Wishart分布とその性質

山北倫太郎

July 9, 2025

目次

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、$x_1, \ldots, x_n$ 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 。
- 7.3節では、Wishart分布の基本的な特性 。

7.2 x̄ 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。

$$f(X) = (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right] \tag{7.1}$$

$$= (2\pi)^{-np/2}|\Sigma|^{-n/2}e^{-\frac{n}{2}\mu'\Sigma^{-1}\mu}\mathrm{etr}\left[-\frac{1}{2}\mathrm{tr}\,\Sigma^{-1}X'X + n\mu'\Sigma^{-1}\bar{x}\right]$$

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

1行目

確率変数ベクトル $X = (x_1, \ldots, x_p)'$ 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。

2行目

$$
\begin{aligned}
&(\mathsf{x_i} - \mu)'\Sigma^{-1}(\mathsf{x_i} - \mu) \\
&= (\mathsf{x_i}' - \mu')\Sigma^{-1}(\mathsf{x_i} - \mu) \\
&= \mathsf{x_i}'\Sigma^{-1}\mathsf{x_i} - \mathsf{x_i}'\Sigma^{-1}\mu - \mu'\Sigma^{-1}\mathsf{x_i} + \mu'\Sigma^{-1}\mu \\
&= \mathsf{x_i}'\Sigma^{-1}\mathsf{x_i} - 2\mathsf{x_i}'\Sigma^{-1}\mu + \mu'\Sigma^{-1}\mu
\end{aligned}
$$

ここで、$\mu'\Sigma^{-1}\mathsf{x_i}$
はスカラー値であり、スカラーの転置は自分自身なので、$\mu'\Sigma^{-1}\mathsf{x_i} = (\mathsf{x_i}'\Sigma^{-1}\mu)'$ です。
したがって、$\mathsf{x_i}'\Sigma^{-1}\mu$ と $\mu'\Sigma^{-1}\mathsf{x_i}$ 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列の行ベクトルになります。
Σ^{-1}は共分散行列Σの逆行列を表し、Σはp*pの正方行列なので、
Σ^{-1}もp*pの正方行列になります。
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、$(\mathsf{x}_\mathsf{i} - \mu)'\Sigma^{-1}(\mathsf{x}_\mathsf{i} - \mu)$はスカラー値であります。

また、$\mu'\Sigma^{-1}\mu$は、$\mathsf{e}^{-\frac{1}{2}\mathsf{n}\mu'\Sigma^{-1}\mu}$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、$\mathrm{tr}(A) = \sum_i a_{ii}$ で定義される。

トレースの性質

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

$$
\begin{aligned}
\sum_{i=1}^{n} x_i' \Sigma^{-1} x_i &= \sum_{i=1}^{n} \mathrm{tr}\left(\Sigma^{-1} x_i x_i'\right) \\
&= \mathrm{tr}\left(\Sigma^{-1} \sum_{i=1}^{n} x_i x_i'\right) \\
&= \mathrm{tr}(\Sigma^{-1} X' X)
\end{aligned}
$$

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。

第2項目

$$\sum_{i=1}^{n} -2x_i'\Sigma^{-1}\mu = -2(\sum_{i=1}^{n} x_i')\Sigma^{-1}\mu$$

$$= -2\left(\sum_{i=1}^{n} x_i'\right)\Sigma^{-1}\mu$$

$$\sum_{i=1}^{n} x_i' = (x_1' + \cdots + x_n') = (n\bar{x})'\text{であるので}$$

$$= -2(n\bar{x})'\Sigma^{-1}\mu$$

$$= -2n\bar{x}'\Sigma^{-1}\mu$$

第3項目

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

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。

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

まとめると

$$
\begin{aligned}
&-\frac{1}{2}\left[\mathrm{tr}(\Sigma^{-1}\mathsf{X}'\mathsf{X}) - 2\mathsf{n}\bar{\mathsf{x}}'\Sigma^{-1}\mu + \mathsf{n}\mu'\Sigma^{-1}\mu\right] \\
&= -\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\mathsf{X}'\mathsf{X}) + \mathsf{n}\bar{\mathsf{x}}'\Sigma^{-1}\mu - \frac{\mathsf{n}}{2}\mu'\Sigma^{-1}\mu
\end{aligned}
$$

となり、2行目の指数部分と完全に一致します。

3行目

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

3行目 (続き)

ここで、最後の項は

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

となるので、

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

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

証明

証明

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

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 $X' = (x_1 \ x_2 \ \ldots \ x_n)$ 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。

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

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

Rao-Blackwellの定理とLehmann-Schefféの定理

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、$\hat{\theta}^* = \mathsf{E}[\hat{\theta} \mid \mathsf{T}]$ と定義すると、

- $\hat{\theta}^*$ は θ の不偏推定量である。
- $\mathrm{Var}(\hat{\theta}^*) \leq \mathrm{Var}(\hat{\theta})$ 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 $\hat{\theta}^* = \mathsf{g}(\mathsf{T})$ 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、

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

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 $\bar{x}$ に集約されています。

多変量正規分布は指数族の一つである

- 多変量正規分布の確率密度関数を指数族の形で表現してみましょう。
- 確率密度関数は次のように書けます：

$$
\begin{aligned}
\mathsf{f}(\mathsf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= (2\pi)^{-\frac{\mathsf{p}}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathsf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathsf{x}-\boldsymbol{\mu})\right\} \\
&= (2\pi)^{-\frac{\mathsf{p}}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathsf{x}'\boldsymbol{\Sigma}^{-1}\mathsf{x}+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathsf{x}-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}
\end{aligned}
$$

多変量正規分布を指数族の形に書き換える

- 上記の密度関数を指数族の一般形に変形します：

$$
\begin{aligned}
\mathsf{f}(\mathsf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= (2\pi)^{-\frac{\mathsf{p}}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathsf{x}'\boldsymbol{\Sigma}^{-1}\mathsf{x}+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathsf{x}-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\} \\
&= \underbrace{(2\pi)^{-\frac{\mathsf{p}}{2}}}_{\mathsf{h}(\mathsf{x})}\exp\left\{\underbrace{\begin{pmatrix}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1} \\ -\frac{1}{2}\boldsymbol{\Sigma}^{-1}\end{pmatrix}'}_{\boldsymbol{\eta}(\boldsymbol{\theta})'}\underbrace{\begin{pmatrix}\mathsf{x} \\ \mathsf{x}\mathsf{x}'\end{pmatrix}}_{\mathsf{T}(\mathsf{x})}-\underbrace{\left(\frac{1}{2}\ln|\boldsymbol{\Sigma}|+\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)}_{\mathsf{A}(\boldsymbol{\theta})}\right\}
\end{aligned}
$$

i.i.d.標本における十分統計量

- 独立同分布な標本 $x_1, x_2, \ldots, x_n$ における同時確率密度関数は：

$$f(x_1, \ldots, x_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(x_i | \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) \right\}$$

- これを展開すると：

$$= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}(\boldsymbol{\Sigma}^{-1} X' X) + n\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \bar{x} - \frac{n}{2} \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\}$$

i.i.d.標本の同時分布を指数族の形で表現

- 同時分布を指数族の形に書き換えると：

$$
\begin{aligned}
& f(x_1, \ldots, x_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \underbrace{(2\pi)^{-\frac{np}{2}}}_{h(x)} \exp\left\{ \underbrace{\begin{pmatrix} n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1} \\ -\frac{1}{2}\boldsymbol{\Sigma}^{-1} \end{pmatrix}'}_{\boldsymbol{\eta}(\boldsymbol{\theta})'} \underbrace{\begin{pmatrix} \bar{x} \\ X'X \end{pmatrix}}_{T(x)} - \underbrace{\left(\frac{n}{2} \ln |\boldsymbol{\Sigma}| + \frac{n}{2} \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \right)}_{A(\boldsymbol{\theta})} \right\}
\end{aligned}
$$

- このことから、$T(x) = \begin{pmatrix} \bar{x} \\ X'X \end{pmatrix}$ が十分統計量であることがわかります。
- また、S と $X'X$ には次の関係があります：

$$
(n-1)S = X'X - n\bar{x}\bar{x}'
$$

- これより、$(S, \bar{x})$ も十分統計量となります。

指数族に対するLehmann-Schefféの定理

Lehmann-Schefféの定理の特別なケース

指数族のモデルにおいて、十分統計量 $T(x)$ がパラメータ $\boldsymbol{\theta}$ の関数 $g(\boldsymbol{\theta})$ の不偏推定量である場合、その推定量は最小分散不偏推定量（MVUE）である。

- 多変量正規分布の場合、$(\bar{x}, S)$ はそれぞれ $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の不偏推定量であることが示されました。
- 指数族である多変量正規分布では、$(\bar{x}, S)$ は十分統計量であるため、Lehmann-Schefféの定理により、$\bar{x}$ は $\boldsymbol{\mu}$ のMVUEであり、S は $\boldsymbol{\Sigma}$ のMVUEであると結論づけることができます。

前提となる事実

- $\mathrm{x_i} \sim \mathrm{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (i.i.d.)
- $\bar{\mathrm{x}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}$
- $(\mathrm{n}-1)\mathrm{S} = \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{x_i} - \bar{\mathrm{x}})(\mathrm{x_i} - \bar{\mathrm{x}})'$ [これは $(\mathrm{n}-1)\mathrm{S} = \mathrm{X'X} - \mathrm{n}\bar{\mathrm{x}}\bar{\mathrm{x}}'$ に等しい]
- $(\bar{\mathrm{x}}, \mathrm{S})$ 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 $\boldsymbol{\mu}$ 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 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うため、$E[x_i] = \boldsymbol{\mu}$ です。

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

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

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

- $E[x_ix_i'] = \mathrm{Cov}(x_i) + E[x_i]E[x_i'] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'$。
- $E[\overline{x}\overline{x}'] = \mathrm{Cov}(\overline{x}) + E[\overline{x}]E[\overline{x}'] = \frac{1}{n}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'$ $(\because \overline{x} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}))$。
- $\sum_{i=1}^{n} E[x_i\overline{x}'] = E[(\sum_{i=1}^{n} x_i)\overline{x}'] = E[n\overline{x}\overline{x}'] = nE[\overline{x}\overline{x}'] = n(\frac{1}{n}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') = \boldsymbol{\Sigma} + n\boldsymbol{\mu}\boldsymbol{\mu}'$。

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。

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

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

結論

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 $\ln|\boldsymbol{\Sigma}| + \mathrm{tr}\frac{1}{n}\mathrm{V}\boldsymbol{\Sigma}^{-1} + (\bar{\mathrm{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathrm{x}} - \boldsymbol{\mu})$ (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 $\boldsymbol{\mu}$ と共分散行列 $\boldsymbol{\Sigma}$ 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の値を推定する統計的手法です.
- 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）をパラメータの関数と見なした「尤度関数」を最大化します.

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:

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

- この尤度関数を直接最大化する代わりに、対数尤度関数を最大化します.
- 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:

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

対数尤度関数の簡略化と最小化 (1)

- この対数尤度関数を $\mathrm{l}(\boldsymbol{\Sigma}, \boldsymbol{\mu})$ と表すとき、MLEを得るためには $\mathrm{l}(\boldsymbol{\Sigma}, \boldsymbol{\mu})$ を最大化する必要があります.
- ここで、定数項である $-\frac{\mathrm{np}}{2}\ln(2\pi)$ はパラメータ $\boldsymbol{\Sigma}$ や $\boldsymbol{\mu}$ 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:

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

- 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 $\frac{\mathrm{n}}{2}$ 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)

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

- （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
  $\boldsymbol{\mu}$ と共分散行列 $\boldsymbol{\Sigma}$
  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: $\hat{\mu} = \bar{\mathrm{x}}$ 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）$\hat{\mu} = \bar{\mathrm{x}}$ であることは明らか」です。
- この「最後の項」とは、$+(\bar{\mathrm{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathrm{x}} - \boldsymbol{\mu})$ のことです。
- この項は、$\boldsymbol{\Sigma}$ が正定値行列（つまり $\boldsymbol{\Sigma}^{-1}$ 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。
- $(\bar{\mathrm{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathrm{x}} - \boldsymbol{\mu}) = 0$ となるのは、$\bar{\mathrm{x}} - \boldsymbol{\mu} = 0$ のとき、すなわち $\boldsymbol{\mu} = \bar{\mathrm{x}}$ のときです。
- したがって、$\boldsymbol{\mu}$ に関する最尤推定値 $\hat{\mu}$ は $\bar{\mathrm{x}}$ であると直ちに分かります。
- $\hat{\mu} = \bar{\mathrm{x}}$ を元の式に代入すると、最後の項は0になります。

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

ステップ2: $\ln|\Sigma|$ の変形と V の導入

- 最小化すべき式：

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

- ここで、テキストでは「$\ln|nV^{-1}\boldsymbol{\Sigma}|$」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）
- $\ln|\boldsymbol{\Sigma}|$ 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。

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

式変形のまとめ (1)

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

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

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

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

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

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 $\lambda_1, \lambda_2, \ldots, \lambda_p$ を用いて次のように表現できます。
  - $\mathrm{tr}(T) = \sum_{j=1}^{p} \lambda_j$
  - $|T| = \prod_{j=1}^{p} \lambda_j$
- したがって、$\ln|T| + \mathrm{tr}\ T^{-1}$ は、固有値の関数として次のように書くことができます。
  - $\ln|T| = \ln(\prod_{j=1}^{p} \lambda_j) = \sum_{j=1}^{p} \ln(\lambda_j)$
  - $\mathrm{tr}(T^{-1}) = \sum_{j=1}^{p} \frac{1}{\lambda_j}$ （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。

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

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、

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

各固有値ごとの最小化 (2)

- $\lambda(\lambda-1)=0$より、$\lambda=0$ 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ではありえません。
- 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。

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

- $\lambda=1$ 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 $f(\lambda_1, \ldots, \lambda_p)$ 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。

正規分布のMLE特性と場所族

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 $\mathrm{f}(\mathrm{x}-\theta)$ 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 $\mathbb{R}^{\mathrm{p}}$ でも成り立ちます [Stadje (1993)]。 すなわち、$\mathrm{f}(\mathrm{x}-\theta)$ 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。

$\bar{x} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma)$ 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 $x_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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。
    - $E[\sum x_i] = \sum E[x_i] = \sum \boldsymbol{\mu} = n\boldsymbol{\mu}$。
    - $\mathrm{Cov}(\sum x_i) = \sum \mathrm{Cov}(x_i) = \sum \boldsymbol{\Sigma} = n\boldsymbol{\Sigma}$（独立性の仮定による）。
    - したがって、$\sum_{i=1}^{n} x_i \sim N_p(n\boldsymbol{\mu}, n\boldsymbol{\Sigma})$ です。

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

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

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

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

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

標本平均の変換

- $\bar{x}$ は x_i の標本平均です。

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

- $x_i = Az_i + \boldsymbol{\mu}$ を代入すると、

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

- したがって、$\bar{x}$ の分布は $A\bar{z} + \boldsymbol{\mu}$ 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。

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

- これを標本分散共分散行列の定義に代入します。

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

$$= \sum_{i=1}^{n} A(z_i - \bar{z})(z_i - \bar{z})'A'$$

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

標本共分散行列の変換 (2)

- 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_zA' の分布と一致します。

「⊥⊥」の意味

- 数学や統計学において、「⊥⊥」という記号は、統計的独立性 (statistical independence) を意味します.
- A ⊥⊥ B と書かれた場合、それは確率変数（またはベクトル、行列）A と B が互いに統計的に独立であることを意味します.
- 統計的独立性とは、一方の変数の値が分かっても、もう一方の変数の値に関する情報か
- つまり、2つの事象や変数が互いに影響し合わない状態を指します.

PZ $\perp\!\!\!\perp$ QZ であることが明らかである理由

- PZ と QZ が独立（$\perp\!\!\!\perp$）である理由は、P と Q が直交射影行列であり、$\mathsf{P}+\mathsf{Q}=\mathsf{I}$ かつ $\mathsf{PQ}=\mathsf{QP}=0$ という性質に基づきます。
- Z の各列は独立な標準正規ベクトルなので、PZ（平均成分）と QZ（偏差成分）はともに正規分布に従います。
- PZ と QZ の和は Z そのものであり、PZ と QZ は直交する部分空間への射影なので、$\mathrm{Cov}(\mathsf{PZ},\mathsf{QZ})=0$ が成り立ちます。
- 多変量正規分布では、無相関（共分散ゼロ）は独立と同値なので、PZ と QZ は独立です。
- したがって、PZ $\perp\!\!\!\perp$ QZ となることは「明らか」です。

$\mathsf{PZ} \perp\!\!\!\perp \mathsf{QZ}$ であることが明らかである理由 (続き)

- 直交射影: P と Q は、互いに直交する射影行列です.
  - $\mathsf{P}^2 = \mathsf{P}$ かつ $\mathsf{Q}^2 = \mathsf{Q}$ （射影性）
  - $\mathsf{PQ} = \mathsf{QP} = 0$ （直交性）.
    これは、平均への射影と平均からの偏差への射影が互いに直交していることを意味します

したがって $\bar{x} \perp\!\!\!\perp S_x$ の理由

- これは、独立性が線形変換によって保たれるという性質に基づきます。
- 先に示したように、$x_i = Az_i + \boldsymbol{\mu}$ から $(\bar{x}, S_x) = (A\bar{z} + \boldsymbol{\mu}, AS_zA')$ となります。
- $\bar{x}$ は $\bar{z}$ の線形変換、S_x は S_z の線形変換です。
- $\bar{z} \perp\!\!\!\perp S_z$ であるため、$\bar{x}$ と S_x も独立です。
- よって、$\bar{x} \perp\!\!\!\perp S_x$ が成り立ちます。

$\bar{z} \perp\!\!\!\perp S_z$ の理由

- 多変量標準正規分布において、標本平均 $\bar{z}$ と標本分散共分散行列 S_z は独立です。
- これは、PZ（平均成分）と QZ（偏差成分）が直交射影行列 P と Q による射影であり、$PZ \perp\!\!\!\perp QZ$ が成り立つためです。
- $\bar{z}$ は PZ の線形変換、S_z は QZ のみから構成されるため、PZ と QZ の独立性から $\bar{z}$ と S_z も独立となります。
-

  この性質は、正規分布の直交射影による分割が独立性をもたらすことに由来します。

$n\bar{z} = Z'P\mathbf{1}$ の導出

- $n\bar{z} = Z'P\mathbf{1}$ の理由:
    - 標本平均ベクトル $\bar{z}$ の定義は $\bar{z} = n^{-1}Z'\mathbf{1}$ です。
    - ここで、$P = n^{-1}\mathbf{1}\mathbf{1}'$ は平均への射影行列です。
    - $Z'P\mathbf{1} = Z'(n^{-1}\mathbf{1}\mathbf{1}')\mathbf{1} = n^{-1}Z'\mathbf{1}(\mathbf{1}'\mathbf{1})$ となります。
    - $\mathbf{1}'\mathbf{1} = n$ なので、$Z'P\mathbf{1} = n^{-1}Z'\mathbf{1} \cdot n = Z'\mathbf{1}$ です。
    - よって、$n\bar{z} = Z'\mathbf{1} = Z'P\mathbf{1}$ となります。

$n\bar{z} = Z'P1$ と $(n-1)S_z = Z'QQZ$ の導出（続き）

- $(n-1)S_z = Z'QQZ$ の理由:
  - 標本分散共分散行列 S_z の定義は

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

  - $Q = I - n^{-1}\mathbf{1}\mathbf{1}'$ は、平均ベクトルへの射影を除去する直交射影行列です。
  - $Qz_i = z_i - \bar{z}$ となるので、QZ は各行から平均を引いた行列になります。
  - よって、QZ の i 行目は $(z_i - \bar{z})'$ です。
  - QZ の各行を縦に並べたものが QZ、その転置が $(QZ)' = Z'Q$ です。
  - したがって、

$$(n-1)S_z = (QZ)'(QZ) = Z'Q'QZ$$

  - Q は射影行列なので $Q' = Q$ かつ $Q^2 = Q$ です。よって

$$(n-1)S_z = Z'QQZ$$

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 。
(Z’ H = U’)

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 $(i = 1, \ldots, n)$、

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

命題 7.2：ウィシャート行列のトレース

- $W \sim W_p(m)$ のトレースの分布は、定義からほぼ直接的に導かれます。

命題 7.2

$W \sim W_p(m) \Rightarrow \mathrm{tr}\, W \sim \chi^2_{mp}$

- 証明. $W_p(m)$ の定義より、

$$\mathrm{tr}\, W = \mathrm{tr}\left(\sum_{i=1}^{m} z_i z_i'\right) = \sum_{i=1}^{m} z_i' z_i,$$

ここで z_i は i.i.d. $N_p(0, I)$。
- 各 $z_i' z_i \sim \chi^2_p$ であり、独立なので和 $\sum_{i=1}^{m} \chi^2_p = \chi^2_{mp}$。(命題4.4を参照)
- よって $\mathrm{tr}\, W \sim \chi^2_{mp}$ となる。(系3.1を参照)

系 3.1：ガンマ分布の加法性

- 系 3.1 もし $x_i, i = 1, \ldots, n,$ が独立に $G(p_i, \theta)$ に従うならば、$\sum_{i=1}^n x_i \sim G\left(\sum_{i=1}^n p_i, \theta\right)$ である[1].
- 特別なケースとして、$p = 1$ の場合、指数分布が得られます。

[1] 例えば、χ_k^2 分布は $G(k/2, 2)$ である。

命題 4.4：カイ二乗分布の定義とモーメント

- 命題 4.4 $y \sim \chi^2_m \iff y \overset{d}{=} \sum_{i=1}^m z_i^2$, ここで $z_1, ..., z_m$ は i.i.d. $N(0, 1)$ に従う。
- 命題 4.3 の表現により、(積分) モーメントを素早く導出できます:
  - n が奇数の場合: $E[z^n] = 0$（すなわち $z^n \overset{d}{=} -z^n$）
  - n が偶数の場合: $E[z^n] = E[z^{2k}] = 2^k E[w^k]$, ここで $w \sim G(\frac{1}{2})$
  - $= 2^k \Gamma(k + \frac{1}{2}) / \Gamma(\frac{1}{2})$
  - $= (n-1)(n-3) \cdots 3 \cdot 1$
- 特に、$E[z] = 0$ および $\mathrm{var}[z] = 1$ です。
-

  より一般的な正規分布は、単純に位置を移動し、スケールを変更することで得られます

補題 7.1：ランダム行列の非特異性

補題 7.1

- 補題 7.1 $Z = (z_{ij}) \in \mathbb{R}^{n \times n}$ で z_{ij} が独立同分布 $N(0, 1)$ に従うとき、$P(|Z| = 0) = 0$ である。

補題 7.1：証明

- 証明. 帰納法で示す。
- $n = 1$ の場合、z_{11} は絶対連続分布なので $P(z_{11} = 0) = 0$。
- Z を $Z = \begin{pmatrix} z_{11} & z'_{12} \\ z_{21} & Z_{22} \end{pmatrix}$ と分割し、$Z_{22} \in \mathbb{R}^{(n-1)\times(n-1)}$ で結果が成り立つと仮定する。
- このとき、

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

- z_{11} は絶対連続分布なので、特定の値をとる確率は0である。

ステップ1: $n = 1$ の場合

- $n = 1$ の行列 Z は、ただ1つの要素 z_{11} からなる行列です。$Z = (z_{11})$
- このとき、行列式 $|Z| = z_{11}$ です。
- z_{11} は $N(0, 1)$
  という正規分布に従うランダムな値です。正規分布は連続的な分布なので、
  ある特定の値（この場合は0）をちょうど取る確率はゼロです。
  (例えば、身長がぴったり170.000...cmになる確率はゼロ、というのと同じです。)
- だから、$P(|Z| = 0) = 0$ は $n = 1$ の場合に真です。

ステップ2: $n-1$ の場合が正しいと仮定し、n の場合を示す (1/3)

- 次に、「もしサイズが $(n-1)\times(n-1)$ の行列 Z_{22} の行列式が0になる確率がゼロだとしたら、サイズ $n\times n$ の行列 Z でも同じことが言えるか？」と考えます。
- 行列 Z を次のように分割します:

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

  - z_{11}: 左上の1つの要素（スカラー）
  - z'_{12}: 1行目から z_{11} を除いた行ベクトル
  - z_{21}: 1列目から z_{11} を除いた列ベクトル
  - Z_{22}: 残りの右下の $(n-1)\times(n-1)$ 行列
- 行列式 $|Z|$ を計算する一つの方法は、ブロック行列の行列式公式を使うことです。
- もし $|Z_{22}|\neq 0$ ならば、$|Z| = |Z_{22}|\cdot(z_{11} - z'_{12}Z_{22}^{-1}z_{21})$ となります。

ステップ2: $n-1$ の場合が正しいと仮定し、n の場合を示す (2/3)

- さて、$P(|Z|=0)$ を求めたいのですが、これを2つの部分に分けます:

$$P(|Z|=0) = P(|Z|=0, |Z_{22}| \neq 0) + P(|Z|=0, |Z_{22}|=0)$$

- $P(|Z|=0, |Z_{22}|=0)$ の部分:
  - 帰納法の仮定により、$P(|Z_{22}|=0)=0$ です。
  - したがって、$P(|Z|=0, |Z_{22}|=0)$
    というイベントの確率は0になります（なぜなら、これは $|Z_{22}|=0$
    という確率0のイベントの一部だからです）。
- $P(|Z|=0, |Z_{22}| \neq 0)$ の部分:
  - この部分が重要です。$|Z|=0$ かつ $|Z_{22}| \neq 0$
    となるためには、先ほどの行列式の公式から、$z_{11} - z_{12}' Z_{22}^{-1} z_{21} = 0$
    つまり、$z_{11} = z_{12}' Z_{22}^{-1} z_{21}$ となる必要があります。

ステップ2: $n-1$ の場合が正しいと仮定し、n の場合を示す (3/3)

- この確率を計算するために、条件付き確率の考え方を使います。

$$E[P(z_{11} = z_{12}' Z_{22}^{-1} z_{21}, |Z_{22}| \neq 0 | z_{12}, z_{21}, Z_{22})]$$

- これは、(z_{12}, z_{21}, Z_{22}) の値がすでに決まっているという条件下で、z_{11} が特定の値 $z_{12}' Z_{22}^{-1} z_{21}$ を取る確率を計算し、その期待値を取る、ということです。
- ここで、z_{11} は $N(0,1)$ に従う独立な確率変数です。そして、$z_{12}' Z_{22}^{-1} z_{21}$ という値は、(z_{12}, z_{21}, Z_{22}) が決まれば定数になります。
- 前述の $n=1$ の場合と同じ理由で、連続分布に従う確率変数（z_{11}）が、ある特定の定数とぴったり一致する確率はゼロです。
- したがって、$P(z_{11} = z_{12}' Z_{22}^{-1} z_{21} | z_{12}, z_{21}, Z_{22}) = 0$ です。
- この条件付き確率が0なので、その期待値も0になります。
- これにより、$P(|Z| = 0, |Z_{22}| \neq 0) = 0$ です。

最終的な結論

- $\mathrm{P}(|\mathrm{Z}| = 0) = 0 + 0 = 0$
- これで、サイズ n
  の場合も行列式が0になる確率がゼロであることが示されました。
- 帰納法により、この結果はすべての自然数 n に対して成り立ちます。

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、

$$P(|Z| = 0) = P(|Z| = 0, |Z_{22}| \neq 0) + P(|Z| = 0, |Z_{22}| = 0)$$

$$= P(z_{11} = z_{12}' Z_{22}^{-1} z_{21}, |Z_{22}| \neq 0)$$

$$= E[P(z_{11} = z_{12}' Z_{22}^{-1} z_{21}, |Z_{22}| \neq 0 | z_{12}, z_{21}, Z_{22})] = 0$$

系 7.1：一般化された行列式の確率

系 7.1

系 7.1 $Z = (z_{ij}) \in \mathbb{R}^{n\times n}$ で z_{ij} が i.i.d. $N(0,1) \Rightarrow P(|Z| = t) = 0, \forall t.$

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

備考

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ことに注意すべきです。これで、[Stein (1969), Dykstra (1970)] の証明ができます.

系 7.1の証明における項の増加理由 (1/2)

- 系7.1の証明における $P(|Z| = t)$ の式で、$\frac{t}{|Z_{22}|}$ の項が増える理由についてですね。
- これは、行列式に関する同じブロック行列の公式を使用し、$|Z| = t$ という条件を式に組み込んでいるためです。
- 補題7.1の証明と同じく、行列 Z を次のようにブロック分割します:

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

- そして、右下のブロック Z_{22} が正則である場合（$|Z_{22}| \neq 0$ の場合）の行列式の公式を使います。

$$|Z| = |Z_{22}| \cdot (z_{11} - z_{12}' Z_{22}^{-1} z_{21})$$

- 系7.1では、$P(|Z| = t)$ という確率を求めています。つまり、行列式 $|Z|$ が特定の値 t になる確率です。

系 7.1の証明における項の増加理由 (2/2)

- したがって、$|Z| = t$
  という条件を前のフレームの公式に代入すると、次のようになります。

$$t = |Z_{22}| \cdot (z_{11} - z'_{12} Z_{22}^{-1} z_{21})$$

- この式を z_{11} について解くと、

$$z_{11} - z'_{12} Z_{22}^{-1} z_{21} = \frac{t}{|Z_{22}|}$$

$$z_{11} = z'_{12} Z_{22}^{-1} z_{21} + \frac{t}{|Z_{22}|}$$

- このように、$P(|Z| = t)$ を計算する際には、z_{11}
  が他のブロックによって定まる値に加えて、t の値と $|Z_{22}|$ の逆数に比例する項 $\frac{t}{|Z_{22}|}$
  を取る必要があるため、この項が増えるのです。
- 補題7.1が $|Z| = 0$ の場合を扱っていたのに対し、系7.1は $|Z| = t$
  というより一般的なケースを扱っているため、この追加の項が必要になります。

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、$Z' = (z_1, \ldots, z_m)$ かつ z_i は独立同分布の $N_p(0, I)$ に従う。

$\mathrm{rank}\, W \stackrel{d}{=} \mathrm{rank}\, Z'Z = \mathrm{rank}\, Z \geq \mathrm{rank}\, (z_1, \ldots, z_p)$ は確率1で p となる。
したがって、$\mathrm{rank}\, W$ は確率1で p となる。

命題 7.3：証明の概要 (1/2)

- 証明は、W の定義とランクの性質を利用しています。
- $W = Z'Z$ という表現:
  - ウィシャート分布の定義（定義7.1）によれば、$W = \sum_{i=1}^{m} z_i z_i'$ です。ここで z_i は i.i.d. $N_p(0, I)$ です。
  - この W は、$Z'Z$ の形で書くことができます。ここで Z' は $z_1, \dots, z_m$ を列ベクトルとして並べた行列です。

$$Z' = \begin{pmatrix} z_1 & z_2 & \dots & z_m \end{pmatrix}$$

  - すると、Z は $m \times p$ 行列です。
- ランクの等式:
  - 一般に、行列 A に対して $\mathrm{rank}(A'A) = \mathrm{rank}(A)$ という性質があります。
  - したがって、$\mathrm{rank}\, W = \mathrm{rank}\, Z'Z = \mathrm{rank}\, Z$ となります。

命題 7.3：証明の概要 (2/2)

- ランクの下限:
  - Z は $m \times p$ 行列であり、その各要素は独立な正規分布に従います。
  - rank Z は、Z
    の列ベクトルのうち線形独立なものの最大数、または行ベクトルのうち線形独立なものの
  - Z の列ベクトルは $z_1, \ldots, z_m$ です。
  - $\mathrm{rank}\, Z \geq \mathrm{rank}(z_1, \ldots, z_p)$ というのは、Z の最初の p 個の列ベクトル $(z_1, \ldots, z_p)$
    からなる行列のランクを考えたものです。
  - 補題7.1や系7.1で証明されたように、独立な正規分布の要素からなる行列が特異になる確
    p 行列 $(z_1, \ldots, z_p)$ は、確率1で非特異、つまりランクが p
    になります。つまり、$\mathrm{rank}(z_1, \ldots, z_p) = p$ です。
  - このため、$\mathrm{rank}\, W \geq p$ となります。
  - しかし、W は $p \times p$ 行列なので、rank W は最大でも p です。
  - したがって、$\mathrm{rank}\, W = p$ が確率1で成立する、つまり W
    は確率1で非特異である、と結論できます。