7.13 より $\overline{X}_i \xrightarrow{p} \mu_i$ が成り立つから，標本数 N を十分大きくした（確率）極限は 0 である．第一項目 $(1/n)\sum_{\alpha=1}^{N}(X_{i\alpha} - \mu_i)(X_{j\alpha} - \mu_j)$

を U_{ij} とし，これに対して Lemma 7.20 を用いる．Lemma 7.20 より，確率変数 $(Y-\mathscr{E}Y)^2$ と定数 $\varepsilon>0$ に対して

$$\mathrm{P}\{(|Y-\mathscr{E}Y|)^2\geq\varepsilon^2\}\leq\frac{\mathscr{E}(|Y-\mathscr{E}Y|)^2}{\varepsilon^2}$$

が成り立つ．これより

$$\mathrm{P}\{|Y-\mathscr{E}Y|\geq\varepsilon\}\leq\frac{\mathscr{V}Y}{\varepsilon^2}\tag{6.55}$$

を得る．U_{ij} の期待値は

$$\mathscr{E}U_{ij}=\mathscr{E}\left[\frac{1}{n}\sum_{\alpha=1}^{N}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)\right]=\frac{1}{n}\sum_{\alpha=1}^{N}\mathscr{E}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)=\frac{N}{n}\sigma_{ij}$$

と求まる．ただし，最後の等号は共分散の定義から導かれる．U_{ij} の分散は，$\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$ の独立性に注意して

$$\mathscr{V}U_{ij}=\mathscr{V}\left[\frac{1}{n}\sum_{\alpha=1}^{N}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)\right]=\frac{1}{n^2}\sum_{\alpha=1}^{N}\mathscr{V}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)$$

と求まる．ここで，$\mathscr{V}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)$ を計算すると，

$$\begin{aligned}\mathscr{V}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)&=\mathscr{E}(X_{i\alpha}-\mu_i)^2(X_{j\alpha}-\mu_j)^2-\left(\mathscr{E}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)\right)^2\\&=\mathscr{E}(X_{i\alpha}-\mu_i)^2(X_{j\alpha}-\mu_j)^2-\sigma_{ij}^2\\&=\sigma_{ii}\sigma_{jj}+2\sigma_{ij}^2-\sigma_{ij}^2\\&=\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2\end{aligned}$$

を得る．ただし，三番目の等号については注意が必要である．今は Theorem 3.1 と同じ仮定を用いており，$\boldsymbol{X}$ に多変量正規分布を仮定している．従って，多変量正規分布の特性関数と結合モーメントの関係から 4 次モーメントが求められる．Theorem 1.7 の証明より，$\mathscr{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ に従う確率変数ベクトル $\boldsymbol{Z}_\alpha$ に対して，$\mathscr{E}Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha}=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$ が成り立っていたため，$\mathscr{E}[(X_{i\alpha}-\mu_i)^2(X_{j\alpha}-\mu_j)^2]=\sigma_{ii}\sigma_{jj}+\sigma_{ij}\sigma_{ij}+\sigma_{ij}\sigma_{ij}=\sigma_{ii}\sigma_{jj}+2\sigma_{ij}^2$ が成り立っていた．従って，$\mathscr{V}U_{ij}$ は，

$$\mathscr{V}U_{ij}=\frac{N}{n^2}(\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2)$$

と求まる．よって，不等式 (6.55) に $Y=U_{ij}$ を適用すれば，

$$\mathrm{P}\left\{\left|U_{ij}-\frac{N}{n}\sigma_{ij}\right|\geq\varepsilon\right\}\leq\frac{(N/n^2)(\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2)}{\varepsilon^2}$$

を得る．$N\to\infty$ とすると，

$$\lim_{N\to\infty}\mathrm{P}\left\{\left|U_{ij}-\frac{N}{n}\sigma_{ij}\right|\geq\varepsilon\right\}\leq\lim_{N\to\infty}\frac{(N/n^2)(\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2)}{\varepsilon^2}$$

を得る．これは $n=N-1$ に注意して $N/n=1/(1-1/N)\to1$，$N/n^2=N/(N^2-2N+1)=1/(N-2+1/N)\to0$ より，$\lim_{N\to\infty}\mathrm{P}\{|U_{ij}-\sigma_{ij}|\geq\varepsilon\}\leq0$ であり，$\lim_{N\to\infty}\mathrm{P}\{|U_{ij}-\sigma_{ij}|\geq\varepsilon\}=0$ が成り立つ，よって，$U_{ij}\xrightarrow{p}\sigma_{ij}$ が成り立つ．

まとめると，

$$\lim_{N\to\infty} s_{ij} = \lim_{N\to\infty}\left[\frac{1}{n}\sum_{\alpha=1}^{N}(X_{i\alpha}-\mu_i)(X_{j\alpha}-\mu_j)-\frac{N}{n}(\overline{X}_i-\mu_i)(\overline{X}_j-\mu_j)\right] = \lim_{N\to\infty} U_{ij}+0=\sigma_{ij}$$

が導ける．従って，$\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$ が独立同一に平均 $\boldsymbol{\mu}$，共分散行列 $\boldsymbol{\Sigma}$ の多変量正規分布に従うとき，標本共分散行列 $\boldsymbol{S}$ の要素 $s_{ij}=(1/n)\sum_{\alpha=1}^{N}(X_{i\alpha}-\overline{X}_i)(X_{j\alpha}-\overline{X}_j)$ は母共分散行列 $\boldsymbol{\Sigma}$ の要素 σ_{ij} に確率収束する．つまり，s_{ij} は σ_{ij} の一致推定量であることがわかった．よって，標本共分散行列 $\boldsymbol{S}$ は母共分散行列 $\boldsymbol{\Sigma}$ の一致推定量である．従って，$\boldsymbol{\Sigma}=B\Lambda B'$ と $\boldsymbol{S}=\boldsymbol{BLB}'$ について，$\boldsymbol{B}\xrightarrow{p}\pm B$ と $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Sigma}$ が確認できたから，$\boldsymbol{L}\xrightarrow{p}\boldsymbol{\Lambda}$ である．特に i 番目の対角成分に注目すれば $l_i\xrightarrow{p}\lambda_i$ であるから，$\boldsymbol{S}$ の固有値 l_i が $\boldsymbol{\Sigma}$ の固有値 λ_i の一致推定量であることの確認ができた． □

Proof. 第 4 章 Theorem 4.2(p.11)

l_i は λ_i の一致推定量であるから，

$$\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}l_i} \tag{6.56}$$

について，Lemma 7.21 より $l_i-\lambda_i\xrightarrow{p}0$ と $\lambda_i\xrightarrow{d}\lambda_i$ から，$l_i-\lambda_i+\lambda_i\xrightarrow{d}\lambda_i+0$ が成り立つため，近似

$$\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}l_i}=\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}(l_i-\lambda_i+\lambda_i)}\approx\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}(0+\lambda_i)}=\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}\lambda_i}$$

が成り立つ．最右辺の漸近分布は $d_i=\sqrt{n}(l_i-\lambda_i)$ が漸近的に $\mathcal{N}(0,2\lambda_i^2)$ にしたがっていたから，その標準化である

$$\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}\lambda_i}=\frac{d_i-\mathscr{E}d_i}{\sqrt{\mathscr{E}d_i}}$$

は $\mathcal{N}(0,1)$ に従う．従って，(6.56) の極限分布は $\mathcal{N}(0,1)$ となる． □

Proof. 第 4 章 Theorem 4.3(p.13)

Lemma 7.18 において，$\{\boldsymbol{U}_N\}=\{l_i\}$，$\boldsymbol{b}=(\lambda_i)=\lambda_i$ とすると，$\sqrt{n}(l_i-\lambda_i)=d_i$ の $n\to\infty$ のときの極限分布は Theorem 3.1 より $\mathcal{N}(0,2\lambda_i^2)$ である．よって $\boldsymbol{T}=2\lambda_i^2$ を得る．$\boldsymbol{f}(l_i)=(\log l_i)=\log l_i$ とすると，その要素 $\log l_i$ が $l_i=\lambda_i$ で非 0 の微分を持つ．実際，

$$\left.\frac{\partial\log l_i}{\partial l_i}\right|_{l_i=\lambda_i}=\left.\frac{1}{l_i}\right|_{l_i=\lambda_i}=\frac{1}{\lambda_i}\neq 0$$

が成り立つ．よって，$\boldsymbol{\Phi_b}$ の定義より，$\boldsymbol{\Phi_b}=1/\lambda_i$ を得る．従って，Lemma 7.18 より，$\sqrt{n}(\log l_i-\log\lambda_i)$ が極限分布 $\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Phi}_{\boldsymbol{b}}'\boldsymbol{T}\boldsymbol{\Phi_b}\right)=\mathcal{N}\left(0,(1/\lambda_i)\cdot2\lambda_i^2\cdot(1/\lambda_i)\right)=\mathcal{N}(0,2)$ を持つ． □

第7章　補填

補填にある Lemma について，詳細な証明がない場合は参考文献などの一般的な数理統計学，確率論，行列代数，線形代数，微分積分の書籍に証明を委ねている．

Definition 7.1 (半正定値と正定値)**.** 変数ベクトル $\boldsymbol{x} = (x_1, \ldots, x_p)'$ と対称行列 $\boldsymbol{A} = (a_{ij})_{p\times p}$ について，二次形式

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i,j=1}^{p} a_{ij}x_i x_j$$

が全ての $\boldsymbol{x}$ に対して 0 以上のとき，行列 $\boldsymbol{A}$ と二次形式を半正定値と呼ぶ．特に，全ての $\boldsymbol{0}$ でない $\boldsymbol{x}$ に対して 0 より大きい場合は正定値と呼ぶ．

Definition 7.2 (ベクトル微分)**.** 変数ベクトル $\boldsymbol{x} = (x_1, \ldots, x_p)'$ の関数 $S(\boldsymbol{x})$ のベクトル微分は，

$$\frac{\partial S(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial S(\boldsymbol{x})}{\partial x_1}, \frac{\partial S(\boldsymbol{x})}{\partial x_2}, \ldots, \frac{\partial S(\boldsymbol{x})}{\partial x_p}\right)'$$

のように定義される．つまり，$\partial/\partial\boldsymbol{x}$ とは $\boldsymbol{x}$ の各要素による偏微分を取って列に並べることを表す．

Lemma 7.1 (正定値行列の固有値)**.** 正定値行列の任意の固有値は正である．

Proof. David[15] を参照. □

Lemma 7.2. n 次正方行列 $\boldsymbol{A}$ と $\boldsymbol{B}$ について，

$$|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$$

が成り立つ．また，

$$\mathrm{tr}\boldsymbol{A}\boldsymbol{B} = \mathrm{tr}\boldsymbol{B}\boldsymbol{A}$$

も成り立つ．

Proof. David[15] を参照. □

Lemma 7.3 (対称行列の対角化可能性)**.** 任意に与えられた対称行列 $\boldsymbol{B}$ に対して，ある直交行列 $\boldsymbol{C}$ があって，

$$\boldsymbol{C}'\boldsymbol{B}\boldsymbol{C} = \boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_p \end{pmatrix}$$

を満たす．もし，$\boldsymbol{B}$ が半正定値ならば $d_i \geq 0$（$i = 1, \ldots, p$）が成り立ち，もし $\boldsymbol{B}$ が正定値ならば $d_i > 0$（$i = 1, \ldots, p$）が成り立つ．

Proof. Theorem 2.1 の証明は $\boldsymbol{B}$ が半正定値の場合の証明になっている．David[15] を参照. □

Lemma 7.4. $\boldsymbol{A}$ を $n \times m$（$n > m$）行列で，$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_m$ を満たすとする．すると，$n \times (n-m)$ 行列 $\boldsymbol{B}$ で $(\boldsymbol{A}, \boldsymbol{B})$ が直交行列となるものが存在する．

Proof. Anderson[11] を参照. □

Lemma 7.5 (確率変数行列に対する期待値を取る操作の線形性)**.** $\boldsymbol{X}$ を $m \times n$ 確率変数行列，$\boldsymbol{D}$ を $l \times m$ 実行列，$\boldsymbol{E}$ を $n \times q$ 実行列，$\boldsymbol{F}$ を $l \times q$ 実行列とすると，

$$\mathscr{E}(\boldsymbol{DXE} + \boldsymbol{F}) = \boldsymbol{D}(\mathscr{E}\boldsymbol{X})\boldsymbol{E} + \boldsymbol{F}$$

が成り立つ．

Proof. (i, j) 成分に注目して確率変数に対する期待値を取る操作の線形性を考えればよい．Anderson[11] を参照. □

Lemma 7.6 (単位行列化)**.** 正定値行列 $\boldsymbol{A}$ に対し，$\boldsymbol{GAG}' = \boldsymbol{I}$ を満たす下三角行列 $\boldsymbol{G}$（$g_{ij} = 0$，$i < j$）が存在する．

Proof. Anderson[11] を参照. □

Lemma 7.7 (逆写像の定理)**.** $U \subset \mathbb{R}^n$ を開集合，$\boldsymbol{F} : U \to \mathbb{R}^p$ を C^1 級関数とすると，$\boldsymbol{F}$ の点 $a \in U$ におけるヤコビ行列 $\boldsymbol{J_F}(a)$ が正則であるとき，$\boldsymbol{F}$ は a の近傍で可逆となり，この逆関数 $\boldsymbol{F}^{-1}$ もまた C^1 級となる。このとき $\boldsymbol{F}^{-1}$ は

$$\boldsymbol{J}_{\boldsymbol{F}^{-1}}(\boldsymbol{F}(a)) = (\boldsymbol{J_F}(a))^{-1}$$

を満たす．ここで A^{-1} は A の逆行列，$\boldsymbol{J_F}(a)$ は $\boldsymbol{F}$ の点 a におけるヤコビ行列である．

Proof. 長瀬，芦野 [6] を参照. □

Lemma 7.8 (独立性と同値ないくつかの性質)**.** 確率変数 $\{X_i\}$（$i = 1, \ldots, p$）に対応する分布関数を $\{F_{X_i}\}$（$i = 1, \ldots, p$），密度関数を $\{f_{X_i}\}$（$i = 1, \ldots, p$）とする．それぞれの分布関数に対応する特性関数を $\{\phi_{F_{X_i}}\}$（$i = 1, \ldots, p$）とする．また，p 次確率変数ベクトル $\boldsymbol{X} = (X_1, \ldots, X_p)$ の同時分布関数を $F_{\boldsymbol{X}}$ として，これが同時密度関数 $f_{\boldsymbol{X}}$ を持つとする．この同時分布関数に対応する特性関数を $\phi_{F_{\boldsymbol{X}}}$ とする．すると，以下は同値になる．

1. $X_1, \ldots, X_p$ は独立である．
2. 任意の $x_1, \ldots, x_p \in \mathbb{R}$ に対して，$F_{\boldsymbol{X}}(x_1, \ldots, x_p) = \prod_{i=1}^{p} F_{X_i}(x_i)$ が成り立つ．
3. 任意の $x_1, \ldots, x_p \in \mathbb{R}$ に対して，$f_{\boldsymbol{X}}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_{X_i}(x_i)$ が成り立つ．
4. 任意の $t_1, \ldots, t_p \in \mathbb{R}$ に対して，$\phi_{F_{\boldsymbol{X}}}(t_1, \ldots, t_p) = \prod_{i=1}^{p} \phi_{F_{X_i}}(t_i)$ が成り立つ．

Proof. 清水 [4] を参照. □

Lemma 7.9 (分布と特性関数の一対一対応)**.** 任意の定数ベクトル $\boldsymbol{t}$ に対して，確率変数ベクトル $\boldsymbol{X}_1$，$\boldsymbol{X}_2$ の特性関数 $\phi_{\boldsymbol{X}_1}$，$\phi_{\boldsymbol{X}_2}$ が等しいならば，$\boldsymbol{X}_1$，$\boldsymbol{X}_2$ の確率分布は等しい．

Proof. 高信 [5] を参照. □

Lemma 7.10. 確率変数 $\{Z_N\}$ $(N = 1,2,\ldots)$ が確率変数 Z に分布収束し，$a, b \in \mathbb{R}$ であるとき，$aZ_N + b$ は $aZ + b$ に分布収束する．

Proof. Z_N の特性関数を ϕ_N，Z の特性関数を ϕ とすると，Lemma 7.11 から，$\phi_N(t) \to \phi(t)$ $(N \to \infty)$ が成り立つ．これを利用して $aZ + b$ の特性関数の極限を求め，再び Lemma 7.11 と Lemma 7.9 を用いて導かれる． □

Lemma 7.11 (確率変数に対する連続性の定理)**.** $\{F_N(x)\}$ $(N = 1,2,\ldots)$ を分布関数たちの列として，それらに対応する特性関数を $\{\phi_N(t)\}$ とする．ただし，t は任意の実数とする．すると，$F_N(x)$ が分布関数 $F(x)$ に収束するための必要十分条件は，全ての t に対して $\phi_N(t)$ が点 $t = 0$ で連続な極限 $\phi(t)$ に収束することである．この条件が満たされているとき，極限 $\phi(t)$ は極限分布 $F(x)$ の特性関数と等しくなる．

Proof. 証明は Lemma 7.12 の証明において確率変数ベクトルの次元が 1 の場合を考えれば良い． □

Lemma 7.12 (確率変数ベクトルに対する連続性の定理)**.** $\{F_N(\boldsymbol{x})\}$ $(N = 1,2,\ldots)$ を分布関数たちの列として，それらに対応する特性関数を $\{\phi_N(\boldsymbol{t})\}$ とする．ただし，$\boldsymbol{t}$ は $\boldsymbol{x}$ と同じ次元の任意の実数ベクトルとする．すると，$F_N(\boldsymbol{x})$ が分布関数 $F(\boldsymbol{x})$ に収束するための必要十分条件は，全ての $\boldsymbol{t}$ に対して $\phi_N(\boldsymbol{t})$ が点 $\boldsymbol{t} = \boldsymbol{0}$ で連続な極限 $\phi(\boldsymbol{t})$ に収束することである．この条件が満たされているとき，極限 $\phi(\boldsymbol{t})$ は極限分布 $F(\boldsymbol{x})$ の特性関数と等しくなる．

Proof. Cramér[14] の Section 10.6，Section 10.7，Rao[17] の p.119，高信 [5]，西尾 [7] を参照． □

Lemma 7.13 (大数の法則)**.** $\{X_N\}$ $(N = 1,2,\ldots)$ を独立同一な確率変数とすると，$\overline{X}_N = (1/N)\sum_{i=1}^{N} X_i$ が母平均 μ に確率収束する．

Proof. 野田，宮岡 [8] を参照．確率変数ベクトルについては $\lim_{N\to\infty} \mathrm{P}(|\|\overline{\boldsymbol{X}}_N - \boldsymbol{\mu}\| - 0| \geq \varepsilon) = 0$ を示す． □

Lemma 7.14. $\boldsymbol{A} = \sum_{i=1}^{N} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)'$ とする．ただし，$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ たちは独立同一に $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとする．すると，$\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ を独立に $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ に従う確率変数ベクトルとしたとき，$\boldsymbol{A}$ は $\sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i'$ と漸近的に同様の分布となる．ただし，$n = N - 1$ とする．

Proof. 形式的な確認を行う．

$$
\begin{aligned}
\boldsymbol{A} &= \sum_{i=1}^{N} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_N)' \\
&= \sum_{i=1}^{N} \left\{ (\boldsymbol{X}_i - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N) \right\} \left\{ (\boldsymbol{X}_i - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N) \right\}' \\
&= \sum_{i=1}^{N} \left\{ (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})' + (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N)' + (\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i - \boldsymbol{\mu})' + (\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N)(\boldsymbol{\mu} - \overline{\boldsymbol{X}}_N)' \right\}
\end{aligned}
$$

と変形できる．ここで，$\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i-\boldsymbol{\mu})(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i-\boldsymbol{\mu})'+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\right\}$ を考える．

$$
\begin{aligned}
&\sum_{i=1}^{N}\left\{(\boldsymbol{X}_i-\boldsymbol{\mu})(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i-\boldsymbol{\mu})'+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\right\}\\
&\quad=N\left\{\frac{(\boldsymbol{X}_i-\boldsymbol{\mu})\sum_{i=1}^{N}(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'}{N}+\frac{(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)\sum_{i=1}^{N}(\boldsymbol{X}_i-\boldsymbol{\mu})'}{N}+\frac{N(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'}{N}\right\}\\
&\quad=N\left\{(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i-\boldsymbol{\mu}\right)+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i'-\boldsymbol{\mu}'\right)+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\right\}\\
&\quad=N\left\{(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'(\overline{\boldsymbol{X}}_N-\boldsymbol{\mu})+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\overline{\boldsymbol{X}}_N'-\boldsymbol{\mu}')+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\right\}\\
&\quad=N(\overline{\boldsymbol{X}}_N-\boldsymbol{\mu})\left\{(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'+(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'-(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\right\}\\
&\quad=N(\overline{\boldsymbol{X}}_N-\boldsymbol{\mu})(\boldsymbol{\mu}-\overline{\boldsymbol{X}}_N)'\to 0\quad(N\to\infty)
\end{aligned}
$$

と求まる．$N\to\infty$ のとき，Lemma 7.13 より $\overline{\boldsymbol{X}}_N\to\boldsymbol{\mu}$ であるから最後の等号が成り立つ．ただし，ここで「形式的な確認」としているのは，$N\to\infty$ が $\sum_{i=1}^{N}(\boldsymbol{X}_i-\boldsymbol{\mu})(\boldsymbol{X}_i-\boldsymbol{\mu})'$ にも影響するからである．この項以外が $\mathbf{0}$ になるという結果を常識的に知っていることからできる形式的な操作になる．従って．$\boldsymbol{A}=\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_N)(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_N)'$ と $\sum_{i=1}^{N}(\boldsymbol{X}_i-\boldsymbol{\mu})(\boldsymbol{X}_i-\boldsymbol{\mu})'=\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i'\approx\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i'$ が漸近的に同様の分布を持つ． □

Lemma 7.15 (ラグランジュ未定乗数法，小西 [3])**.** 目的関数 $f(\boldsymbol{x})$ を最小化する点 $\boldsymbol{x}$ を，p 個の等号制約 $h_i(\boldsymbol{x})=0$（$i=1,2,\dots,p$）の下で求める．

この等号制約付の最適化問題に対し，目的関数と等号制約式を合わせたラグランジュ関数と呼ばれる実数値関数

$$\phi(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})+\sum_{i=1}^{p}\lambda_i h_i(\boldsymbol{x})$$

を定義する．ただし，等号制約式の係数 λ_i はラグランジュ乗数と呼ばれ，$\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_p)'$ とする．この等号制約式付最適化問題の解は，ラグランジュ関数を $\boldsymbol{x}$ と λ_i で偏微分した方程式

$$\frac{\partial\phi(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\boldsymbol{x}}=\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}+\sum_{i=1}^{p}\lambda_i\frac{\partial h_i(\boldsymbol{x})}{\partial\boldsymbol{x}}=\mathbf{0},$$

$$\frac{\partial\phi(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\lambda_i}=h_i(\boldsymbol{x})=0,\quad(i=1,\dots,p)$$

を満たす解を求める問題に帰着される．

Lemma 7.16 (根 λ たちの大小付け)**.** 方程式 (6.24) の根は実数であり，大小付け可能である．

Proof. $\boldsymbol{\beta}$ を複素列ベクトルとする．$\boldsymbol{\beta}^*$ を $\boldsymbol{\beta}$ の各要素の複素共役を取って転置した行列（随伴行列）とする．$\boldsymbol{\beta}$ が実ベクトルならただの転置と一致する．$(\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta})^*$ は，

$$(\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta})^*=\boldsymbol{\beta}^*\boldsymbol{\Sigma}^*\boldsymbol{\beta}^{**}=\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta}$$

を満たす．ただし，$\boldsymbol{\Sigma}$ が実対称行列であることから，$\boldsymbol{\Sigma}^*=\boldsymbol{\Sigma}$ が成り立つことを用いた．$(\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta})^*=\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta}$ であるから，$\boldsymbol{\beta}^*\boldsymbol{\Sigma}\boldsymbol{\beta}$ は実数である．また，これは λ に等しいことが (6.25) で証明されていた．従って，転置を随伴行列と捉えて考えれば $\boldsymbol{\beta}$ が複素列ベクトルでも λ は実数であると証明できた． □

Lemma 7.17. 確率変数ベクトル $\boldsymbol{X} = (X_1, \ldots, X_N)'$ が多変量正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとする．すると，$i \neq j$ である全ての X_i と X_j の組が無相関であるならば X_i たちは独立である．

Proof. $i \neq j$ となる全ての X_i と X_j の組が無相関のとき，つまりその共分散が 0 であるから $\boldsymbol{X}$ の共分散行列 $\boldsymbol{\Sigma}$ は対角行列 $\mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ となる．ただし，σ_i^2 は X_i の分散で，$\boldsymbol{\Sigma}$ の i 番目の対角成分である．$\boldsymbol{\Sigma}$ の行列式は $|\boldsymbol{\Sigma}| = \sigma_1^2 \cdot \sigma_2^2 \cdots \sigma_p^2$ で，逆行列は $\boldsymbol{\Sigma}^{-1} = \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_p^2)$ である．よって，多変量正規分布に従う $\boldsymbol{X}$ の密度は

$$\begin{aligned}
\frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] &= \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma_1^2\cdot\sigma_2^2\cdots\sigma_p^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\left\{\mathrm{diag}(1/\sigma_1^2,\ldots,1/\sigma_p^2)\right\}(\boldsymbol{x}-\boldsymbol{\mu})\right] \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma_1\cdot\sigma_2\cdots\sigma_p)}\exp\left[-\frac{1}{2}\sum_{i=1}^{p}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right] \\
&= \prod_{i=1}^{p}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right] \qquad (7.1)
\end{aligned}$$

を満たす．(7.1) は一変量正規分布，つまり正規分布 $\mathcal{N}(\mu_i, \sigma_i^2)$ に従う X_i たちの密度の積の形となっている．$\boldsymbol{X}$ の密度が X_i たちの密度関数の積の形に分解できたため，Lemma 7.8 より，X_i たちは独立である． □

Lemma 7.18 (デルタ法)**.** p 次確率変数ベクトル $\{\boldsymbol{U}_N = (u_{N1}, \ldots, u_{Np})'\}$（$N = 1, 2, \ldots$）に対し，定数ベクトル $\boldsymbol{b} \in \mathbb{R}^p$ を $\sqrt{N}[\boldsymbol{U}_N - \boldsymbol{b}]$ が $N \to \infty$ のときに極限分布 $\mathcal{N}(\boldsymbol{0}, \boldsymbol{T})$ を持つように固定したベクトルとする．$\boldsymbol{f}(\boldsymbol{U}_N) = (f_1(\boldsymbol{U}_N), f_2(\boldsymbol{U}_N), \ldots, f_q(\boldsymbol{U}_N))' \in \mathbb{R}^q$ を要素 $f_j(\boldsymbol{U}_N)$ が $\boldsymbol{U}_N = \boldsymbol{b}$ で非 0 の微分を持つ，つまり

$$\left.\frac{\partial f_j(\boldsymbol{U}_N)}{\partial u_{Ni}}\right|_{\boldsymbol{U}_N=\boldsymbol{b}} \neq 0, \quad (i = 1, \ldots, p,\ j = 1, \ldots, q)$$

となるような $\boldsymbol{U}_N$ のベクトル値関数とする．行列 $\boldsymbol{\Phi}_{\boldsymbol{b}}$ を (i, j) 成分が

$$\left.\frac{\partial f_j(\boldsymbol{U}_N)}{\partial u_{Ni}}\right|_{\boldsymbol{U}_N=\boldsymbol{b}}$$

である $p \times q$ 行列とする．つまり，行列 $\boldsymbol{\Phi}_{\boldsymbol{b}}$ は

$$\boldsymbol{\Phi}_{\boldsymbol{b}} = \left(\left.\frac{\partial f_j(\boldsymbol{U}_N)}{\partial u_{Ni}}\right|_{\boldsymbol{U}_N=\boldsymbol{b}}\right)_{p\times q}$$

と表せる．すると，$\sqrt{N}\{\boldsymbol{f}[\boldsymbol{U}_N] - \boldsymbol{f}(\boldsymbol{b})\}$ が極限分布 $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Phi}_{\boldsymbol{b}}'\boldsymbol{T}\boldsymbol{\Phi}_{\boldsymbol{b}})$ を持つ．

Proof. ただし，使っているのは $p = 1$，$q = 1$ の場合である．$p = 1$，$q = 1$ の場合の証明は野田，宮岡 [8] を参照． □

Lemma 7.19. $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ を p 次確率変数ベクトルとして，$\overline{\boldsymbol{X}}$ を $\overline{\boldsymbol{X}} = (1/N)\sum_{i=1}^{N}\boldsymbol{X}_i$ で定義する．すると，任意のベクトル $\boldsymbol{b}$ に対して，

$$\sum_{i=1}^{N}(\boldsymbol{X}_i - \boldsymbol{b})(\boldsymbol{X}_i - \boldsymbol{b})' = \sum_{i=1}^{N}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})' + N(\overline{\boldsymbol{X}} - \boldsymbol{b})(\overline{\boldsymbol{X}} - \boldsymbol{b})'$$

が成り立つ．

Proof.

$$\begin{aligned}\sum_{i=1}^{N}(\boldsymbol{X}_i-\boldsymbol{b})(\boldsymbol{X}_i-\boldsymbol{b})' &= \sum_{i=1}^{N}\left[(\boldsymbol{X}_i-\overline{\boldsymbol{X}})+(\overline{\boldsymbol{X}}-\boldsymbol{b})\right]\left[(\boldsymbol{X}_i-\overline{\boldsymbol{X}})+(\overline{\boldsymbol{X}}-\boldsymbol{b})\right]' \\ &= \sum_{i=1}^{N}\left[(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'+(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}-\boldsymbol{b})'+(\overline{\boldsymbol{X}}-\boldsymbol{b})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'+(\overline{\boldsymbol{X}}-\boldsymbol{b})(\overline{\boldsymbol{X}}-\boldsymbol{b})'\right] \\ &= \sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'+\left[\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})\right](\overline{\boldsymbol{X}}-\boldsymbol{b})' \\ &\quad +(\overline{\boldsymbol{X}}-\boldsymbol{b})\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'+N(\overline{\boldsymbol{X}}-\boldsymbol{b})(\overline{\boldsymbol{X}}-\boldsymbol{b})'\end{aligned}$$

と求まる．ただし，最後の第二，第三項目は $\overline{\boldsymbol{X}}$ の定義より $\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})=\sum_{i=1}^{N}\boldsymbol{X}_i-N\overline{\boldsymbol{X}}=\boldsymbol{0}$ が成り立つ．従って，非 $\boldsymbol{0}$ の項を見ると

$$\sum_{i=1}^{N}(\boldsymbol{X}_i-\boldsymbol{b})(\boldsymbol{X}_i-\boldsymbol{b})'=\sum_{i=1}^{N}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})'+N(\overline{\boldsymbol{X}}-\boldsymbol{b})(\overline{\boldsymbol{X}}-\boldsymbol{b})'$$

が成り立つ． □

Lemma 7.20 (マルコフの不等式)**.** 確率変数 X と実定数 $a>0$ に対し，

$$\mathrm{P}\{|X|\geq a\}\leq\frac{\mathscr{E}|X|}{a}$$

が成り立つ．

Proof.

$$\begin{aligned}\mathscr{E}|X| &= \int_{-\infty}^{\infty}|z|f(z)dz \\ &\geq \int_{-\infty}^{-a}|z|f(z)dx+\int_{a}^{\infty}|z|f(z)dz \\ &\geq a\int_{-\infty}^{-a}f(z)dz+a\int_{a}^{\infty}f(z)dz \\ &= a\mathrm{P}(X\leq -a)+a\mathrm{P}(X\geq a) \\ &= a\mathrm{P}(|X|\geq a)\end{aligned}$$

が成り立つ．従って，両辺を a で割ることで不等式

$$\mathrm{P}(|X|\geq a)\leq\frac{\mathscr{E}|X|}{a}$$

を得る． □

Lemma 7.21 (スラツキーの定理)**.** 確率変数 $\{U_N\}$，$\{V_N\}$（$N=1,2,\ldots$）と確率変数 U，定数 a について $U_N\xrightarrow{d}U$，$V_N\xrightarrow{p}a$ とする．このとき，$U_N+V_N\xrightarrow{d}U+a$，$U_NV_N\xrightarrow{d}aU$ が成り立つ．

Proof. 野田，宮岡 [8] を参照． □

参考文献

(1) 赤平昌文（2003）．統計解析入門．森北出版．

(2) 内田治（2013）．主成分分析の基本と活用．日科技連．

(3) 小西貞則（2010）．多変量解析入門．岩波書店．

(4) 清水泰隆（2019）．統計学への確率論，その先へ．内田老鶴圃．

(5) 高信敏（2015）．確率論．共立出版．

(6) 長瀬道弘，芦野隆一（1999）．微分積分概説．サイエンス社．

(7) 西尾真喜子（1978）．確率論．実教出版．

(8) 野田一雄，宮岡悦良（1992）．数理統計学の基礎．共立出版．

(9) 藤澤洋徳（2006）．確率と統計．朝倉書店．

(10) 吉田朋広（2006）．講座 数学の考え方〈21〉数理統計学．朝倉書店．

(11) Anderson, T. W.（2003）．*An Introduction to Multivariate Statistical Analysis Third Edition*．John Wiley & Sons．New York．

(12) Bishop, C. M.（2006）．*Pattern Recognition and Machine Learning*．Springer．

(13) Chen, Y.，Samworth, R.（2022）． scar: Shape-Constrained Additive Regression: a Maximum Likelihood Approach． R package version 0.2-2． https://CRAN.R-project.org/package=scar．

(14) Cramér, H.（1946）．*Mathematical Methods of Statistics*．Princeton University．Princeton．

(15) David, A. Harville（1997）．*Matrix Algebra From a Statistician's Perspective*．Springer．

(16) LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.（1998）．*Gradient-based lerning applied to document recognition*．Priceedings of the IEEE．

(17) Rao, C. R.（1973）．*Linear Statistical Inference and Its Applications* (2nd ed.)．John Wiley & Sons. New York. Princeton.