行列輪講: 第 8 回 ガウス分布 2

杉浦 圭祐

慶應義塾大学理工学部情報工学科 松谷研究室

September 30, 2023

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

このスライドの概要

- ガウス分布について確認する
  - Isserlis の定理
- 以下の資料を参考に作成しました:
  - パターン認識と機械学習 (上巻)
  - State Estimation For Robotics
- 重要な分布なので, 考えることがたくさんある

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

Isserlis の定理

- $\mathbf{x} = (x_1, \ldots, x_D)$ が, 平均 $\mathbf{0}$ のガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \mathbf{\Sigma})$ に従うとき, Isserlis の定理が成り立つ:

$$\mathbb{E}\left[x_1 x_2 \ldots x_n\right] = \sum_{p \in P_n^2} \prod_{(i,j) \in p} \mathbb{E}\left[x_i x_j\right] = \sum_{p \in P_n^2} \prod_{(i,j) \in p} \mathrm{Cov}\left(x_i, x_j\right)$$

- 総和: $\{1, \ldots, n\}$ を, 互いに素な $\dfrac{n}{2}$ 個のペアに分割する方法について
- $n = 4$ であれば, 以下の 3 通りある:

$$\{1,2,3,4\} \to \{\{1,2\},\{3,4\}\},\ \{\{1,3\},\{2,4\}\},\ \{\{1,4\},\{2,3\}\} \equiv P_4^2$$

- $n = 4$ のときは, Isserlis の定理は次のようになる:

$$\mathbb{E}\left[x_1 x_2 x_3 x_4\right] = \mathbb{E}\left[x_1 x_2\right]\mathbb{E}\left[x_3 x_4\right] + \mathbb{E}\left[x_1 x_3\right]\mathbb{E}\left[x_2 x_4\right] + \mathbb{E}\left[x_1 x_4\right]\mathbb{E}\left[x_2 x_3\right]$$

Isserlis の定理

- $\mathbf{x} = (x_1, \ldots, x_D)$ が, 平均 $\mathbf{0}$ のガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \mathbf{\Sigma})$ に従うとき, Isserlis の定理が成り立つ:

$$\mathbb{E}\left[x_1 x_2 \ldots x_n\right] = \sum_{p \in P_n^2} \prod_{(i,j) \in p} \mathbb{E}\left[x_i x_j\right] = \sum_{p \in P_n^2} \prod_{(i,j) \in p} \mathrm{Cov}\left(x_i, x_j\right)$$

- 総和: $\{1, \ldots, n\}$ を, 互いに素な $\frac{n}{2}$ 個のペアに分割する方法について
- n が奇数 $(n = 2m + 1)$ のときは, ペアを作れないので, 0 になる.
- n が偶数 $(n = 2m)$ であれば, 項の総数 (分割の方法) は

$$\frac{(2m)!}{(2^m m!)} = (2m-1)(2m-3) \cdots 3 \cdot 1 = (2m-1)!!$$

- $n = 4$ のとき 3, $n = 6$ のとき 15, $n = 8$ のとき 105 個の項が出現する.

Isserlis の定理

- 簡単な例として, x がガウス分布 $\mathcal{N}(x \mid 0, \sigma^2)$ に従うとする.
- $\mathbb{E}\left[x^4\right]$ は次のように分かる:

$$\mathbb{E}\left[x^4\right] = 3\,\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right] = 3\sigma^2\sigma^2 = 3\sigma^4$$

- $\mathbb{E}\left[x^6\right]$ は

$$\mathbb{E}\left[x^6\right] = 15\,\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right] = 15\sigma^2\sigma^2\sigma^2 = 15\sigma^6$$

- $\mathbb{E}\left[x^8\right]$ は

$$\mathbb{E}\left[x^8\right] = 105\,\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right] = 105\sigma^2\sigma^2\sigma^2\sigma^2 = 105\sigma^8$$

- 奇数の n に対しては, $\mathbb{E}\left[x^n\right] = 0$.
- 分布の平均は 0 だから, $\mathbb{E}\left[x^2\right] = \mathrm{Var}\left[x\right] + \mathbb{E}\left[x\right]^2 = \mathrm{Var}\left[x\right] = \sigma^2$.

Isserlis の定理の応用 1

- 続いて, $\mathbf{x} = (x_1, \ldots, x_D)$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \mathbf{\Sigma})$ に従うとする.
- 分布の平均は $\mathbf{0}$ だから, 次が成り立つことに注意しよう:

$$\begin{aligned} \mathrm{Cov}\left(x_i, x_j\right) &= \mathbb{E}\left[\left(x_i - \mathbb{E}\left[x_i\right]\right)\left(x_j - \mathbb{E}\left[x_j\right]\right)\right] = \mathbb{E}\left[x_i x_j\right] \\ \mathrm{Var}\left[x_i\right] &= \mathrm{Cov}\left(x_i, x_i\right) = \mathbb{E}\left[x_i^2\right] \end{aligned}$$

- このとき, 共分散 $\mathbf{\Sigma}$ の (i, j) 成分は, $\mathrm{Cov}\left(x_i, x_j\right) = \mathbb{E}\left[x_i x_j\right]$ となる:

$$\left(\mathbf{\Sigma}\right)_{ij} = \mathrm{Cov}\left(x_i, x_j\right) = \mathbb{E}\left[x_i x_j\right]$$

- 以上を踏まえて, 期待値 $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top \mathbf{x}\mathbf{x}^\top\right]$ を求めよう.

Isserlis の定理の応用 1

- 期待値 $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\mathbf{x}\mathbf{x}^\top\right]$ を求めよう. (i,j) 成分は

$$\mathbb{E}\left[\left(\mathbf{x}\left(\mathbf{x}^\top\mathbf{x}\right)\mathbf{x}^\top\right)_{ij}\right] = \mathbb{E}\left[x_i\left(\sum_{k=1}^{D}x_k^2\right)x_j\right] = \sum_k \mathbb{E}\left[x_i x_j x_k^2\right]$$

$$= \sum_k \mathbb{E}\left[x_i x_j\right]\mathbb{E}\left[x_k^2\right] + 2\,\mathbb{E}\left[x_i x_k\right]\mathbb{E}\left[x_j x_k\right]$$

$$= \mathbb{E}\left[x_i x_j\right]\left(\sum_k \mathbb{E}\left[x_k^2\right]\right) + 2\sum_k \mathbb{E}\left[x_i x_k\right]\mathbb{E}\left[x_k x_j\right]$$

$$= (\boldsymbol{\Sigma})_{ij}\left(\sum_k (\boldsymbol{\Sigma})_{kk}\right) + 2\sum_k (\boldsymbol{\Sigma})_{ik}(\boldsymbol{\Sigma})_{kj} = (\boldsymbol{\Sigma})_{ij}\,\mathrm{tr}(\boldsymbol{\Sigma}) + 2\left(\boldsymbol{\Sigma}^2\right)_{ij}$$

であるから, $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\mathbf{x}\mathbf{x}^\top\right] = \boldsymbol{\Sigma}\,\mathrm{tr}(\boldsymbol{\Sigma}) + 2\boldsymbol{\Sigma}^2 = \boldsymbol{\Sigma}\left(\mathrm{tr}(\boldsymbol{\Sigma}) + 2\boldsymbol{\Sigma}\right).$

Isserlis の定理の応用 2

- 続いて，期待値 $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top \mathbf{A}\mathbf{x}\mathbf{x}^\top\right]$ を求めよう．(i,j) 成分は

$$
\begin{aligned}
\mathbb{E}\left[\left(\mathbf{x}\left(\mathbf{x}^\top \mathbf{A}\mathbf{x}\right)\mathbf{x}^\top\right)_{ij}\right] &= \mathbb{E}\left[x_i\left(\sum_{k=1}^{D}\sum_{l=1}^{D} x_k a_{kl} x_l\right)x_j\right] \\
&= \sum_k \sum_l a_{kl}\,\mathbb{E}\left[x_i x_j x_k x_l\right] \\
&= \sum_k \sum_l a_{kl}\left(\mathbb{E}\left[x_i x_j\right]\mathbb{E}\left[x_k x_l\right] + \mathbb{E}\left[x_i x_k\right]\mathbb{E}\left[x_j x_l\right] + \mathbb{E}\left[x_i x_l\right]\mathbb{E}\left[x_j x_k\right]\right) \\
&= \mathbb{E}\left[x_i x_j\right]\sum_k \sum_l a_{kl}\,\mathbb{E}\left[x_l x_k\right] \\
&\quad + \sum_k \sum_l \mathbb{E}\left[x_i x_k\right] a_{kl}\,\mathbb{E}\left[x_l x_j\right] + \sum_k \sum_l \mathbb{E}\left[x_i x_l\right] a_{kl}\,\mathbb{E}\left[x_k x_j\right]
\end{aligned}
$$

Isserlis の定理の応用 2

- $\mathbb{E}[x_i x_j] = (\mathbf{\Sigma})_{ij}$ を使って，式変形を続けると，

$$
\begin{aligned}
&\mathbb{E}[x_i x_j] \sum_k \sum_l a_{kl}\, \mathbb{E}[x_l x_k] \\
&\quad + \sum_k \sum_l \mathbb{E}[x_i x_k]\, a_{kl}\, \mathbb{E}[x_l x_j] + \sum_k \sum_l \mathbb{E}[x_i x_l]\, a_{kl}\, \mathbb{E}[x_k x_j] \\
&= (\mathbf{\Sigma})_{ij} \sum_k \sum_l a_{kl} (\mathbf{\Sigma})_{lk} \\
&\quad + \sum_k \sum_l (\mathbf{\Sigma})_{ik}\, a_{kl}\, (\mathbf{\Sigma})_{lj} + \sum_k \sum_l (\mathbf{\Sigma})_{il} \left(\mathbf{A}^\top\right)_{lk} (\mathbf{\Sigma})_{kj} \\
&= (\mathbf{\Sigma})_{ij} \sum_k (\mathbf{A\Sigma})_{kk} + (\mathbf{\Sigma A \Sigma})_{ij} + \left(\mathbf{\Sigma A}^\top \mathbf{\Sigma}\right)_{ij} \\
&= (\mathbf{\Sigma})_{ij} \operatorname{tr}(\mathbf{A\Sigma}) + (\mathbf{\Sigma A \Sigma})_{ij} + \left(\mathbf{\Sigma A}^\top \mathbf{\Sigma}\right)_{ij}
\end{aligned}
$$

Isserlis の定理の応用 2

- 期待値 $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top \mathbf{A}\mathbf{x}\mathbf{x}^\top\right]$ の (i, j) 成分は

$$\mathbb{E}\left[\left(\mathbf{x}\left(\mathbf{x}^\top \mathbf{A}\mathbf{x}\right)\mathbf{x}^\top\right)_{ij}\right] = (\boldsymbol{\Sigma})_{ij}\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + (\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma})_{ij} + \left(\boldsymbol{\Sigma}\mathbf{A}^\top\boldsymbol{\Sigma}\right)_{ij}$$

となるから,

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top \mathbf{A}\mathbf{x}\mathbf{x}^\top\right] &= \boldsymbol{\Sigma}\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^\top\boldsymbol{\Sigma} \\ &= \boldsymbol{\Sigma}\left(\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\mathbf{I} + \left(\mathbf{A} + \mathbf{A}^\top\right)\boldsymbol{\Sigma}\right)\end{aligned}$$

Isserlis の定理の応用 3

- 続いて, $\mathbf{x} = (x_1, \ldots, x_D)$ が標準正規分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \mathbf{I})$ に従うとする.
- $\mathbf{x}$ の各成分は互いに独立であり, x_i は $\mathcal{N}(x_i \mid 0, 1)$ に従う.
- このとき, $z = \mathbf{x}^\top \mathbf{x} = \sum_{i=1}^{D} x_i^2$ は, 自由度 D のカイ二乗分布に従う.
- 期待値は D, 分散は $2D$ になる.
- 期待値 $\mathbb{E}[z] = \mathbb{E}\left[\mathbf{x}^\top \mathbf{x}\right]$ を求めよう:

$$\mathbb{E}\left[\mathbf{x}^\top \mathbf{x}\right] = \mathbb{E}\left[\sum_i x_i^2\right] = \sum_i \mathbb{E}\left[x_i^2\right] = \sum_i 1 = D$$

- 分布の平均は 0 だから, $\mathbb{E}\left[x_i^2\right] = \mathrm{Var}\left[x_i\right] + \mathbb{E}\left[x_i\right]^2 = 1$.

Isserlis の定理の応用 3

- 続いて，分散 $\mathrm{Var}\left[z\right] = \mathbb{E}\left[\left(z - \mathbb{E}\left[z\right]\right)^2\right]$ を求めよう ($z = \mathbf{x}^\top \mathbf{x}$):

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathbf{x}^\top \mathbf{x} - \mathbb{E}\left[\mathbf{x}^\top \mathbf{x}\right]\right)^2\right] = \mathbb{E}\left[\left(\mathbf{x}^\top \mathbf{x} - D\right)^2\right] = \mathbb{E}\left[\left(\sum_i x_i^2 - D\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_i x_i^2\right)^2 - 2D\left(\sum_i x_i^2\right) + D^2\right] \\
&= \sum_i \sum_j \mathbb{E}\left[x_i^2 x_j^2\right] - 2D \sum_i \mathbb{E}\left[x_i^2\right] + D^2 \\
&= \sum_i \sum_j \left(\mathbb{E}\left[x_i^2\right] \mathbb{E}\left[x_j^2\right] + 2\,\mathbb{E}\left[x_i x_j\right]^2\right) - 2D \sum_i 1 + D^2
\end{aligned}
$$

- 最後の式変形で，Isserlis の定理を用いた．

Isserlis の定理の応用 3

- さらに式変形を続けると，自由度 D のカイ二乗分布の分散は，

$$
\begin{aligned}
&= \sum_i \sum_j \left(\mathbb{E}\left[x_i^2\right] \mathbb{E}\left[x_j^2\right] + 2\,\mathbb{E}\left[x_i x_j\right]^2 \right) - 2D \sum_i 1 + D^2 \\
&= \sum_i \sum_{j \neq i} \left(\mathbb{E}\left[x_i^2\right] \mathbb{E}\left[x_j^2\right] + 2\,\mathbb{E}\left[x_i x_j\right]^2 \right) + \sum_i 3\,\mathbb{E}\left[x_i^2\right]^2 - 2D^2 + D^2 \\
&= \sum_i \sum_{j \neq i} (1 \cdot 1 + 2 \cdot 0^2) + 3 \sum_i 1^2 - D^2 \\
&= \sum_i \sum_{j \neq i} 1 + 3 \sum_i 1 - D^2 \\
&= D(D-1) + 3D - D^2 = 2D
\end{aligned}
$$

- $\mathbf{x}$ の各成分は独立だから，$\mathbb{E}\left[x_i x_j\right] = \mathbb{E}\left[x_i\right] \mathbb{E}\left[x_j\right]$．平均は 0 なので，$\mathbb{E}\left[x_i x_j\right] = 0$．また，$\mathbb{E}\left[x_i^2\right] = 1$．

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

条件付きガウス分布と周辺ガウス分布

- $\mathbf{x}, \mathbf{y}$ の同時分布が, ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}\right)$$

- まずは, 形から確認してみよう.

条件付きガウス分布と周辺ガウス分布

- $\mathbf{x}, \mathbf{y}$ の同時分布が，ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}\right)$$

- $\mathbf{x}$ を M 次元，$\mathbf{y}$ を N 次元とすると，$\boldsymbol{\mu}_x$ は M 次，$\boldsymbol{\mu}_y$ は N 次ベクトル．
- $\boldsymbol{\Sigma}_{xx}$ は $M \times M$，$\boldsymbol{\Sigma}_{yy}$ は $N \times N$，$\boldsymbol{\Sigma}_{xy}$ は $M \times N$，$\boldsymbol{\Sigma}_{yx}$ は $N \times M$ 行列．

$$\begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} = \mathbb{E}\left[\begin{pmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - \boldsymbol{\mu}_y \end{pmatrix}\begin{pmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - \boldsymbol{\mu}_y \end{pmatrix}^\top\right] = \begin{pmatrix} \mathrm{Cov}\left(\mathbf{x}, \mathbf{x}\right) & \mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) \\ \mathrm{Cov}\left(\mathbf{y}, \mathbf{x}\right) & \mathrm{Cov}\left(\mathbf{y}, \mathbf{y}\right) \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{xx} = \mathbb{E}\left[\left(\mathbf{x} - \boldsymbol{\mu}_x\right)\left(\mathbf{x} - \boldsymbol{\mu}_x\right)^\top\right], \quad \boldsymbol{\Sigma}_{yy} = \mathbb{E}\left[\left(\mathbf{y} - \boldsymbol{\mu}_y\right)\left(\mathbf{y} - \boldsymbol{\mu}_y\right)^\top\right]$$

$$\boldsymbol{\Sigma}_{xy} = \mathbb{E}\left[\left(\mathbf{x} - \boldsymbol{\mu}_x\right)\left(\mathbf{y} - \boldsymbol{\mu}_y\right)^\top\right], \quad \boldsymbol{\Sigma}_{yx} = \mathbb{E}\left[\left(\mathbf{y} - \boldsymbol{\mu}_y\right)\left(\mathbf{x} - \boldsymbol{\mu}_x\right)^\top\right]$$

条件付きガウス分布と周辺ガウス分布

- $\mathbf{x}, \mathbf{y}$ の同時分布が，ガウス分布で表されるとする：

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix}\mathbf{x} \\ \mathbf{y}\end{pmatrix} \middle| \begin{pmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}\right)$$

- 共分散は対称行列なので，

$$\begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}^\top = \begin{pmatrix}\boldsymbol{\Sigma}_{xx}^\top & \boldsymbol{\Sigma}_{yx}^\top \\ \boldsymbol{\Sigma}_{xy}^\top & \boldsymbol{\Sigma}_{yy}^\top\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}$$

- 以上より，$\boldsymbol{\Sigma}_{xx}^\top = \boldsymbol{\Sigma}_{xx}$，$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{yy}^\top$ だから，$\boldsymbol{\Sigma}_{xx}$ と $\boldsymbol{\Sigma}_{yy}$ は対称行列.
- また，右上と左下に関しては，$\boldsymbol{\Sigma}_{xy}^\top = \boldsymbol{\Sigma}_{yx}$ が成り立つ.

条件付きガウス分布と周辺ガウス分布

- ガウス分布では, $\exp(\cdot)$ の中に, 共分散の逆行列 $\boldsymbol{\Sigma}^{-1}$ が現れる.
- $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}$ を精度行列 (Precision Matrix) とよぶ.
- 前回は $\boldsymbol{\Lambda}$ を, $\boldsymbol{\Sigma}$ の固有値を並べた対角行列としたので, 混同に注意.
- ブロック行列の逆行列の関係から (第 1 回で扱った),

$$\begin{aligned}
\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} &= \begin{pmatrix} \mathbf{X}^{-1} & -\mathbf{X}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{X}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Y}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Y}^{-1} \\ -\mathbf{Y}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Y}^{-1} \end{pmatrix}
\end{aligned}$$

  $\mathbf{X} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$, $\mathbf{Y} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ は, $\mathbf{A}, \mathbf{D}$ のシューア補行列.
- 上の関係を使って, $\boldsymbol{\Lambda}$ の各成分を調べてみよう.

条件付きガウス分布と周辺ガウス分布

- 精度行列の各成分を, 次のように表す:

$$\begin{pmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}^{-1}$$

- $\boldsymbol{\Sigma}_{xx}$ のシューア補行列 $\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}$ を使えば,

$$\begin{aligned}
\boldsymbol{\Lambda}_{xx} &= \left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)^{-1} \\
\boldsymbol{\Lambda}_{xy} &= -\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \\
\boldsymbol{\Lambda}_{yx} &= -\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)^{-1} \\
\boldsymbol{\Lambda}_{yy} &= \boldsymbol{\Sigma}_{yy}^{-1} + \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}
\end{aligned}$$

条件付きガウス分布と周辺ガウス分布

- 精度行列の各成分を, 次のように表す:

$$\begin{pmatrix} \mathbf{\Lambda}_{xx} & \mathbf{\Lambda}_{xy} \\ \mathbf{\Lambda}_{yx} & \mathbf{\Lambda}_{yy} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{pmatrix}^{-1}$$

- 共分散 $\mathbf{\Sigma}$ は対称行列なので, $\mathbf{\Lambda} = \mathbf{\Sigma}^{-1}$ も対称行列になる:

$$\mathbf{\Lambda}^\top = \left(\mathbf{\Sigma}^{-1}\right)^\top = \left(\mathbf{\Sigma}^\top\right)^{-1} = \mathbf{\Sigma}^{-1} = \mathbf{\Lambda}$$

- $\mathbf{\Lambda}$ は対称行列だから,

$$\begin{pmatrix} \mathbf{\Lambda}_{xx} & \mathbf{\Lambda}_{xy} \\ \mathbf{\Lambda}_{yx} & \mathbf{\Lambda}_{yy} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{\Lambda}_{xx}^\top & \mathbf{\Lambda}_{yx}^\top \\ \mathbf{\Lambda}_{xy}^\top & \mathbf{\Lambda}_{yy}^\top \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_{xx} & \mathbf{\Lambda}_{xy} \\ \mathbf{\Lambda}_{yx} & \mathbf{\Lambda}_{yy} \end{pmatrix}$$

- 以上より, $\mathbf{\Lambda}_{xx}^\top = \mathbf{\Lambda}_{xx}$, $\mathbf{\Lambda}_{yy}^\top = \mathbf{\Lambda}_{yy}$ だから, $\mathbf{\Lambda}_{xx}$ と $\mathbf{\Lambda}_{yy}$ は対称行列.
- また, $\mathbf{\Lambda}_{xy}^\top = \mathbf{\Lambda}_{yx}$.

条件付きガウス分布と周辺ガウス分布

- $\mathbf{x}, \mathbf{y}$ の同時分布が，ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \middle| \begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}\right)$$

- $\exp(\cdot)$ の中身を調べてみよう ($-\frac{1}{2}$ を除く):

$$\begin{aligned}
&\begin{pmatrix}\mathbf{x}-\boldsymbol{\mu}_x\\ \mathbf{y}-\boldsymbol{\mu}_y\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}^{-1} \begin{pmatrix}\mathbf{x}-\boldsymbol{\mu}_x\\ \mathbf{y}-\boldsymbol{\mu}_y\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{x}-\boldsymbol{\mu}_x\\ \mathbf{y}-\boldsymbol{\mu}_y\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy}\\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy}\end{pmatrix} \begin{pmatrix}\mathbf{x}-\boldsymbol{\mu}_x\\ \mathbf{y}-\boldsymbol{\mu}_y\end{pmatrix}\\
&= (\mathbf{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Lambda}_{xx} (\mathbf{x}-\boldsymbol{\mu}_x) + (\mathbf{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Lambda}_{xy} (\mathbf{y}-\boldsymbol{\mu}_y)\\
&\quad + (\mathbf{y}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Lambda}_{yx} (\mathbf{x}-\boldsymbol{\mu}_x) + (\mathbf{y}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Lambda}_{yy} (\mathbf{y}-\boldsymbol{\mu}_y)
\end{aligned}$$

条件付きガウス分布と周辺ガウス分布

- $\mathbf{\Sigma}_{xx}, \mathbf{\Sigma}_{yy}, \mathbf{\Lambda}_{xx}, \mathbf{\Lambda}_{yy}$ は対称．$\mathbf{\Sigma}_{xy}^\top = \mathbf{\Sigma}_{yx}$，$\mathbf{\Lambda}_{xy}^\top = \mathbf{\Lambda}_{yx}$ であるから，

$$
\begin{aligned}
\mathbf{\Lambda}_{xx} &= \left(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\right)^{-1} \\
\mathbf{\Lambda}_{xy} &= -\left(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\right)^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1} = -\mathbf{\Lambda}_{xx}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1} \\
\mathbf{\Lambda}_{yx} &= -\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\left(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\right)^{-1} = -\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{xy}^\top\mathbf{\Lambda}_{xx} \\
\mathbf{\Lambda}_{yy} &= \mathbf{\Sigma}_{yy}^{-1} + \mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\left(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\right)^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1} \\
&= \mathbf{\Sigma}_{yy}^{-1} + \mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{xy}^\top\mathbf{\Lambda}_{xx}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}
\end{aligned}
$$

- これらを使うと，$\exp(\cdot)$ の中身は，

$$
\begin{aligned}
&(\mathbf{x} - \boldsymbol{\mu}_x)^\top \mathbf{\Lambda}_{xx}(\mathbf{x} - \boldsymbol{\mu}_x) - (\mathbf{x} - \boldsymbol{\mu}_x)^\top \mathbf{\Lambda}_{xx}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \\
&\quad - (\mathbf{y} - \boldsymbol{\mu}_y)^\top \mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{xy}^\top\mathbf{\Lambda}_{xx}(\mathbf{x} - \boldsymbol{\mu}_x) \\
&\quad + (\mathbf{y} - \boldsymbol{\mu}_y)^\top \left(\mathbf{\Sigma}_{yy}^{-1} + \mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{xy}^\top\mathbf{\Lambda}_{xx}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\right)(\mathbf{y} - \boldsymbol{\mu}_y)
\end{aligned}
$$

条件付きガウス分布と周辺ガウス分布

- 続いて，項を頑張って整理すると，2 つの二次形式の和となる:

$$
\begin{aligned}
&(\mathbf{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Lambda}_{xx}(\mathbf{x}-\boldsymbol{\mu}_x) - (\mathbf{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Lambda}_{xx}\left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right) \\
&\quad - \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right)^\top \boldsymbol{\Lambda}_{xx}(\mathbf{x}-\boldsymbol{\mu}_x) \\
&\quad + \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right)^\top \boldsymbol{\Lambda}_{xx}\left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right) \\
&\quad + (\mathbf{y}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y) \\
&= \left(\mathbf{x}-\boldsymbol{\mu}_x - \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right)\right)^\top \\
&\quad \boldsymbol{\Lambda}_{xx}\left(\mathbf{x}-\boldsymbol{\mu}_x - \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right)\right) + (\mathbf{y}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)
\end{aligned}
$$

- 上の変形で，$\boldsymbol{\Sigma}_{yy}$ が対称だから，$\boldsymbol{\Sigma}_{yy}^{-1}$ も対称であることを使った．
- $\exp(\cdot)$ の中身が 2 つの二次形式の和だから，同時ガウス分布 $p(\mathbf{x}, \mathbf{y})$ は，2 つのガウス分布に分解できることを示唆する．

条件付きガウス分布と周辺ガウス分布

- $\exp(\cdot)$ の中身は, 2 つの二次形式の和として表現できる:

$$\left(\mathbf{x} - \boldsymbol{\mu}_x - \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right)\right)^\top \boldsymbol{\Lambda}_{xx}\left(\mathbf{x} - \boldsymbol{\mu}_x - \left(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right)\right) + (\mathbf{y} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$$

- 第 1 項: $p(\mathbf{x} \mid \mathbf{y}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Lambda}_{xx}^{-1})$ に対応
- 第 2 項: $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$ に対応
- 同時ガウス分布 $p(\mathbf{x}, \mathbf{y})$ は次のように分解できる:

$$\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x} \mid \mathbf{y})p(\mathbf{y}) \\
p(\mathbf{x} \mid \mathbf{y}) &= \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) \\
p(\mathbf{y}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})
\end{aligned}$$

- 定数係数の部分について確認しておこう.

条件付きガウス分布と周辺ガウス分布

- 同時ガウス分布 $p(\mathbf{x}, \mathbf{y})$ は次のように分解できる:

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \\
p(\mathbf{x} \mid \mathbf{y}) &= \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_y\right), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}\right) \\
p(\mathbf{y}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})
\end{aligned}
$$

- $\mathbf{x}$ を M 次元, $\mathbf{y}$ を N 次元とする. ガウス分布 $p(\mathbf{x}, \mathbf{y})$ の定数係数は,

$$
\begin{aligned}
\frac{1}{(2\pi)^{\frac{M+N}{2}} \sqrt{\det \boldsymbol{\Sigma}}} &= \frac{1}{(2\pi)^{\frac{M+N}{2}} \sqrt{\det \boldsymbol{\Sigma}_{yy} \det\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}\right)}} \\
&= \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \boldsymbol{\Sigma}_{yy}}} \times \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}\right)}}
\end{aligned}
$$

- ガウス分布 $p(\mathbf{y})$, $p(\mathbf{x} \mid \mathbf{y})$ の定数係数の積となっている.

条件付きガウス分布と周辺ガウス分布

- ただし，ブロック行列の行列式に関する，次の式を使った:

$$\det\left(\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}\right) = \det(\mathbf{A})\det\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right) \qquad \mathbf{A}\text{ が正則}$$

$$\det\left(\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}\right) = \det(\mathbf{D})\det\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right) \qquad \mathbf{D}\text{ が正則}$$

- 上記より，

$$\begin{aligned}\det\mathbf{\Sigma} &= \det\left(\begin{pmatrix}\mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy}\\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy}\end{pmatrix}\right)\\ &= \det\mathbf{\Sigma}_{yy}\times\det\left(\mathbf{\Sigma}_{xx}-\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\right)\\ &= \det\mathbf{\Sigma}_{xx}\times\det\left(\mathbf{\Sigma}_{yy}-\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\right)\end{aligned}$$

条件付きガウス分布と周辺ガウス分布

- 同時ガウス分布 $p(\mathbf{x}, \mathbf{y})$ について, 次が分かった:
- $p(\mathbf{x}, \mathbf{y})$ の指数部の中身は, $p(\mathbf{x} \mid \mathbf{y})$, $p(\mathbf{y})$ の指数部の和.
- $p(\mathbf{x}, \mathbf{y})$ の定数係数は, $p(\mathbf{x} \mid \mathbf{y})$, $p(\mathbf{y})$ の定数係数の積.
- よって $p(\mathbf{x}, \mathbf{y})$ は, 2 つのガウス分布 $p(\mathbf{x} \mid \mathbf{y})$, $p(\mathbf{y})$ に分解できる.

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \\
p(\mathbf{x} \mid \mathbf{y}) &= \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_y\right), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}\right) \\
p(\mathbf{y}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})
\end{aligned}
$$

- 2 つのガウス分布 $p(\mathbf{y} \mid \mathbf{x})$, $p(\mathbf{x})$ にも分解できる (練習問題):

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \\
p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}\left(\mathbf{y} \mid \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}\right) \\
p(\mathbf{x}) &= \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})
\end{aligned}
$$

条件付きガウス分布と周辺ガウス分布

- 上記の議論より，周辺ガウス分布についても簡単に得られる:

$$
\begin{aligned}
\int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{y} &= \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})\, \mathrm{d}\mathbf{y} = p(\mathbf{x}) \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x})\, \mathrm{d}\mathbf{y} \\
&= p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\
\int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{x} &= \int_{-\infty}^{\infty} p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y})\, \mathrm{d}\mathbf{x} = p(\mathbf{y}) \int_{-\infty}^{\infty} p(\mathbf{x} \mid \mathbf{y})\, \mathrm{d}\mathbf{x} \\
&= p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})
\end{aligned}
$$

- 同時分布 $p(\mathrm{x}, \mathrm{y})$ がガウス分布であれば，条件付き分布と周辺分布もガウス分布となる.

条件付きガウス分布と周辺ガウス分布

条件付きガウス分布と周辺ガウス分布

$\mathbf{x}, \mathbf{y}$ の同時分布が, ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix}\mathbf{x} \\ \mathbf{y}\end{pmatrix} \,\middle|\, \begin{pmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}\right)$$

このとき, 条件付き分布 $p(\mathbf{x} \mid \mathbf{y})$ と周辺分布 $p(\mathbf{y})$ もガウス分布となる:

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \\ p(\mathbf{x} \mid \mathbf{y}) &= \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_y\right), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right) \\ p(\mathbf{y}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) \end{aligned}$$

条件付きガウス分布と周辺ガウス分布

条件付きガウス分布と周辺ガウス分布

$\mathbf{x}, \mathbf{y}$ の同時分布が, ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}\right)$$

このとき, 条件付き分布 $p(\mathbf{y} \mid \mathbf{x})$ と周辺分布 $p(\mathbf{x})$ もガウス分布となる:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})$$
$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy})$$
$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

ガウス分布の線形変換

- $\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとする.
- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ の変数変換を施したとき, $\mathbf{y}$ が従う分布を考える.
- 次のように, $\mathbf{y}$ の期待値と共分散が求まる:

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{y}\right] &= \mathbb{E}\left[\mathbf{A}\mathbf{x} + \mathbf{b}\right] = \mathbf{A}\,\mathbb{E}\left[\mathbf{x}\right] + \mathbf{b} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b} \\
\mathrm{Var}\left[\mathbf{y}\right] &= \mathbb{E}\left[\left(\mathbf{y} - \mathbb{E}\left[\mathbf{y}\right]\right)\left(\mathbf{y} - \mathbb{E}\left[\mathbf{y}\right]\right)^\top\right] \\
&= \mathbb{E}\left[\left(\left(\mathbf{A}\mathbf{x} + \mathbf{b}\right) - \left(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}\right)\right)\left(\left(\mathbf{A}\mathbf{x} + \mathbf{b}\right) - \left(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}\right)\right)^\top\right] \\
&= \mathbb{E}\left[\mathbf{A}\left(\mathbf{x} - \boldsymbol{\mu}\right)\left(\mathbf{x} - \boldsymbol{\mu}\right)^\top \mathbf{A}^\top\right] \\
&= \mathbf{A}\,\mathbb{E}\left[\left(\mathbf{x} - \boldsymbol{\mu}\right)\left(\mathbf{x} - \boldsymbol{\mu}\right)^\top\right]\mathbf{A}^\top = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top
\end{aligned}
$$

- よって, $\mathbf{y}$ はガウス分布 $\mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ に従う.

ガウス分布の線形変換

- モーメント母関数 (確率分布と一対一で対応) からも確認できる:
- $\mathbf{x}$ のモーメント母関数は,

$$\mathbb{E}\left[\exp\left(\mathbf{t}^\top \mathbf{X}\right)\right] = \exp\left(\boldsymbol{\mu}^\top \mathbf{t} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}\right)$$

- 続いて, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ のモーメント母関数は,

$$\begin{aligned}\mathbb{E}\left[\exp\left(\mathbf{t}^\top (\mathbf{A}\mathbf{X} + \mathbf{b})\right)\right] &= \mathbb{E}\left[\exp\left(\mathbf{t}^\top \mathbf{A}\mathbf{X}\right)\right]\exp\left(\mathbf{t}^\top \mathbf{b}\right) \\ &= \exp\left(\boldsymbol{\mu}^\top \mathbf{A}^\top \mathbf{t} + \frac{1}{2}\mathbf{t}^\top \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top \mathbf{t}\right)\exp\left(\mathbf{b}^\top \mathbf{t}\right) \\ &= \exp\left((\mathbf{A}\boldsymbol{\mu} + \mathbf{b})^\top \mathbf{t} + \frac{1}{2}\mathbf{t}^\top \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top \mathbf{t}\right)\end{aligned}$$

これは, $\mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ に対応する.

ガウス分布の線形変換

- $\mathbf{A}$ が正則であれば, 次のようにも確認できる:
- $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$ だから, ヤコビアンは

$$\left| \det \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = \left| \det \mathbf{A}^{-1} \right| = \left| (\det \mathbf{A})^{-1} \right| = \left| \det \mathbf{A} \right|^{-1}$$

- 次を満たすから, $\mathbf{y}$ はガウス分布 $\mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ に従う.

$$1 = \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}\mathbf{x} = \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top) \, \mathrm{d}\mathbf{y}$$

- この式を確かめよう:

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}\mathbf{x} \\
&= \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \mathrm{d}\mathbf{x} \\
&= \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \int_{-\infty}^{\infty} \exp\Bigg(-\frac{1}{2} \left(\mathbf{A}^{-1} (\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}\right)^\top \\
&\quad \boldsymbol{\Sigma}^{-1} \left(\mathbf{A}^{-1} (\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}\right) \Bigg) |\det \mathbf{A}|^{-1}\, \mathrm{d}\mathbf{y}
\end{aligned}
$$

ガウス分布の線形変換

- $\exp(\cdot)$ の中身は ($-\frac{1}{2}$ は除く), $\boldsymbol{\Sigma}^{-1}$ の対称性と, $(\mathbf{A}^{-1})^\top = \left(\mathbf{A}^\top\right)^{-1}$ を使えば,

$$
\begin{aligned}
&\left(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})-\boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})-\boldsymbol{\mu}\right) \\
&= \left(\mathbf{A}^{-1}(\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}))\right)^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}^{-1}(\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b})) \\
&= (\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}))^\top \left(\mathbf{A}^{-1}\right)^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}^{-1}(\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b})) \\
&= (\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}))^\top \left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\right)^{-1}(\mathbf{y}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}))
\end{aligned}
$$

- $|\det \mathbf{A}| = \sqrt{(\det \mathbf{A})^2} = \sqrt{\det \mathbf{A} \det \mathbf{A}^\top}$ だから,

$$
\frac{1}{\sqrt{\det \boldsymbol{\Sigma}}\,|\det \mathbf{A}|} = \frac{1}{\sqrt{\det \mathbf{A} \det \boldsymbol{\Sigma} \det \mathbf{A}^\top}} = \frac{1}{\sqrt{\det(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)}}
$$

ガウス分布の線形変換

- 以上の結果より, $\mathbf{y}$ はガウス分布 $\mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ に従う.

$$
\begin{aligned}
1 &= \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \left(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}\right)^\top \right. \\
&\qquad \left. \boldsymbol{\Sigma}^{-1} \left(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}\right) \right) |\det \mathbf{A}|^{-1} \, \mathrm{d}\mathbf{y} \\
&= \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} (\mathbf{y} - (\mathbf{A}\boldsymbol{\mu} + \mathbf{b}))^\top \right. \\
&\qquad \left. \left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\right)^{-1} (\mathbf{y} - (\mathbf{A}\boldsymbol{\mu} + \mathbf{b})) \right) \mathrm{d}\mathbf{y} \\
&= \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top) \, \mathrm{d}\mathbf{y} = \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}\mathbf{x}
\end{aligned}
$$

ガウス分布の線形変換

ガウス分布の線形変換

$\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとする.
$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ の変数変換を施すと, $\mathbf{y}$ はガウス分布に従う:

$$\mathcal{N}(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$$

ガウス分布の線形変換

- $\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとする.
- このとき, $\mathbf{x}$ の各成分 x_i は, ガウス分布 $\mathcal{N}(x_i \mid \mu_i, \sigma_i^2)$ に従う.
- 先程の線形変換の例において, 次のように定める:

$$\mathbf{A} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathbf{b} = \mathbf{0}$$

- $\mathbf{A}$ は, 第 i 要素のみが 1, それ以外が 0 の行ベクトルである.
- $x_i = \mathbf{A}\mathbf{x} + \mathbf{b}$ の平均と分散は, 次のように求まる:

$$\begin{aligned} \mathbf{A}\boldsymbol{\mu} + \mathbf{b} &= \mu_i \\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top &= (\boldsymbol{\Sigma})_{ii} = \mathrm{Cov}\left(x_i, x_i\right) = \mathrm{Var}\left[x_i\right] = \sigma_i^2 \end{aligned}$$

- 先程の, 周辺分布の議論からも確認できる (x_i 以外の全ての変数を, 周辺化によって取り除く).

ガウス分布の線形変換

- 続いて, $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$
$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$$

- 例えば, 状態 $\mathbf{x}$ を, 観測データ $\mathbf{y}$ から推定したいとする.
- $p(\mathbf{x})$ は $\mathbf{x}$ の事前分布である.
- $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ を, 観測データに加わる, ガウスノイズとする.
- $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} + \mathbf{n}$ のように, センサをモデル化したとする.
- このとき, $p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$ になる (注意: $\mathbf{x}$ は定数扱い).
- $p(\mathbf{y} \mid \mathbf{x})$ は, 状態 $\mathbf{x}$ の下での観測データ $\mathbf{y}$ の尤もらしさ (尤度).

ガウス分布の線形変換

- $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$
\begin{aligned}
p(\mathbf{x}) &= \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\
p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})
\end{aligned}
$$

- 状態 $\mathbf{x}$ を, 観測データ $\mathbf{y}$ から推定したい. $p(\mathbf{x})$ は $\mathbf{x}$ の事前分布.
- $p(\mathbf{y} \mid \mathbf{x})$ は, 状態 $\mathbf{x}$ の下での観測データ $\mathbf{y}$ の尤もらしさ (尤度).
- ベイズの定理より, 観測データ $\mathbf{y}$ のもとでの状態 $\mathbf{x}$ の確率分布, $p(\mathbf{x} \mid \mathbf{y})$ が得られる:

$$
p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})}{\displaystyle\int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \, \mathrm{d}\mathbf{x}}
$$

- $p(\mathbf{x} \mid \mathbf{y})$ と, $p(\mathbf{y})$ の一般形を求めてみよう.

ガウス分布の線形変換

- $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$
$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$$

- 最初に, 同時分布 $p(\mathbf{x}, \mathbf{y})$ を求めよう:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})$$

- $\exp(\cdot)$ の中身は ($-\frac{1}{2}$ を除く),

$$(\mathbf{y} - (\mathbf{A}\mathbf{x} + \mathbf{b}))^\top \mathbf{R}^{-1} (\mathbf{y} - (\mathbf{A}\mathbf{x} + \mathbf{b})) + (\mathbf{x} - \boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x)$$

ガウス分布の線形変換

- $\exp(\cdot)$ の中身を変形して, 2 次の項, 1 次の項, 定数項に分けると,

$$
\begin{aligned}
&(\mathbf{y}-(\mathbf{A}\mathbf{x}+\mathbf{b}))^\top \mathbf{R}^{-1}(\mathbf{y}-(\mathbf{A}\mathbf{x}+\mathbf{b}))+(\mathbf{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x) \\
&=\mathbf{y}^\top\mathbf{R}^{-1}\mathbf{y}-\mathbf{y}^\top\mathbf{R}^{-1}\mathbf{A}\mathbf{x}-\mathbf{x}^\top\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{y}+\mathbf{x}^\top\left(\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}+\boldsymbol{\Sigma}_{xx}^{-1}\right)\mathbf{x} \\
&\quad-(\mathbf{y}-\mathbf{A}\mathbf{x})^\top\mathbf{R}^{-1}\mathbf{b}-\mathbf{b}^\top\mathbf{R}^{-1}(\mathbf{y}-\mathbf{A}\mathbf{x})-\boldsymbol{\mu}_x^\top\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{x}-\mathbf{x}^\top\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\
&\quad+\mathbf{b}^\top\mathbf{R}^{-1}\mathbf{b}+\boldsymbol{\mu}_x^\top\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\
&=\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top\begin{pmatrix}\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}+\boldsymbol{\Sigma}_{xx}^{-1} & -\mathbf{A}^\top\mathbf{R}^{-1}\\ -\mathbf{R}^{-1}\mathbf{A} & \mathbf{R}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \\
&\quad-\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top\begin{pmatrix}-\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}+\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\\ \mathbf{R}^{-1}\mathbf{b}\end{pmatrix} \\
&\quad-\begin{pmatrix}-\mathbf{b}^\top\mathbf{R}^{-1}\mathbf{A}+\boldsymbol{\mu}_x^\top\boldsymbol{\Sigma}_{xx}^{-1} & \mathbf{b}^\top\mathbf{R}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}+C
\end{aligned}
$$

- $\mathbf{x}, \mathbf{y}$ によらない項をまとめて, C とおいた.

ガウス分布の線形変換

- 共分散 $\mathbf{R}, \boldsymbol{\Sigma}_{xx}$ は対称行列だから，その逆行列も対称．よって，

$$
\begin{aligned}
\begin{pmatrix} -\mathbf{b}^\top \mathbf{R}^{-1}\mathbf{A} + \boldsymbol{\mu}_x^\top \boldsymbol{\Sigma}_{xx}^{-1} & \mathbf{b}^\top \mathbf{R}^{-1} \end{pmatrix} &= \begin{pmatrix} \left(-\mathbf{b}^\top \mathbf{R}^{-1}\mathbf{A} + \boldsymbol{\mu}_x^\top\right)^\top \\ \left(\mathbf{b}^\top \mathbf{R}^{-1}\right)^\top \end{pmatrix}^\top \\
= \begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \left(\boldsymbol{\Sigma}_{xx}^{-1}\right)^\top \boldsymbol{\mu}_x \\ \left(\mathbf{R}^{-1}\right)^\top \mathbf{b} \end{pmatrix}^\top &= \begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}^\top
\end{aligned}
$$

- 同時分布 $p(\mathbf{x}, \mathbf{y})$ の $\exp(\cdot)$ の中身は，

$$
\begin{aligned}
&\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \begin{pmatrix} \mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A} + \boldsymbol{\Sigma}_{xx}^{-1} & -\mathbf{A}^\top \mathbf{R}^{-1} \\ -\mathbf{R}^{-1}\mathbf{A} & \mathbf{R}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \\
&\quad - \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix} - {\color{red}\begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}^\top} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} + C
\end{aligned}
$$

ガウス分布の線形変換

- ブロック行列の逆行列の関係から,

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{Y}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{Y}^{-1} \\ -\mathbf{Y}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{Y}^{-1} \end{pmatrix}$$

$\mathbf{Y} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ は, $\mathbf{A}_{22}$ のシューア補行列.

- この式を逆方向に使うと,

$$\begin{aligned}
\mathbf{\Sigma} &= \begin{pmatrix} \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A} + \mathbf{\Sigma}_{xx}^{-1} & -\mathbf{A}^\top\mathbf{R}^{-1} \\ -\mathbf{R}^{-1}\mathbf{A} & \mathbf{R}^{-1} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \mathbf{\Sigma}_{xx}^{-1} + \mathbf{\Sigma}_{xx}^{-1}\left(\mathbf{\Sigma}_{xx}\mathbf{A}^\top\right)\mathbf{R}^{-1}\left(\mathbf{A}\mathbf{\Sigma}_{xx}\right)\mathbf{\Sigma}_{xx}^{-1} & -\mathbf{\Sigma}_{xx}^{-1}\left(\mathbf{\Sigma}_{xx}\mathbf{A}^\top\right)\mathbf{R}^{-1} \\ -\mathbf{R}^{-1}\left(\mathbf{A}\mathbf{\Sigma}_{xx}\right)\mathbf{\Sigma}_{xx}^{-1} & \mathbf{R}^{-1} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\mathbf{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix} \\
&\quad \left(\because \mathbf{R} = \mathbf{A}_{22} - \left(\mathbf{A}\mathbf{\Sigma}_{xx}\right)\mathbf{\Sigma}_{xx}^{-1}\left(\mathbf{\Sigma}_{xx}\mathbf{A}^\top\right)\right)
\end{aligned}$$

ガウス分布の線形変換

- 同時分布 $p(\mathbf{x}, \mathbf{y})$ の $\exp(\cdot)$ の中身に戻ると,

$$
\begin{aligned}
&\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top \begin{pmatrix}\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}+\mathbf{\Sigma}_{xx}^{-1} & -\mathbf{A}^\top\mathbf{R}^{-1}\\ -\mathbf{R}^{-1}\mathbf{A} & \mathbf{R}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\\
&\quad -\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top \begin{pmatrix}-\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}+\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\\ \mathbf{R}^{-1}\mathbf{b}\end{pmatrix} - \begin{pmatrix}-\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}+\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\\ \mathbf{R}^{-1}\mathbf{b}\end{pmatrix}^\top \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} + C\\
&= \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top \mathbf{\Sigma}^{-1}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} - \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^\top \mathbf{\Sigma}^{-1}\mathbf{\Sigma}\begin{pmatrix}-\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}+\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\\ \mathbf{R}^{-1}\mathbf{b}\end{pmatrix}\\
&\quad - \begin{pmatrix}-\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}+\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\\ \mathbf{R}^{-1}\mathbf{b}\end{pmatrix}^\top \mathbf{\Sigma}\mathbf{\Sigma}^{-1}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} + C
\end{aligned}
$$

ガウス分布の線形変換

- 共分散 $\mathbf{R}, \mathbf{\Sigma}_{xx}$ は対称行列だから, $\mathbf{\Sigma}$ も対称になる ($\mathbf{\Sigma}^{-1}$ も対称).

$$\mathbf{\Sigma}^\top = \begin{pmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\mathbf{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix}^\top = \begin{pmatrix} \mathbf{\Sigma}_{xx}^\top & (\mathbf{A}\mathbf{\Sigma}_{xx})^\top \\ (\mathbf{\Sigma}_{xx}\mathbf{A}^\top)^\top & (\mathbf{R} + \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^\top)^\top \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{\Sigma}_{xx}^\top & \mathbf{\Sigma}_{xx}^\top\mathbf{A}^\top \\ \mathbf{A}^\top\mathbf{\Sigma}_{xx}^\top & \mathbf{R}^\top + \mathbf{A}\mathbf{\Sigma}_{xx}^\top\mathbf{A}^\top \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}^\top\mathbf{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix} = \mathbf{\Sigma}$$

- $\mathbf{\Sigma}$ は対称だから, $\mathbf{X}^\top\mathbf{\Sigma} = \mathbf{X}^\top\mathbf{\Sigma}^\top = (\mathbf{\Sigma}\mathbf{X})^\top$.
- 同時分布 $p(\mathbf{x}, \mathbf{y})$ の $\exp(\cdot)$ の中身に戻ると,

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \mathbf{\Sigma}^{-1} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} - \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \mathbf{\Sigma}^{-1}\mathbf{\Sigma} \begin{pmatrix} -\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b} + \mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}$$
$$- \left(\mathbf{\Sigma}\begin{pmatrix} -\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b} + \mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}\right)^\top \mathbf{\Sigma}^{-1} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} + C$$

ガウス分布の線形変換

- また,

$$
\boldsymbol{\mu} = \boldsymbol{\Sigma} \begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}
$$

$$
= \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix} \begin{pmatrix} -\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\ \mathbf{R}^{-1}\mathbf{b} \end{pmatrix}
$$

$$
= \begin{pmatrix} -\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} \\ -\mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} + \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \end{pmatrix}
$$

- 同時分布 $p(\mathbf{x}, \mathbf{y})$ の $\exp(\cdot)$ の中身を平方完成させると,

$$
\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} - \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^\top \boldsymbol{\Sigma}^{-1} {\color{red}\boldsymbol{\mu}} - {\color{red}\boldsymbol{\mu}^\top} \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} + C
$$

$$
= \left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} - \boldsymbol{\mu} \right)^\top \boldsymbol{\Sigma}^{-1} \left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} - \boldsymbol{\mu} \right) + C'
$$

ガウス分布の線形変換

- $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$
$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$$

- $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})$ は, 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ のガウス分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix}$$

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A} + \boldsymbol{\Sigma}_{xx}^{-1} & -\mathbf{A}^\top\mathbf{R}^{-1} \\ -\mathbf{R}^{-1}\mathbf{A} & \mathbf{R}^{-1} \end{pmatrix}$$

ガウス分布の線形変換

- ここで，条件付きガウス分布と周辺ガウス分布の関係を用いる．
- 同時分布 $p(\mathbf{x}, \mathbf{y})$ は，2 つのガウス分布 $p(\mathbf{x} \mid \mathbf{y})$，$p(\mathbf{y})$ に分解できる:

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix}\right) \\
p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \\
p(\mathbf{x} \mid \mathbf{y}) &= \mathcal{N}\left(\mathbf{x} \middle| \boldsymbol{\Sigma}_{x|y}\left(\mathbf{A}^\top \mathbf{R}^{-1}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{x|y}\right) \\
p(\mathbf{y}) &= \mathcal{N}\left(\mathbf{y} \middle| \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}, \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top\right)
\end{aligned}
$$

- ただし，

$$
\boldsymbol{\Sigma}_{x|y} = \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}\right)^{-1}
$$

ガウス分布の線形変換

- 条件付きガウス分布 $p(\mathbf{x} \mid \mathbf{y})$ の共分散と平均は,

$$\begin{aligned}
\boldsymbol{\Sigma}_{x|y} &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \left(\mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top\right)^{-1} \mathbf{A}\boldsymbol{\Sigma}_{xx} = \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}\right)^{-1} \\
\boldsymbol{\mu}_{x|y} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \left(\mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top\right)^{-1} \left(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_x - \mathbf{b}\right) \\
&= \boldsymbol{\mu}_x + \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}\right)^{-1} \mathbf{A}^\top\mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_x - \mathbf{b}\right) \\
&= \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}\right)^{-1} \left(\mathbf{A}^\top\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{b}\right) + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\right) \\
&= \boldsymbol{\Sigma}_{x|y} \left(\mathbf{A}^\top\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{b}\right) + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\right)
\end{aligned}$$

- Woodbury の公式を用いた:

$$\left(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B}\right)^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1}$$

$$\mathbf{A}\mathbf{B}\left(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B}\right)^{-1} = \left(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1}$$

ガウス分布の線形変換

ガウス分布の線形変換

$p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}), \quad p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$$

このとき, 同時分布 $p(\mathbf{x}, \mathbf{y})$, 条件付き分布 $p(\mathbf{x} \mid \mathbf{y})$, 周辺分布 $p(\mathbf{y})$ は:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Sigma}_{xx} & \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top \end{pmatrix}\right)$$

$$p(\mathbf{x} \mid \mathbf{y}) = \mathcal{N}\left(\mathbf{x} \middle| \boldsymbol{\Sigma}_{x|y}\left(\mathbf{A}^\top\mathbf{R}^{-1}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{x|y}\right)$$

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y} \middle| \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}, \mathbf{R} + \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top\right)$$

ただし, $\boldsymbol{\Sigma}_{x|y} = \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}\right)^{-1}$.

ガウス分布の非線形変換 (線形化)

- 続いて, $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$
$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R})$$

- $\mathbf{g}(\mathbf{x})$ は非線形関数とする.
- 例えば, 状態 $\mathbf{x}$ を, 観測データ $\mathbf{y}$ から推定したいとする.
- $p(\mathbf{x})$ は $\mathbf{x}$ の事前分布である.
- $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ を, 観測データに加わる, ガウスノイズとする.
- $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{n}$ のように, センサをモデル化したとする.
- このとき, $p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R})$ になる (注意: $\mathbf{x}$ は定数扱い).
- $p(\mathbf{y} \mid \mathbf{x})$ は, 状態 $\mathbf{x}$ の下での観測データ $\mathbf{y}$ の尤もらしさ (尤度).

ガウス分布の非線形変換 (線形化)

- $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\ p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R}) \end{aligned}$$

- 状態 $\mathbf{x}$ を, 観測データ $\mathbf{y}$ から推定したい. $p(\mathbf{x})$ は $\mathbf{x}$ の事前分布.
- $p(\mathbf{y} \mid \mathbf{x})$ は, 状態 $\mathbf{x}$ の下での観測データ $\mathbf{y}$ の尤もらしさ (尤度).
- ベイズの定理より, 観測データ $\mathbf{y}$ のもとでの状態 $\mathbf{x}$ の確率分布, $p(\mathbf{x} \mid \mathbf{y})$ が得られる:

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})}{\displaystyle\int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \, \mathrm{d}\mathbf{x}}$$

- $p(\mathbf{x} \mid \mathbf{y})$ と, $p(\mathbf{y})$ がどうなるか調べてみよう.

ガウス分布の非線形変換 (線形化)

- $p(\mathbf{x})$, $p(\mathbf{y} \mid \mathbf{x})$ が次のように定まっているとする:

$$
\begin{aligned}
p(\mathbf{x}) &= \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\
p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R})
\end{aligned}
$$

- 一般の $\mathbf{g}(\mathbf{x})$ に対して, $p(\mathbf{y} \mid \mathbf{x})$, $p(\mathbf{y})$ が解析的に求まるとは限らない.
- そこで, $\mathbf{g}(\mathbf{x})$ を $\boldsymbol{\mu}_x$ のまわりで線形化する:

$$
\begin{aligned}
\mathbf{g}(\mathbf{x}) &\approx \boldsymbol{\mu}_y + \mathbf{G}\left(\mathbf{x} - \boldsymbol{\mu}_x\right) \\
\boldsymbol{\mu}_y &= \mathbf{g}(\boldsymbol{\mu}_x) \\
\mathbf{G} &= \left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\boldsymbol{\mu}_x}
\end{aligned}
$$

- $p(\mathbf{y} \mid \mathbf{x}) \approx \mathcal{N}(\boldsymbol{\mu}_y + \mathbf{G}\left(\mathbf{x} - \boldsymbol{\mu}_x\right), \mathbf{R})$ となって, 先ほどの例と同様になる.

ガウス分布の非線形変換 (線形化)

- 線形化によって, $p(\mathbf{y} \mid \mathbf{x})$ が次のようになる:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}), \quad p(\mathbf{y} \mid \mathbf{x}) \approx \mathcal{N}(\boldsymbol{\mu}_y + \mathbf{G}\left(\mathbf{x} - \boldsymbol{\mu}_x\right), \mathbf{R})$$

- このとき, 同時分布 $p(\mathbf{x}, \mathbf{y})$, 条件付き分布 $p(\mathbf{x} \mid \mathbf{y})$, 周辺分布 $p(\mathbf{y})$ は:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{G}^\top \\ \mathbf{G}\boldsymbol{\Sigma}_{xx} & \mathbf{R} + \mathbf{G}\boldsymbol{\Sigma}_{xx}\mathbf{G}^\top \end{pmatrix}\right)$$

$$p(\mathbf{x} \mid \mathbf{y}) = \mathcal{N}\left(\mathbf{x} \middle| \boldsymbol{\Sigma}_{x|y}\left(\mathbf{G}^\top\mathbf{R}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_y - \mathbf{G}\boldsymbol{\mu}_x\right) + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{x|y}\right)$$

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y} \middle| \boldsymbol{\mu}_y, \mathbf{R} + \mathbf{G}\boldsymbol{\Sigma}_{xx}\mathbf{G}^\top\right)$$

- ただし, $\boldsymbol{\Sigma}_{x|y} = \left(\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{G}^\top\mathbf{R}^{-1}\mathbf{G}\right)^{-1}$, $\boldsymbol{\mu}_y = \mathbf{g}(\boldsymbol{\mu}_x)$.

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

ガウス分布の和

- $\mathbf{x}$, $\mathbf{y}$ は互いに独立で，ガウス分布 $\mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$, $\mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$ に従うとする．
- 和 $\mathbf{z} = \mathbf{x} + \mathbf{y}$ は，ガウス分布 $\mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{yy})$ に従う．
- この性質を再生性という．
- 和 $\mathbf{z} = \mathbf{x} + \mathbf{y}$ の平均と共分散を求めよう．

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{z}\right] &= \mathbb{E}\left[\mathbf{x} + \mathbf{y}\right] = \mathbb{E}\left[\mathbf{x}\right] + \mathbb{E}\left[\mathbf{y}\right] = \boldsymbol{\mu}_x + \boldsymbol{\mu}_y \\
\mathrm{Var}\left[\mathbf{z}\right] &= \mathrm{Var}\left[\mathbf{x} + \mathbf{y}\right] \\
&= \mathrm{Var}\left[\mathbf{x}\right] + \mathrm{Var}\left[\mathbf{y}\right] + \mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) + \mathrm{Cov}\left(\mathbf{y}, \mathbf{x}\right) \\
&= \mathrm{Var}\left[\mathbf{x}\right] + \mathrm{Var}\left[\mathbf{y}\right] = \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{yy}
\end{aligned}
$$

- $\mathbf{x}$, $\mathbf{y}$ は独立 (無相関) なので，$\mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) = \mathbf{0}$, $\mathrm{Cov}\left(\mathbf{y}, \mathbf{x}\right) = \mathbf{0}$.

ガウス分布の和

- $\mathbf{x}$, $\mathbf{y}$ は互いに独立で, ガウス分布 $\mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$, $\mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$ に従うとする.
- 和 $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}$ は, ガウス分布 $\mathcal{N}(\mathbf{A}\boldsymbol{\mu}_x + \mathbf{B}\boldsymbol{\mu}_y, \mathbf{A}\boldsymbol{\Sigma}_{xx}\mathbf{A}^\top + \mathbf{B}\boldsymbol{\Sigma}_{yy}\mathbf{B}^\top)$ に従う.
- 練習問題.

ガウス分布の和

- K 個の独立な確率変数 $\mathbf{x}_k$ があり，ガウス分布 $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ に従うとする．
- このとき，重み付き和 $\mathbf{x} = \sum_k \mathbf{W}_k \mathbf{x}_k$ は，以下のガウス分布に従う：

$$\mathbf{x} \sim \mathcal{N}\left(\sum_k \mathbf{W}_k \boldsymbol{\mu}_k, \sum_k \mathbf{W}_k \boldsymbol{\Sigma}_k \mathbf{W}_k^\top\right)$$

- 重み付き和の平均を求めよう：

$$\mathbb{E}\left[\mathbf{x}\right] = \mathbb{E}\left[\sum_k \mathbf{W}_k \mathbf{x}_k\right] = \sum_k \mathbf{W}_k \, \mathbb{E}\left[\mathbf{x}_k\right] = \sum_k \mathbf{W}_k \boldsymbol{\mu}_k$$

ガウス分布の和

- 重み付き和の共分散は,

$$
\begin{aligned}
\mathrm{Var}\left[\mathbf{x}\right] &= \mathbb{E}\left[\left(\mathbf{x} - \mathbb{E}\left[\mathbf{x}\right]\right)\left(\mathbf{x} - \mathbb{E}\left[\mathbf{x}\right]\right)^{\top}\right] \\
&= \mathbb{E}\left[\left(\sum_{k} \mathbf{W}_k \left(\mathbf{x}_k - \boldsymbol{\mu}_k\right)\right)\left(\sum_{l} \mathbf{W}_l \left(\mathbf{x}_l - \boldsymbol{\mu}_l\right)\right)^{\top}\right] \\
&= \sum_{k}\sum_{l} \mathbf{W}_k \mathbb{E}\left[\left(\mathbf{x}_k - \boldsymbol{\mu}_k\right)\left(\mathbf{x}_l - \boldsymbol{\mu}_l\right)\right] \mathbf{W}_l^{\top} \\
&= \sum_{k}\sum_{l} \mathbf{W}_k \mathrm{Cov}\left(\mathbf{x}_k, \mathbf{x}_l\right) \mathbf{W}_l^{\top} = \sum_{k} \mathbf{W}_k \mathrm{Cov}\left(\mathbf{x}_k, \mathbf{x}_k\right) \mathbf{W}_k^{\top} \\
&= \sum_{k} \mathbf{W}_k \mathrm{Var}\left[\mathbf{x}_k\right] \mathbf{W}_k^{\top} = \sum_{k} \mathbf{W}_k \boldsymbol{\Sigma}_k \mathbf{W}_k^{\top}
\end{aligned}
$$

- $\mathbf{x}_k$ は互いに独立なので, $i \neq j$ のとき $\mathrm{Cov}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{0}$.

ガウス分布の和

- K 個の独立な確率変数 $\mathbf{x}_k$ があり，ガウス分布 $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ に従うとする．
- このとき，重み付き和 $\mathbf{x} = \sum_k w_k \mathbf{x}_k$ は，以下のガウス分布に従う：

$$\mathbf{x} \sim \mathcal{N}\left(\sum_k w_k \boldsymbol{\mu}_k, \sum_k w_k^2 \boldsymbol{\Sigma}_k\right)$$

- 練習問題．

ガウス分布の積

- $\mathbf{x}$ に関する 2 つのガウス分布 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ を考える.
- 2 つのガウス分布の積を, 積分が 1 になるように正規化すれば, 新たなガウス分布となる.
- 積の指数部分 $\exp(\cdot)$ に着目すると ($-\frac{1}{2}$ は除く),

$$
\begin{aligned}
&(\mathbf{x}-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{x}-\boldsymbol{\mu}_1) + (\mathbf{x}-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1} (\mathbf{x}-\boldsymbol{\mu}_2) \\
&= \mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right) \mathbf{x} - \mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right) \\
&\quad - \left(\boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1}\right) \mathbf{x} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2 \\
&= \mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right) \mathbf{x} - \mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right) \\
&\quad - \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right)^\top \mathbf{x} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2
\end{aligned}
$$

ガウス分布の積

- $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}$ (対称), $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2$ とおいて,

$$
\begin{aligned}
&\mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)\mathbf{x} - \mathbf{x}^\top \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right) \\
&\quad - \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right)^\top \mathbf{x} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
&= \mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
&= (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \boldsymbol{\mu}^\top\boldsymbol{\mu} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
&= (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \mathrm{Const.}
\end{aligned}
$$

- $\mathbf{x}$ によらない定数項をまとめて, $\mathrm{Const.}$ とした.
- よって, 2 つのガウス分布 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ の積は,

$$
\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \propto \exp\Big((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\Big) \propto \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})
$$

ガウス分布の積

正規化されたガウス分布の積

$\mathbf{x}$ に関する 2 つのガウス分布 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ を考える. 2 つのガウス分布の積を正規化すれば, 新たなガウス分布となる.

$$\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \propto \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

定数項を取り除いて, 指数部分だけを抜き出すと,

$$\begin{aligned}
&\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\right)\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)
\end{aligned}$$

ただし, 新たな平均 $\boldsymbol{\mu}$ と共分散 $\boldsymbol{\Sigma}$ は,

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}, \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2$$

ガウス分布の積

- $\mathbf{x}$ に関する K 個のガウス分布 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, \mathcal{N}(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)$ を考える.
- K 個のガウス分布の積を, 積分が 1 になるように正規化すれば, 新たなガウス分布となる.
- 積の指数部分 $\exp(\cdot)$ に着目すると ($-\dfrac{1}{2}$ は除く),

$$
\begin{aligned}
&\sum_k (\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \\
&= \mathbf{x}^\top \left(\sum_k \boldsymbol{\Sigma}_k^{-1} \right) \mathbf{x} - \mathbf{x}^\top \left(\sum_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \right) \\
&\quad - \left(\sum_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k^{-1} \right) \mathbf{x} + \sum_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k
\end{aligned}
$$

ガウス分布の積

- $\boldsymbol{\Sigma}^{-1} = \sum_k \boldsymbol{\Sigma}_k^{-1}$, $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \sum_k \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k$ とおけば,

$$\begin{aligned}
&\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} + \sum_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k \\
&= (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \boldsymbol{\mu}^\top \boldsymbol{\mu} + \sum_k \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k \\
&= (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \mathrm{Const.}
\end{aligned}$$

- $\mathbf{x}$ によらない定数項をまとめて, $\mathrm{Const.}$ とした.
- よって, K 個のガウス分布の積は,

$$\prod_i \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \propto \exp\Big((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\Big) \propto \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

ガウス分布の積

正規化されたガウス分布の積

$\mathbf{x}$ に関する K 個のガウス分布 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, \mathcal{N}(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)$ を考える．K 個のガウス分布の積を正規化すれば，新たなガウス分布となる．

$$\prod_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \propto \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

定数項を取り除いて，指数部分だけを抜き出すと，

$$\prod_k \exp\left(-\frac{1}{2}\left(\mathbf{x}_k - \boldsymbol{\mu}_k\right)^\top \boldsymbol{\Sigma}_k^{-1}\left(\mathbf{x}_k - \boldsymbol{\mu}_k\right)\right) \propto \exp\left(-\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right)\right)$$

ただし，新たな平均 $\boldsymbol{\mu}$ と共分散 $\boldsymbol{\Sigma}$ は，

$$\boldsymbol{\Sigma}^{-1} = \sum_k \boldsymbol{\Sigma}_k^{-1}, \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \sum_k \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k$$

ガウス分布の積

- 同様に, 次が成り立つ (練習問題):

$$
\begin{aligned}
&\prod_k \exp\left(-\frac{1}{2}\left(\mathbf{G}_k\mathbf{x}-\boldsymbol{\mu}_k\right)^\top \boldsymbol{\Sigma}_k^{-1}\left(\mathbf{G}_k\mathbf{x}-\boldsymbol{\mu}_k\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right)
\end{aligned}
$$

- ただし,

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \sum_k \mathbf{G}_k^\top \boldsymbol{\Sigma}_k^{-1}\mathbf{G}_k \\
\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} &= \sum_k \mathbf{G}_k^\top \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k
\end{aligned}
$$

目次

1 Isserlis の定理

2 条件付きガウス分布と周辺ガウス分布

3 ガウス分布の線形変換

4 ガウス分布の和と積

5 ガウス分布の最頻値 (モード)

6 ガウス分布の無相関性と独立性

7 ガウス分布の無相関化と標準化

ガウス分布の最頻値 (モード)

- $\mathbf{x}$ についてのガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ を考える.
- ガウス分布の最頻値が, 平均 $\boldsymbol{\mu}$ と一致することを確かめよう.
- 最頻値: 確率密度が最大になるような $\mathbf{x}$
- 極値を調べるために, $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ を $\mathbf{x}$ で微分して $\mathbf{0}$ とおく:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \mathbf{0} \\ \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \frac{\partial}{\partial \mathbf{x}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right) &= \mathbf{0} \\ -\frac{1}{2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right) \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) &= \mathbf{0} \\ -\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right) (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} &= \mathbf{0} \end{aligned}$$

- よって, 極値は $\mathbf{x} = \boldsymbol{\mu}$.

ガウス分布の最頻値 (モード)

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ のヘッセ行列を調べよう.

$$
\begin{aligned}
&\frac{\partial^2}{\partial \mathbf{x}^\top \partial \mathbf{x}} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \frac{\partial^2}{\partial \mathbf{x}^\top \partial \mathbf{x}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \\
&= -\frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \frac{\partial}{\partial \mathbf{x}^\top} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \\
&= -\frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \left\{ \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \boldsymbol{\Sigma}^{-1} \right. \\
&\quad \left. - \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \right\}
\end{aligned}
$$

ガウス分布の最頻値 (モード)

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ のヘッセ行列は,

$$
\begin{aligned}
&-\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{\det \boldsymbol{\Sigma}}}\left\{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\boldsymbol{\Sigma}^{-1}\right.\\
&\quad\left.-\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^2 \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\right\}
\end{aligned}
$$

- ヘッセ行列を $\mathbf{x} = \boldsymbol{\mu}$ で評価し, その負定値性を確認しよう.
- $\mathbf{x} = \boldsymbol{\mu}$ を代入すると, $-\dfrac{1}{(2\pi)^{\frac{D}{2}}\sqrt{\det \boldsymbol{\Sigma}}}\boldsymbol{\Sigma}^{-1}$.
- $\boldsymbol{\Sigma}^{-1}$ は正定値だから, 任意の $\mathbf{y} \neq \mathbf{0}$ について, $\mathbf{y}\boldsymbol{\Sigma}^{-1}\mathbf{y} > 0$.
- 上式のヘッセ行列は負定値 (< 0) となるから, $\mathbf{x} = \boldsymbol{\mu}$ は極大点である.
- よって, $\mathbf{x} = \boldsymbol{\mu}$ はガウス分布の最頻値となって, 平均に一致する.

目次

1 Isserlis の定理

2 条件付きガウス分布と周辺ガウス分布

3 ガウス分布の線形変換

4 ガウス分布の和と積

5 ガウス分布の最頻値 (モード)

6 ガウス分布の無相関性と独立性

7 ガウス分布の無相関化と標準化

ガウス分布の無相関性と独立性

- (独立) $\Longrightarrow$ (無相関) は成り立つが，その逆は一般に成り立たない．
- ガウス分布では，(独立) $\Longleftrightarrow$ (無相関) であることを確認しよう．
- $\mathbf{x}, \mathbf{y}$ の同時分布が，ガウス分布で表されるとする:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}\right)$$

- このとき，条件付き分布と周辺分布もガウス分布となる:

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \\ p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}\left(\mathbf{y} \mid \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}\right) \\ p(\mathbf{x}) &= \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\ p(\mathbf{y}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) \end{aligned}$$

ガウス分布の無相関性と独立性

- $\mathbf{x}, \mathbf{y}$ が無相関であるとき, $\mathbb{E}\left[\mathbf{x}\mathbf{y}^\top\right] = \mathbb{E}\left[\mathbf{x}\right]\mathbb{E}\left[\mathbf{y}\right]^\top$ だから,

$$\begin{aligned}\boldsymbol{\Sigma}_{xy} &= \mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) = \mathbb{E}\left[\mathbf{x}\right]\mathbb{E}\left[\mathbf{y}\right]^\top - \mathbb{E}\left[\mathbf{x}\right]\mathbb{E}\left[\mathbf{y}\right]^\top = \mathbf{0} \\ \boldsymbol{\Sigma}_{yx} &= \mathrm{Cov}\left(\mathbf{y}, \mathbf{x}\right) = \left(\mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right)\right)^\top = \mathbf{0}\end{aligned}$$

- よって, 以下より $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) = p(\mathbf{x})p(\mathbf{y})$ が成り立つから, $\mathbf{x}, \mathbf{y}$ は独立.

$$\begin{aligned}p(\mathbf{y} \mid \mathbf{x}) &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}_x\right), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}) \\ &= \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) = p(\mathbf{y})\end{aligned}$$

- 上記の議論を逆に進めることで, (独立) $\Longrightarrow$ (無相関) も示せる.
- よって, ガウス分布では, 独立性と無相関性が同値となる.

目次

1. Isserlis の定理
2. 条件付きガウス分布と周辺ガウス分布
3. ガウス分布の線形変換
4. ガウス分布の和と積
5. ガウス分布の最頻値 (モード)
6. ガウス分布の無相関性と独立性
7. ガウス分布の無相関化と標準化

ガウス分布の無相関化と標準化

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ の共分散 $\boldsymbol{\Sigma}$ を，直交行列 $\mathbf{U}$ で対角化する ($\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$)．$\boldsymbol{\Lambda}$ は，$\boldsymbol{\Sigma}$ の固有値を斜めに並べた，対角行列．
- $\mathbf{y} = \mathbf{U}^\top (\mathbf{x} - \boldsymbol{\mu})$ の変数変換を施すと，平均と共分散は

$$\mathbb{E}[\mathbf{y}] = \mathbf{U}^\top (\boldsymbol{\mu} - \boldsymbol{\mu}) = \mathbf{0}, \quad \mathrm{Var}[\mathbf{y}] = \mathbf{U}^\top \boldsymbol{\Sigma} \mathbf{U} = \boldsymbol{\Lambda}$$

- よって，$\mathbf{y}$ はガウス分布 $\mathcal{N}(\mathbf{y} \mid \mathbf{0}, \boldsymbol{\Lambda})$ に従う．
- 共分散は対角行列だから，$\mathbf{y}$ の各成分は互いに無相関で，独立．
- ガウス分布なので，無相関性と独立性は同値．
- このような手続きを，無相関化という．
- $\mathbf{y}$ の各成分 y_i は，平均 0，分散 $\sqrt{\lambda_i}$ のガウス分布に従う (λ_i は，$\boldsymbol{\Sigma}$ の対応する固有値)．各成分の分散には，固有値によるばらつきが生じる．

ガウス分布の無相関化と標準化

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ の共分散 $\boldsymbol{\Sigma}$ を, 直交行列 $\mathbf{U}$ で対角化する ($\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$).
- $\boldsymbol{\Lambda}$ は, $\boldsymbol{\Sigma}$ の固有値を斜めに並べた, 対角行列 ($\boldsymbol{\Lambda}^{\frac{1}{2}}, \boldsymbol{\Lambda}^{-\frac{1}{2}}$ も対角行列).
- $\mathbf{y} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top (\mathbf{x} - \boldsymbol{\mu})$ の変数変換を施すと, 平均と共分散は

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{y}\right] &= \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top (\boldsymbol{\mu} - \boldsymbol{\mu}) = \mathbf{0} \\
\mathrm{Var}\left[\mathbf{y}\right] &= \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\boldsymbol{\Sigma}\left(\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\right)^\top \\
&= \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\boldsymbol{\Sigma}\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \mathbf{I}
\end{aligned}
$$

- よって, $\mathbf{y}$ は標準正規分布 $\mathcal{N}(\mathbf{y} \mid \mathbf{0}, \mathbf{I})$ に従う.
- このような手続きを, 白色化という.
- $\mathbf{y}$ の各成分 y_i は, 平均 0, 分散 1 の標準正規分布に従う. 無相関化の場合とは異なり, 固有値による分散のばらつきが解消されている.