

Wishart分布とその性質

多変量正規分布の推定とBox-Cox変換

山北倫太郎

June 25, 2025

1 Wishart分布

7.1 はじめに

- $X = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$ は標本行列を表す。
- \bar{x} と S は、それぞれ μ と Σ の一貫性のある不偏推定量を提供する。
 - $n\bar{x} = X'1$ (標本平均ベクトル(\bar{x}))
 - ここで、 1 は n 次元のベクトルで、すべての要素が 1 である。
 - $(n-1)S = X'X - n\bar{x}\bar{x}'$ (標本共分散行列(S))
 - S は母集団の共分散行列の不偏推定量であり、 $X'X$ は標本行列の転置と自身の積である。

7.1 はじめに

- 7.2節では、 x_1, \dots, x_n が独立同分布で $x \sim N_p(\mu, \Sigma)$ かつ $\Sigma > 0$ の場合の μ と Σ の最尤推定量が導出される。
- \bar{x} と S の同時分布に関する基本的な結果が命題7.1で証明される。
- 7.3節では、Wishart分布の基本的な特性が研究される。
- 7.4節では、データの多変量正規性を高めるためのBox-Cox変換が提示される。

7.2 \bar{x} と S の同時分布

- 正規性がある場合、 \bar{x} と S はいくつかの点で「最適」である。
- $V = (n - 1)S$ とする。
- X の確率密度関数は様々な方法で記述できる。

$$\begin{aligned} f(X) &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right] \\ &= (2\pi)^{-np/2} |\Sigma|^{-n/2} e^{-\frac{n}{2} \mu' \Sigma^{-1} \mu} \text{etr} \left[-\frac{1}{2} \text{tr} \Sigma^{-1} X' X + n \mu' \Sigma^{-1} \bar{x} \right] \\ &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \text{etr} \left[-\frac{1}{2} [V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1} \right] \end{aligned} \quad (7.1)$$

確率変数ベクトル $X = (x_1, \dots, x_p)'$ が平均ベクトル μ と共分散行列 Σ を持つ多変量正規分布に従う場合、その確率密度関数 $f(x)$ は次のように表される。

p次元多変量正規分布の確率密度関数 (p.d.f.)

$$f(x) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right)$$

- ここで、 x は確率変数ベクトル X がとりうる値を示す p 次元の列ベクトルである。
- μ は、各確率変数の平均値を要素とする p 次元の平均ベクトルである。
- Σ は、 $p \times p$ の共分散行列であり、各確率変数間の分散と共分散を表す対称な正定値行列である。

$\exp[\text{tr}(\boxtimes)] = \text{etr}(\boxtimes)$ の記法を用いると、確率密度関数は次のように表される。

$$\begin{aligned}
 & (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \\
 &= (\mathbf{x}_i' - \boldsymbol{\mu}') \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \\
 &= \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i - \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
 &= \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i - 2\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}
 \end{aligned}$$

- ここで、 $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$ はスカラー値であり、スカラーの転置は自分自身なので、 $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i = (\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})'$ です。
したがって、 $\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ と $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$ は同じスカラー値を表します。

2行目 (続き)

$\exp[\text{tr}(\boxtimes)] = \text{etr}(\boxtimes)$ の記法を用いると、確率密度関数は次のように表される。

$$\begin{aligned} & (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \\ &= (\mathbf{x}_i' - \boldsymbol{\mu}') \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \\ &= \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i - \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ &= \mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i - 2\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{aligned}$$

ここで、 $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$ はスカラー値であり、スカラーの転置は自分自身なので、 $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i = (\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})'$ です。したがって、 $\mathbf{x}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ と $\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$ は同じスカラー値を表します。

2行目 (さらに続き)

- μ' は平均ベクトル μ の転置を表し、 μ は p 次元の列ベクトルなので、 μ' は1行 p 列の行ベクトル。
 - Σ^{-1} は共分散行列 Σ の逆行列を表し、 Σ は $p \times p$ の正方行列なので、 Σ^{-1} も $p \times p$ の正方行列。
 - x_i は p 次元の列ベクトルであり、 x_i' はその転置で1行 p 列の行ベクトルになります。
- よって、 $(x_i - \mu)' \Sigma^{-1} (x_i - \mu)$ はスカラー値であります。
- また、 $\mu' \Sigma^{-1} \mu$ は、 $e^{-\frac{1}{2} n \mu' \Sigma^{-1} \mu}$ の形で指数関数に含まれます。

1項目

- 行列のトレースは、行列の対角成分の総和であり、 $\text{tr}(\mathbf{A}) = \sum_i a_{ii}$ で定義される。

トレースの性質

- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$
- $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$, ここで $\mathbf{A} = (a_{ij})$ は $n \times n$ 行列である。

$$\begin{aligned}\sum_{i=1}^n \mathbf{x}_i' \Sigma^{-1} \mathbf{x}_i &= \sum_{i=1}^n \text{tr}(\Sigma^{-1} \mathbf{x}_i \mathbf{x}_i') \\ &= \text{tr}\left(\Sigma^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right) \\ &= \text{tr}(\Sigma^{-1} \mathbf{X}' \mathbf{X})\end{aligned}$$

1項目 (続き)

- $X = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$ であるから、 $X'X = \sum_{i=1}^n x_i x'_i$ となる。
- ここで、 x'_i は x_i の転置を表し、 $x_i x'_i$ は x_i の外積を表す。
- Σ^{-1} は共分散行列の逆行列であり、 $x_i x'_i$ は x_i の外積を表す。

$$\begin{aligned}\sum_{i=1}^n -2\mathbf{x}_i' \Sigma^{-1} \mu &= -2 \left(\sum_{i=1}^n \mathbf{x}_i' \right) \Sigma^{-1} \mu \\ &= -2 \left(\sum_{i=1}^n \mathbf{x}_i' \right) \Sigma^{-1} \mu \\ \sum_{i=1}^n \mathbf{x}_i' &= (\mathbf{x}_1' + \cdots + \mathbf{x}_n') = (n\bar{\mathbf{x}})' \text{であるので} \\ &= -2(n\bar{\mathbf{x}})' \Sigma^{-1} \mu \\ &= -2n\bar{\mathbf{x}}' \Sigma^{-1} \mu\end{aligned}$$

$\sum_{i=1}^n \mu' \Sigma^{-1} \mu$ はスカラー値であり、これが n 回足されます。

$$\sum_{i=1}^n \mu' \Sigma^{-1} \mu = n \mu' \Sigma^{-1} \mu$$

したがって、これら3つの項を合わせると、指数部分は次のようになります。

$$\begin{aligned} \sum_{i=1}^n (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) &= \sum_{i=1}^n \mathbf{x}_i' \Sigma^{-1} \mathbf{x}_i - 2n \bar{\mathbf{x}}' \Sigma^{-1} \mu + n \mu' \Sigma^{-1} \mu \\ &= \text{tr}(\Sigma^{-1} \mathbf{X}' \mathbf{X}) - 2n \bar{\mathbf{x}}' \Sigma^{-1} \mu + n \mu' \Sigma^{-1} \mu \end{aligned}$$

$$\begin{aligned} & -\frac{1}{2} [\text{tr}(\Sigma^{-1}\mathbf{X}'\mathbf{X}) - 2\mathbf{n}\bar{\mathbf{x}}'\Sigma^{-1}\boldsymbol{\mu} + \mathbf{n}\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}] \\ & = -\frac{1}{2}\text{tr}(\Sigma^{-1}\mathbf{X}'\mathbf{X}) + \mathbf{n}\bar{\mathbf{x}}'\Sigma^{-1}\boldsymbol{\mu} - \frac{\mathbf{n}}{2}\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} \end{aligned}$$

となり、2行目の指数部分と完全に一致します。

$$\begin{aligned}\sum_{i=1}^n (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) &= \sum_{i=1}^n ((\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mu))' \Sigma^{-1} ((\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mu)) \\ &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + \sum_{i=1}^n (\bar{\mathbf{x}} - \mu)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu) \\ &\quad + 2 \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma^{-1} (\bar{\mathbf{x}} - \mu)\end{aligned}$$

3行目 (続き)

ここで、最後の項は

$$2 \left(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \right)' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) = 2(n\bar{\mathbf{x}} - n\bar{\mathbf{x}})' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) = 0$$

となるので、

$$\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$
$$\mathbf{V} = (n-1)\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

証明

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' &= \sum_{i=1}^n [x_i x_i' - x_i \bar{x}' - \bar{x} x_i' + \bar{x} \bar{x}'] \\&= \sum_{i=1}^n x_i x_i' - \sum_{i=1}^n x_i \bar{x}' - \sum_{i=1}^n \bar{x} x_i' + \sum_{i=1}^n \bar{x} \bar{x}' \\&= X'X - n\bar{x}\bar{x}' - n\bar{x}\bar{x}' + n\bar{x}\bar{x}' \\&= X'X - n\bar{x}\bar{x}' \\&= (n-1)S\end{aligned}$$

したがって、 $\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' = (n-1)S = V$ である。

証明（各項の詳細な計算）

- $\sum_{i=1}^n x_i x_i'$: 標本行列 X は各行が x_i' (x_i

は列ベクトル) として定義されるので、 $X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$ 。その転置は $X' = (x_1 \ x_2 \ \dots \ x_n)$

となる。したがって、 $X'X = \sum_{i=1}^n x_i x_i'$ 。

- $\sum_{i=1}^n x_i \bar{x}'$: \bar{x}' は和のインデックス i に依存しないので、 $\sum_{i=1}^n x_i \bar{x}' = (\sum_{i=1}^n x_i) \bar{x}' = n\bar{x}\bar{x}'$ 。
- $\sum_{i=1}^n \bar{x} x_i'$: \bar{x} も和のインデックス i に依存しないので、 $\sum_{i=1}^n \bar{x} x_i' = \bar{x}(\sum_{i=1}^n x_i') = n\bar{x}\bar{x}'$ 。
- $\sum_{i=1}^n \bar{x}\bar{x}'$: これは n 回足し合わせるので、 $n\bar{x}\bar{x}'$ 。

7.2 \bar{x} と S の同時分布

- $(X'X, \bar{x})$ (または (S, \bar{x}) のような一対一関数) は、 (Σ, μ) の最小十分完全統計量である。
- Rao-Blackwell/Lehmann-Schefféの定理により、不偏推定量の中で (\bar{x}, S) は最小分散を持つ。
- $n - 1 \geq p$ の場合の最尤推定量 (MLE) $\hat{\mu}$ と $\hat{\Sigma}$ を得るには、対数尤度関数を最小化する。

$$\ln |\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} + (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) \quad (7.2)$$

- 最尤推定量は $\hat{\mu} = \bar{x}$ と $\hat{\Sigma} = \frac{1}{n} V$ である。

定理 (Rao-Blackwell)

$\hat{\theta}$ がパラメータ θ の不偏推定量であり、 T が十分統計量であるとする。このとき、 $\hat{\theta}^* = E[\hat{\theta} \mid T]$ と定義すると、

- $\hat{\theta}^*$ は θ の不偏推定量である。
- $\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$ が成り立つ。

定理 (Lehmann-Scheffé)

T がパラメータ θ の完全かつ十分な統計量であるとする。もし $\hat{\theta}^* = g(T)$ が T の関数であり、かつ θ の不偏推定量であるならば、 $\hat{\theta}^*$ は θ の最小分散不偏推定量 (MVUE) である。

なぜ $(X'X, \bar{x})$ または (S, \bar{x}) なのか？

多変量正規分布の確率密度関数（尤度関数）の指数部分を見返すと、 μ と Σ を含む項が、 $X'X$ と \bar{x} の形で表現されていることがわかります。特に、

$$\text{etr} \left\{ -\frac{1}{2} [V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1} \right\}$$

という形で書けることから、観測されたデータ X の情報のうち、パラメータ μ と Σ に影響を与える部分は、本質的に V （または S ）と \bar{x} に集約されていることが読み取れます。これにより、これらが十分統計量であることが示唆

前提となる事実

- $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (i.i.d.)
- $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$
- $(n-1)S = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ [これは $(n-1)S = \mathbf{X}'\mathbf{X} - n\bar{\mathbf{x}}\bar{\mathbf{x}}'$ に等しい]
- $(\bar{\mathbf{x}}, S)$ が $(\boldsymbol{\Sigma}, \boldsymbol{\mu})$ に対して最小十分かつ完全な統計量である。

ステップ1: 不偏性 (Unbiasedness) の確認

a. \bar{x} が μ の不偏推定値であること

$$\begin{aligned} E[\bar{x}] &= E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] \\ &= \frac{1}{n} \sum_{i=1}^n E[x_i] \end{aligned}$$

各 x_i は $N_p(\mu, \Sigma)$ に従うため、 $E[x_i] = \mu$ です。

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \frac{1}{n} (n\mu) \\ &= \mu \end{aligned}$$

ステップ1: 不偏性 (Unbiasedness) の確認

b. S が Σ の不偏推定値であること

$$\begin{aligned} E[S] &= E \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \right] \\ &= \frac{1}{n-1} E \left[\sum_{i=1}^n (x_i x_i' - x_i \bar{x}' - \bar{x} x_i' + \bar{x} \bar{x}') \right] \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E[x_i x_i'] - \sum_{i=1}^n E[x_i \bar{x}'] - \sum_{i=1}^n E[\bar{x} x_i'] + \sum_{i=1}^n E[\bar{x} \bar{x}'] \right) \end{aligned}$$

ここで、各項を評価します。

- $E[x_i x_i'] = \text{Cov}(x_i) + E[x_i] E[x_i'] = \Sigma + \mu \mu'$.
- $E[\bar{x} \bar{x}'] = \text{Cov}(\bar{x}) + E[\bar{x}] E[\bar{x}'] = \frac{1}{n} \Sigma + \mu \mu' (\because \bar{x} \sim N_p(\mu, \frac{1}{n} \Sigma))$.
- $\sum_{i=1}^n E[x_i \bar{x}'] = E[(\sum_{i=1}^n x_i) \bar{x}'] = E[n \bar{x} \bar{x}'] = n E[\bar{x} \bar{x}'] = n(\frac{1}{n} \Sigma + \mu \mu') = \Sigma + n \mu \mu'$

ステップ1: 不偏性 (Unbiasedness) の確認 (続き)

b. S が Σ の不偏推定値であること (続き)

これらを $E[S]$ の式に代入します。

$$\begin{aligned} E[S] &= \frac{1}{n-1} \left(\sum_{i=1}^n (\Sigma + \mu\mu') - (\Sigma + n\mu\mu') - (\Sigma + n\mu\mu') + n\left(\frac{1}{n}\Sigma + \mu\mu'\right) \right) \\ &= \frac{1}{n-1} (n\Sigma + n\mu\mu' - \Sigma - n\mu\mu' - \Sigma - n\mu\mu' + \Sigma + n\mu\mu') \\ &= \frac{1}{n-1} ((n-1)\Sigma) \\ &= \Sigma \end{aligned}$$

したがって、 $E[S] = \Sigma$ であり、 S は Σ の不偏推定値です。

ステップ2: Rao-Blackwell / Lehmann-Scheffé の定理の適用

Lehmann-Schefféの定理の記述

Lehmann-Schefféの定理は、「もし、ある統計量 T が完全かつ十分 (Complete and Sufficient) であり、 $\hat{\theta}^* = g(T)$ が T の関数であり、かつパラメータ θ の不偏推定値であるならば、 $\hat{\theta}^*$ は θ の最小分散不偏推定量 (MVUE) である」と述べています。

定理の適用

- パラメータ θ は (μ, Σ) に対応します。
- 統計量 T は (\bar{x}, S) に対応します。テキストには、 $(X'X, \bar{x})$ (または (S, \bar{x}) のような1対1関数) が (Σ, μ) に対して最小十分かつ完全であることが述べられています。
- \bar{x} は $T = (\bar{x}, S)$ の関数 (具体的には第一成分) であり、ステップ1で μ の不偏推定値であることを示しました。
- S は $T = (\bar{x}, S)$ の関数 (具体的には第二成分) であり、ステップ1で Σ

上記の条件がすべて満たされるため、Lehmann-Schefféの定理により、 \bar{x} は μ のMVUEであり、 S は Σ のMVUEであると結論付けられます。したがって、 (\bar{x}, S) は (Σ, μ) のMVUEであると述べることができます。

最尤推定値 (MLE) の目的

- この式 $\ln|\Sigma| + \text{tr}\frac{1}{n}V\Sigma^{-1} + (\bar{x} - \mu)' \Sigma^{-1}(\bar{x} - \mu)$ (7.2) を最小化するのは、未知のパラメータである平均ベクトル μ と共分散行列 Σ の最尤推定値 (Maximum Likelihood Estimates, MLE) を求めるためです.
- 最尤推定法は、観測されたデータが最も「もっともらしい」と思われるようなパラメータ
- これを数学的に行うには、データの確率密度関数（または確率質量関数）をパラメータ

尤度関数から対数尤度関数へ

- 多変量正規分布の場合、観測された標本行列 X の同時確率密度関数（尤度関数）は、以下のような形をしていました:

$$f(X) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \text{etr} \left\{ -\frac{1}{2} [V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1} \right\} \quad (7.1)$$

-
- この尤度関数を直接最大化する代わりに、通常は計算が容易な対数尤度関数を最大化し
- 上記の確率密度関数に自然対数 \ln を取ると、以下のようになります:

$$\begin{aligned} \ln f(X) &= \ln \left((2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \text{etr} \left\{ -\frac{1}{2} [V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1} \right\} \right) \\ &= -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} ([V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1}) \end{aligned}$$

対数尤度関数の簡略化と最小化 (1)

- この対数尤度関数を $l(\Sigma, \mu)$ と表すとき、MLEを得るためには $l(\Sigma, \mu)$ を最大化する必要があります.
- ここで、定数項である $-\frac{np}{2}\ln(2\pi)$ はパラメータ Σ や μ に依存しないため、最大化には影響しません.
- したがって、最大化すべきは残りの項です:

$$l(\Sigma, \mu) \propto -\frac{n}{2}\ln|\Sigma| - \frac{1}{2}\text{tr}([V + n(\bar{x} - \mu)(\bar{x} - \mu)']\Sigma^{-1})$$

- この式を最大化することは、符号を反転させて最小化することと同じです.
- そして、全体を $\frac{n}{2}$ で割っても最大化/最小化の結果は変わらないため、以下の式を最小化することになり

対数尤度関数の簡略化と最小化 (2)

$$\begin{aligned} & \frac{1}{n} \text{tr} ([V + n(\bar{x} - \mu)(\bar{x} - \mu)'] \Sigma^{-1}) + \ln |\Sigma| \\ &= \frac{1}{n} \text{tr}(V \Sigma^{-1}) + \frac{1}{n} \text{tr}(n(\bar{x} - \mu)(\bar{x} - \mu)' \Sigma^{-1}) + \ln |\Sigma| \\ &= \frac{1}{n} \text{tr}(V \Sigma^{-1}) + \text{tr}((\bar{x} - \mu)(\bar{x} - \mu)' \Sigma^{-1}) + \ln |\Sigma| \\ &= \ln |\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} + (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) \quad (7.2) \end{aligned}$$

- (最後の項はスカラーなので tr を外すことができます。)
- この式は、テキストに示されている式 (7.2) と完全に一致します。

●

したがって、この式を最小化する目的は、観測されたデータの下で、母集団パラメータ μ と共分散行列 Σ

が最も「もっともらしい」値（すなわち最尤推定値）を見つけるためです。

ステップ1: $\hat{\mu} = \bar{x}$ の特定と最後の項の除去（まとめて解説）

- テキストにあるように、「（最後の項は ≥ 0 なので） $\hat{\mu} = \bar{x}$ であることは明らか」です。
- この「最後の項」とは、 $+(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)$ のことです。
- この項は、 Σ が正定値行列（つまり Σ^{-1} も正定値行列）であるため、常に0以上（ ≥ 0 ）です。
- この項を最小化するためには、その値を0にするのが最も小さい値です。
- $(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) = 0$ となるのは、 $\bar{x} - \mu = 0$ のとき、すなわち $\mu = \bar{x}$ のときです。
- したがって、 μ に関する最尤推定値 $\hat{\mu}$ は \bar{x} であると直ちに分かります。
- $\hat{\mu} = \bar{x}$ を元の式に代入すると、最後の項は0になります。

$$\begin{aligned} & \ln|\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} + (\bar{x} - \bar{x})' \Sigma^{-1} (\bar{x} - \bar{x}) \\ &= \ln|\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} + 0' \Sigma^{-1} 0 \\ &= \ln|\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} \end{aligned}$$

ステップ2: $\ln|\Sigma|$ の変形と V の導入

- 最小化すべき式：

$$\ln|\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1}$$

●

ここで、テキストでは「 $\ln|nV^{-1}\Sigma|$ 」という項が導入されています。これは、最小化の

- 行列式の性質を利用します。
 - $|AB| = |A||B|$
 - $|cA| = c^p|A|$ （ここで c はスカラー、 A は $p \times p$ 行列）
- $\ln|\Sigma|$ を V を含む形に変換するために、恒等式 $I = V^{-1}V$ を利用することを考えます。

$$\begin{aligned}\ln|\Sigma| &= \ln|V^{-1}V\Sigma| \\ &= \ln|V^{-1}(V\Sigma)| \\ &= \ln|nV^{-1}| + \ln\left|\frac{1}{n}V\Sigma\right|\end{aligned}$$

式変形のまとめ (1)

- $$\ln|nV^{-1}\Sigma| + \text{tr}\frac{1}{n}V\Sigma^{-1}$$

- これは、以下の恒等式（定数を追加・削除しても最小化の問題は変わらない）に基づいて

$$\ln|\Sigma| = \ln|nV^{-1}\Sigma| - \ln|nV^{-1}|$$

- ここで、 $\ln|nV^{-1}|$ は Σ に依存しない定数です。
- したがって、最小化すべき式 $\ln|\Sigma| + \text{tr}\frac{1}{n}V\Sigma^{-1}$ は、定数項 $\ln|nV^{-1}|$ を追加（または削除）しても、 Σ の最適値は変わりません。

式変形のまとめ (2)



$$\begin{aligned} & \ln|\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} \\ &= (\ln|nV^{-1}\Sigma| - \ln|nV^{-1}|) + \text{tr} \frac{1}{n} V \Sigma^{-1} \\ &= \ln|nV^{-1}\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1} - \ln|nV^{-1}| \end{aligned}$$

- テキストでは、この定数項 $-\ln|nV^{-1}|$ を「追加された定数」として無視し、以下の式を最小化することに焦点を当てています

$$\ln|nV^{-1}\Sigma| + \text{tr} \frac{1}{n} V \Sigma^{-1}$$

ステップ3: Σ の MLE の最終導出

定数 $\ln |nV^{-1}|$ は Σ の最適化には影響しません。条件 $n - 1 \geq p$ は V が確率1で非特異（正則）であることを保証します（これは後の系7.2で証明されます）。変数変換 $T = nV^{-1}\Sigma$ を導入すると、最小化すべき式は

$$\ln |T| + \text{tr}(T^{-1})$$

となります。この式は T のすべての固有値が1のとき最小値をとります。したがって、 $\hat{\Sigma} = \frac{1}{n}V$ が最尤推定値となります。

固有値による表現

- 対称行列 T の場合、そのトレースと行列式は固有値 $\lambda_1, \lambda_2, \dots, \lambda_p$ を用いて次のように表現できます。
 - $\text{tr}(T) = \sum_{j=1}^p \lambda_j$
 - $|T| = \prod_{j=1}^p \lambda_j$
- したがって、 $\ln|T| + \text{tr } T^{-1}$ は、固有値の関数として次のように書くことができます。
 - $\ln|T| = \ln(\prod_{j=1}^p \lambda_j) = \sum_{j=1}^p \ln(\lambda_j)$
 - $\text{tr}(T^{-1}) = \sum_{j=1}^p \frac{1}{\lambda_j}$ (T の固有値が λ_j なら、 T^{-1} の固有値は $1/\lambda_j$)
- これにより、最小化すべき関数は、各固有値 λ_j の関数として次のように分解できます。

$$f(\lambda_1, \dots, \lambda_p) = \sum_{j=1}^p \left(\ln(\lambda_j) + \frac{1}{\lambda_j} \right)$$

各固有値ごとの最小化 (1)

- この関数は、各固有値 λ_j について独立に最小化できます。

$$g(\lambda) = \ln(\lambda) + \frac{1}{\lambda}$$

- この関数 $g(\lambda)$ を最小化するために、 λ について微分し、導関数を0と置きます。

$$g'(\lambda) = \frac{d}{d\lambda}(\ln(\lambda) + \lambda^{-1})$$

$$g'(\lambda) = \frac{1}{\lambda} - \frac{1}{\lambda^2}$$

- $g'(\lambda) = 0$ とすると、

$$\frac{1}{\lambda} - \frac{1}{\lambda^2} = 0$$

$$\frac{1}{\lambda} = \frac{1}{\lambda^2}$$

$$\lambda^2 = \lambda$$

$$\lambda(\lambda - 1) = 0$$

各固有値ごとの最小化 (2)

- $\lambda = 0$ または $\lambda = 1$ 。
- しかし、行列式 $|T|$ は非ゼロでなければならない（ウィシャート分布の文脈で V は非特異なので）ため、固有値 λ は0ではありません。
- したがって、唯一の極値点は $\lambda = 1$ です。
- 二階微分を調べて、これが最小値であることを確認します。

$$g''(\lambda) = \frac{d}{d\lambda}(\lambda^{-1} - \lambda^{-2})$$

$$g''(\lambda) = -\lambda^{-2} + 2\lambda^{-3} = -\frac{1}{\lambda^2} + \frac{2}{\lambda^3}$$

- $\lambda = 1$ のとき、 $g''(1) = -1 + 2 = 1 > 0$
なので、これは極小値であり、唯一の最小値です。

各固有値ごとの最小化 (3) および結論

- 関数 $f(\lambda_1, \dots, \lambda_p)$ は、すべての固有値 λ_j が 1 のときに最小値を取ります。
- すべての固有値が1である対称行列は、単位行列 I だけです。
- よって、 $T=I$ のとき最小値となり、 $nV^{-1}\hat{\Sigma} = I$ 、すなわち $\hat{\Sigma} = \frac{1}{n}V$ となります。
- これが多変量正規分布における共分散行列の最尤推定値です。

備考

Gaussに遡るよく知られた結果として、 \mathbb{R} 上の確率密度関数 $f(x - \theta)$ で、 x が θ の最尤推定量 (MLE) となる唯一の場所族は正規密度に由来します。

この正規密度のMLE特性は \mathbb{R}^p でも成り立ちます [Stadje (1993)]。すなわち、 $f(x - \theta)$ という形の密度で、 x が常に θ のMLEとなるのは正規分布の場合のみです。

$\bar{x} \sim N_p(\mu, \frac{1}{n}\Sigma)$ が「明らか」な理由 (1)



この記述が「明らか」とされるのは、多変量正規分布の線形変換と標本平均の性質に関

- 前提: 各観測ベクトル x_i は独立同分布 (i.i.d.) で $x_i \sim N_p(\mu, \Sigma)$ に従います。

- 標本平均の定義: 標本平均 \bar{x} は、 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ です。

- 正規分布の和の性質:

- 独立な正規分布に従う確率変数の和も正規分布に従います。
- $E[\sum x_i] = \sum E[x_i] = \sum \mu = n\mu$ 。
- $\text{Cov}(\sum x_i) = \sum \text{Cov}(x_i) = \sum \Sigma = n\Sigma$ (独立性の仮定による)。
- したがって、 $\sum_{i=1}^n x_i \sim N_p(n\mu, n\Sigma)$ です。

$\bar{x} \sim N_p(\mu, \frac{1}{n}\Sigma)$ が「明らか」な理由 (2)

- 正規分布の定数倍の性質: $c \cdot Y \sim N_p(c \cdot E[Y], c^2 \cdot \text{Cov}(Y))$ です。
 - ここで、 $Y = \sum x_i$ 、 $c = \frac{1}{n}$ とすると、

$$\bar{x} = \frac{1}{n} \sum x_i \sim N_p \left(\frac{1}{n}(n\mu), \left(\frac{1}{n}\right)^2 (n\Sigma) \right)$$

$$\bar{x} \sim N_p \left(\mu, \frac{1}{n^2} n\Sigma \right)$$

$$\bar{x} \sim N_p \left(\mu, \frac{1}{n} \Sigma \right)$$

標本平均の変換

- \bar{x} は x_i の標本平均です。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- $x_i = Az_i + \mu$ を代入すると、

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n (Az_i + \mu) \\ &= \frac{1}{n} \left(A \sum_{i=1}^n z_i + \sum_{i=1}^n \mu \right) \\ &= \frac{1}{n} (A(n\bar{z}) + n\mu) \\ &= A\bar{z} + \mu\end{aligned}$$

- したがって、 \bar{x} の分布は $A\bar{z} + \mu$ の分布と一致します。

標本分散共分散行列の変換 (1/2)

- S_x は x_i の標本分散共分散行列です。

$$(n-1)S_x = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

- $x_i - \bar{x}$ の部分を変換します。

$$\begin{aligned}x_i - \bar{x} &= (Az_i + \mu) - (A\bar{z} + \mu) \\&= Az_i - A\bar{z} \\&= A(z_i - \bar{z})\end{aligned}$$

- これを標本分散共分散行列の定義に代入します。

$$\begin{aligned}(n-1)S_x &= \sum_{i=1}^n (A(z_i - \bar{z}))(A(z_i - \bar{z}))' \\&= \sum_{i=1}^n A(z_i - \bar{z})(z_i - \bar{z})'A'\end{aligned}$$

標本共分散行列の変換 (1)

- 標本共分散行列 S は

$$(n-1)S_z = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

で定義されます。

- $x_i = Az_i + \mu$ および $\bar{x} = A\bar{z} + \mu$ を代入すると、

$$x_i - \bar{x} = A(z_i - \bar{z})$$

- よって、

$$\begin{aligned}(n-1)S &= \sum_{i=1}^n A(z_i - \bar{z})(z_i - \bar{z})'A' \\ &= A \left(\sum_{i=1}^n (z_i - \bar{z})(z_i - \bar{z})' \right) A' \\ &= A(n-1)S_z A'\end{aligned}$$

- したがって、 S の分布は $AS_z A'$ の分布と一致します。

7.3 Wishart分布の性質 - 命題

命題 7.3

$W \sim W_p(m)$ かつ $m \geq p$ ならば、 W は確率1で非特異である。

$W \stackrel{d}{=} Z'Z$ であり、 $Z' = (z_1, \dots, z_m)$ かつ z_i は独立同分布の $N_p(0, I)$ に従う。

$\text{rank } W \stackrel{d}{=} \text{rank } Z'Z = \text{rank } Z \geq \text{rank } (z_1, \dots, z_p)$ は確率1で p となる。

したがって、 $\text{rank } W$ は確率1で p となる。

7.2 \bar{x} と S の同時分布 (続き)

- 一般的な結果として、 \bar{x} は $N_p(\mu, \Sigma/n)$ に従う。
- 標本行列 X を Z を用いて表す。 $X \stackrel{d}{=} ZA' + 1\mu'$, ここで $Z \sim N_n^p(0, I_n \otimes I_p)$ かつ $\Sigma = AA'$ 。
- \bar{x} と S_x の分布は \bar{x} と S_z の分布に等しい。
- $P = n^{-1}11'$ と $Q = I - n^{-1}11'$ は直交射影である。
- $PZ \perp QZ$ であるため、 $\bar{z} \perp S_z$ である。
- $Q = HH'$ は直交基底を与えるため、 $(n-1)S_z = Z'HH'Z = U'U$ となる。

7.2 \bar{x} と S の同時分布 (続き)

定義 7.1 Wishart分布

$W \sim W_p(m)$ ならば $W \stackrel{d}{=} \sum_{i=1}^m z_i z_i'$, ここで z_i は独立同分布で $N_p(0, I)$ に従う。

$V \sim W_p(m, \Sigma)$ ならば $V \stackrel{d}{=} AWA'$, ここで $\Sigma = AA'$ かつ $W \sim W_p(m)$ 。

命題 7.1

x_i が独立同分布で $N_p(\mu, \Sigma)$ に従う場合 ($i = 1, \dots, n$)、

- $\bar{x} \sim N_p(\mu, \Sigma/n)$
- $(n-1)S \sim W_p(n-1, \Sigma)$
- $\bar{x} \perp S$

7.3 Wishart分布の性質 - 補題

補題 7.1

$Z = (z_{ij}) \in \mathbb{R}^{n \times n}$ が独立同分布の $N(0, 1)$ に従う場合、 $P(|Z| = 0) = 0$ 。

$n = 1$ の場合は z_{11} が絶対連続分布を持つため、結果は成立する。
 Z を以下のように分割する。

$$Z = \begin{pmatrix} z_{11} & z'_{12} \\ z_{21} & Z_{22} \end{pmatrix}$$

$Z_{22} \in \mathbb{R}^{(n-1) \times (n-1)}$ に対して結果が成立すると仮定すると、

$$\begin{aligned} P(|Z| = 0) &= P(|Z| = 0, |Z_{22}| \neq 0) + P(|Z| = 0, |Z_{22}| = 0) \\ &= P(z_{11} = z'_{12} Z_{22}^{-1} z_{21}, |Z_{22}| \neq 0) \\ &= E[P(z_{11} = z'_{12} Z_{22}^{-1} z_{21}, |Z_{22}| \neq 0 | z_{12}, z_{21}, Z_{22})] = 0 \end{aligned}$$

7.3 Wishart分布の性質 - 系

系 7.1

$Z = (z_{ij}) \in \mathbb{R}^{n \times n}$ が独立同分布の $N(0, 1)$ に従う場合、 $P(|Z| = t) = 0, \forall t$ 。

$$P(|Z| = t) = E[P(z_{11} = z'_{12} Z_{22}^{-1} z_{21} + t/|Z_{22}|, |Z_{22}| \neq 0 | z_{12}, z_{21}, Z_{22})] = 0.$$

補題7.1と系7.1は、 Z が任意の絶対連続分布を持つ場合にも有効である。

7.3 Wishart分布の性質 - 命題

命題 7.3

$W \sim W_p(m)$ かつ $m \geq p$ ならば、 W は確率1で非特異である。

$W \stackrel{d}{=} Z'Z$ であり、 $Z' = (z_1, \dots, z_m)$ かつ z_i は独立同分布の $N_p(0, I)$ に従う。

$\text{rank } W \stackrel{d}{=} \text{rank } Z'Z = \text{rank } Z \geq \text{rank } (z_1, \dots, z_p)$ は確率1で p となる。

したがって、 $\text{rank } W$ は確率1で p となる。