

証明 公式 2.3 を確かめてみましょう. 式 (2.38) の積分は $\boldsymbol{\mu}_1$ の場所によりませんので*6,

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right) \quad (2.39)$$

すなわち,

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \quad (2.40)$$

と定義すれば

$$p \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \propto \exp \left(-\frac{1}{2} \underbrace{\left\{ \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \right\}}_L \right) \quad (2.41)$$

の場合について証明すれば十分です. このとき $\int p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2$ を計算すればいいのですが, 基本的な戦略は, 積分したいベクトル $\mathbf{x}_2$ について平方完成することです. 式 (2.41) の右辺を $\exp(-\frac{1}{2}L)$ とおくと, $\boldsymbol{\Lambda}^T = \boldsymbol{\Lambda}$ なので

$$L = \mathbf{x}_1^T \boldsymbol{\Lambda}_{11} \mathbf{x}_1 + \mathbf{x}_1^T \boldsymbol{\Lambda}_{12} \mathbf{x}_2 + \mathbf{x}_2^T \boldsymbol{\Lambda}_{21} \mathbf{x}_1 + \mathbf{x}_2^T \boldsymbol{\Lambda}_{22} \mathbf{x}_2 \quad (2.42)$$

$$= \mathbf{x}_1^T \boldsymbol{\Lambda}_{11} \mathbf{x}_1 + 2\mathbf{x}_2^T \boldsymbol{\Lambda}_{21} \mathbf{x}_1 + \mathbf{x}_2^T \boldsymbol{\Lambda}_{22} \mathbf{x}_2 \quad (2.43)$$

と表すことができます.

ここで一般に, 対称行列 $\boldsymbol{\Lambda}$ について

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Lambda} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Lambda} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Lambda} \boldsymbol{\mu} \quad (2.44)$$

$$= \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} - 2\mathbf{x}^T \boldsymbol{\Lambda} \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\Lambda} \boldsymbol{\mu} \quad (2.45)$$

が成り立つことに注意しましょう. 式 (2.45) では, $(\boldsymbol{\mu}^T \boldsymbol{\Lambda}) \mathbf{x} = \mathbf{x}^T (\boldsymbol{\mu}^T \boldsymbol{\Lambda})^T = \mathbf{x}^T \boldsymbol{\Lambda} \boldsymbol{\mu}$ であることを用いました. よって L は, 次の形に平方完成することができます.

$$L = (\mathbf{x}_2 + \boldsymbol{\Lambda}_{22}^{-1} \boldsymbol{\Lambda}_{21} \mathbf{x}_1)^T \boldsymbol{\Lambda}_{22} (\mathbf{x}_2 + \cdots) - \mathbf{x}_1^T \boldsymbol{\Lambda}_{21}^T \boldsymbol{\Lambda}_{22}^{-1} \boldsymbol{\Lambda}_{21} \mathbf{x}_1 + \mathbf{x}_1^T \boldsymbol{\Lambda}_{11} \mathbf{x}_1 \quad (2.46)$$

式 (2.46) の第 1 項は, 多変量ガウス分布の正規化定数を求める式 (2.24) と

*6 数学的には, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu} + \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ なので, $\int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} = \boldsymbol{\mu} + \int \mathcal{N}(\mathbf{x}|\mathbf{0}, \boldsymbol{\Sigma}) d\mathbf{x}$ となるためです.

同じ形をしており, $\mathbf{x}_2$ について積分すると真ん中の $\mathbf{\Lambda}_{22}$ だけに依存する定数となることに注意しましょう. よって,

$$p(\mathbf{x}_1) = \int p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = \int \exp\left(-\frac{1}{2}L\right) d\mathbf{x}_2 \quad (2.47)$$

$$= (\text{定数}) \times \exp\left(-\frac{1}{2} \left\{ -\mathbf{x}_1^T \mathbf{\Lambda}_{21}^T \mathbf{\Lambda}_{22}^{-1} \mathbf{\Lambda}_{21} \mathbf{x}_1 + \mathbf{x}_1^T \mathbf{\Lambda}_{11} \mathbf{x}_1 \right\}\right) \quad (2.48)$$

$$\propto \exp\left(-\frac{1}{2} \mathbf{x}_1^T (\mathbf{\Lambda}_{11} - \mathbf{\Lambda}_{21}^T \mathbf{\Lambda}_{22}^{-1} \mathbf{\Lambda}_{21}) \mathbf{x}_1\right) \quad (2.49)$$

となります. これを公式 2.2 と見比べれば, $\mathbf{x}_1$ の分布は, $\mathbf{\Lambda}$ で表せば

$$p(\mathbf{x}_1) = \mathcal{N}(\mathbf{0}, (\mathbf{\Lambda}_{11} - \mathbf{\Lambda}_{21}^T \mathbf{\Lambda}_{22}^{-1} \mathbf{\Lambda}_{21})^{-1}) \quad (2.50)$$

となることがわかります.

最後に, 式 (2.50) を $\mathbf{\Sigma}$ で表してみましょう. $\mathbf{\Lambda}$ の定義は式 (2.40) でしたから, 付録「行列の分割と逆行列」の式 (1) から,

$$(\mathbf{\Lambda}_{11} - \mathbf{\Lambda}_{21}^T \mathbf{\Lambda}_{22}^{-1} \mathbf{\Lambda}_{21})^{-1} = \mathbf{\Sigma}_{11} \quad (2.51)$$

が成り立っています. よって, 式 (2.50) を $\mathbf{\Sigma}$ で表せば

$$p(\mathbf{x}_1) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{11}) \quad (2.52)$$

となります. □

2.3.4 多変量ガウス分布の条件つき分布

ベクトル $\mathbf{x}$ が多変量ガウス分布に従うとき, $\mathbf{x}$ の一部の次元を固定したときの分布, すなわちある次元で「切った」ときの切り口の分布も, やはりガウス分布になります. たとえば図 2.7 のように, 2 次元のガウス分布 $p(x_1, x_2)$ において, $x_1=c$ (定数) で切ったときの x_2 の分布 $p(x_2|x_1)$ は, 値 c によって形の異なるガウス分布になります.

このことを, 式で表してみましょう. D 次元のガウス分布に従うベクトル

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma}) \quad (2.53)$$

を, 前節の式 (2.37) と同様に $\mathbf{x}_1, \mathbf{x}_2$ と分けて表すことにします. このとき, $\mathbf{x}_1$ を固定したときの $\mathbf{x}_2$ の分布, すなわち同時分布 $p(\mathbf{x}_1, \mathbf{x}_2)$ を $\mathbf{x}_1$ で切っ

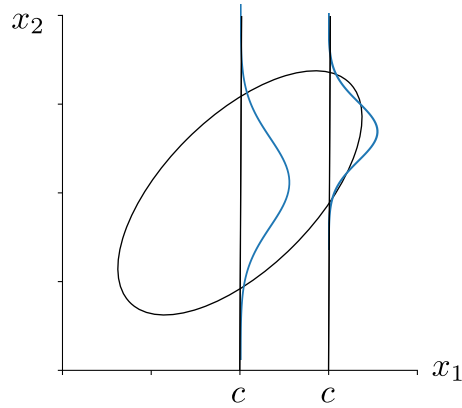


図 2.7 多変量ガウス分布の条件つき分布を表した図. 楕円で示した 2 次元のガウス分布 $p(x_1, x_2)$ において, $x_1=c$ を固定したときの分布 $p(x_2|x_1)$ は, c によって異なる平均および分散のガウス分布となり, 式 (2.54) で与えられます.

たときの $\mathbf{x}_2$ の条件つき分布は

公式 2.4 (多変量ガウス分布の条件つき分布)

$$p(\mathbf{x}_2|\mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) \quad (2.54)$$

となります. 若干計算が面倒ですが, この公式は前節と同様に条件部の $\mathbf{x}_1$ について平方完成することで, 以下のような計算から得られます.*⁷

証明 条件つき確率 $p(\mathbf{x}_2|\mathbf{x}_1)$ は, $\mathbf{x}_2$ の関数としては同時確率 $p(\mathbf{x}_1, \mathbf{x}_2)$ に比例していますので,

$$p(\mathbf{x}_2|\mathbf{x}_1) \propto p(\mathbf{x}_1, \mathbf{x}_2) \quad (2.55)$$

$$\propto \exp\left(-\frac{1}{2}\left\{\begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}\right\}\right) \quad (2.56)$$

*⁷ [59] の付録 A11.3 にも, 別の興味深い導出が載っています.

$$\begin{aligned}
&= \exp \left(-\frac{1}{2} \left\{ (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{11} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{12} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \right. \right. \\
&\quad \left. \left. (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\} \right) \\
&\propto \exp \left(-\frac{1}{2} \left\{ (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + 2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{21} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\} \right) \\
&\propto \exp \left(-\frac{1}{2} \left\{ \mathbf{x}_2^T \boldsymbol{\Lambda}_{22} \mathbf{x}_2 - \mathbf{x}_2^T \boldsymbol{\Lambda}_{22} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_2^T \boldsymbol{\Lambda}_{22} \mathbf{x}_2 + 2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{21} \mathbf{x}_2 \right\} \right) \\
&\propto \exp \left(-\frac{1}{2} \left\{ \mathbf{x}_2^T \boldsymbol{\Lambda}_{22} \mathbf{x}_2 - 2\mathbf{x}_2^T (\boldsymbol{\Lambda}_{22} \boldsymbol{\mu}_2 - \boldsymbol{\Lambda}_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1)) \right\} \right) \\
&\propto \exp \left(-\frac{1}{2} \left\{ (\mathbf{x}_2 - \boldsymbol{\Lambda}_{22}^{-1} (\boldsymbol{\Lambda}_{22} \boldsymbol{\mu}_2 - \boldsymbol{\Lambda}_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1)))^T \boldsymbol{\Lambda}_{22} (\mathbf{x}_2 - \dots) \right\} \right)
\end{aligned} \tag{2.57}$$

と平方完成することができます。よって $\mathbf{x}_2$ の分布は、 $\boldsymbol{\Lambda}$ で表せば

$$p(\mathbf{x}_2 | \mathbf{x}_1) \sim \mathcal{N}(\boldsymbol{\Lambda}_{22}^{-1} (\boldsymbol{\Lambda}_{22} \boldsymbol{\mu}_2 - \boldsymbol{\Lambda}_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1)), \boldsymbol{\Lambda}_{22}^{-1}) \tag{2.58}$$

$$= \mathcal{N}(\boldsymbol{\mu}_2 - \boldsymbol{\Lambda}_{22}^{-1} \boldsymbol{\Lambda}_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Lambda}_{22}^{-1}) \tag{2.59}$$

となることがわかります。

後は、上を $\boldsymbol{\Sigma}$ を使って表すだけです。 $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ でしたから、付録「行列の分割と逆行列」の式 (2) より、 $\mathbf{M} = (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{21})^{-1}$ とおいて

$$\begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \tag{2.60}$$

$$= \begin{pmatrix} \cdots & \cdots \\ -\mathbf{M} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{M} \end{pmatrix} \tag{2.61}$$

が成り立っています。よって、

$$\boldsymbol{\Lambda}_{22} = (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{21})^{-1} \tag{2.62}$$

$$\boldsymbol{\Lambda}_{22}^{-1} \boldsymbol{\Lambda}_{21} = -\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \tag{2.63}$$

ですから、式 (2.59) は $\boldsymbol{\Sigma}$ を使って表せば、

$$p(\mathbf{x}_2 | \mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{21}) \tag{2.64}$$

となります。□