

行列論講: 第 8 回 ガウス分布 2

松谷研究室

June 2, 2024

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

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

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

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

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

- $n = 4$ のときは、Isserlis の定理は次のようになる:

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

Isserlis の定理

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

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

- 総和: $\{1, \dots, n\}$ 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。

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

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

- $\mathbb{E}[x^6]$ は

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

- $\mathbb{E}[x^8]$ は

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

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

- 続いて, $\mathbf{x} = (x_1, \dots, x_D)$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \Sigma)$ 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:

$$\text{Cov}(x_i, x_j) = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])] = \mathbb{E}[x_i x_j]$$

$$\text{Var}[x_i] = \text{Cov}(x_i, x_i) = \mathbb{E}[x_i^2]$$

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

$$(\Sigma)_{ij} = \text{Cov}(x_i, x_j) = \mathbb{E}[x_i x_j]$$

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

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

$$\begin{aligned} \mathbb{E} \left[\left(\mathbf{x} \left(\mathbf{x}^\top \mathbf{x} \right) \mathbf{x}^\top \right)_{ij} \right] &= \mathbb{E} \left[x_i \left(\sum_{k=1}^D x_k^2 \right) x_j \right] = \sum_k \mathbb{E} [x_i x_j x_k^2] \\ &= \sum_k \mathbb{E} [x_i x_j] \mathbb{E} [x_k^2] + 2 \mathbb{E} [x_i x_k] \mathbb{E} [x_j x_k] \\ &= \mathbb{E} [x_i x_j] \left(\sum_k \mathbb{E} [x_k^2] \right) + 2 \sum_k \mathbb{E} [x_i x_k] \mathbb{E} [x_k x_j] \\ &= (\boldsymbol{\Sigma})_{ij} \left(\sum_k (\boldsymbol{\Sigma})_{kk} \right) + 2 \sum_k (\boldsymbol{\Sigma})_{ik} (\boldsymbol{\Sigma})_{kj} = (\boldsymbol{\Sigma})_{ij} \text{tr}(\boldsymbol{\Sigma}) + 2 (\boldsymbol{\Sigma}^2)_{ij} \end{aligned}$$

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

- 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

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

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

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

- 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

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

となるから,

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

- 続いて, $\mathbf{x} = (x_1, \dots, x_D)$ 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 $\mathbb{E}[z] = \mathbb{E}[\mathbf{x}^\top \mathbf{x}]$ を求めよう:

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

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

- 続いて、分散 $\text{Var}[z] = \mathbb{E}[(z - \mathbb{E}[z])^2]$ 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。

- 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、

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

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

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.

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

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

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

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 Σ^{-1} が現れる.
- $\Sigma^{-1} = \Lambda$ を**精度行列** (Precision Matrix) とよぶ.
- 前回は Λ を, Σ 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, Λ 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:

$$\begin{pmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}^{-1}$$

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

$$\Lambda_{xx} = (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}$$

$$\Lambda_{xy} = -(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} \Sigma_{xy}\Sigma_{yy}^{-1}$$

$$\Lambda_{yx} = -\Sigma_{yy}^{-1}\Sigma_{yx} (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}$$

$$\Lambda_{yy} = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx} (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} \Sigma_{xy}\Sigma_{yy}^{-1}$$

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:

$$\begin{pmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}^{-1}$$

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

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

- Λ は対称行列だから,

$$\begin{pmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{pmatrix}^{\top} = \begin{pmatrix} \Lambda_{xx}^{\top} & \Lambda_{yx}^{\top} \\ \Lambda_{xy}^{\top} & \Lambda_{yy}^{\top} \end{pmatrix} = \begin{pmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{pmatrix}$$

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

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

- $\Sigma_{xx}, \Sigma_{yy}, \Lambda_{xx}, \Lambda_{yy}$ は対称. $\Sigma_{xy}^\top = \Sigma_{yx}^\top, \Lambda_{xy}^\top = \Lambda_{yx}$ であるから,

$$\Lambda_{xx} = (\Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1}$$

$$\Lambda_{xy} = -(\Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} = -\Lambda_{xx} \Sigma_{xy} \Sigma_{yy}^{-1}$$

$$\Lambda_{yx} = -\Sigma_{yy}^{-1} \Sigma_{yx} (\Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1} = -\Sigma_{yy}^{-1} \Sigma_{xy}^\top \Lambda_{xx}$$

$$\begin{aligned} \Lambda_{yy} &= \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1} \Sigma_{yx} (\Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \\ &= \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1} \Sigma_{xy}^\top \Lambda_{xx} \Sigma_{xy} \Sigma_{yy}^{-1} \end{aligned}$$

- これらを使うと, $\exp(\cdot)$ の中身は,

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

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:

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

- 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:

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

- 第1項: $p(\mathbf{x} | \mathbf{y}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Lambda}_{xx}^{-1})$ に対応
- 第2項: $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$ 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:

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

- 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:

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

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

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$$

- \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,

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

- ガウス分布 $p(\mathbf{y})$, $p(\mathbf{x} | \mathbf{y})$ 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:

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

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

- 上記より,

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

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、 $p(\mathbf{x} | \mathbf{y})$, $p(\mathbf{y})$ の指数部の和。
- $p(\mathbf{x}, \mathbf{y})$ の定数係数は、 $p(\mathbf{x} | \mathbf{y})$, $p(\mathbf{y})$ の定数係数の積。
- よって $p(\mathbf{x}, \mathbf{y})$ は、2つのガウス分布 $p(\mathbf{x} | \mathbf{y})$, $p(\mathbf{y})$ に分解できる。

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

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

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$$

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

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

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

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$

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:

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

- 同時分布 $p(\mathbf{x}, \mathbf{y})$ 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 $p(\mathbf{x} | \mathbf{y})$ と周辺分布 $p(\mathbf{y})$ もガウス分布となる:

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

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

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$$

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 $p(\mathbf{y} | \mathbf{x})$ と周辺分布 $p(\mathbf{x})$ もガウス分布となる:

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

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

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$

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:

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x} + \mathbf{b}] = \mathbf{A} \mathbb{E}[\mathbf{x}] + \mathbf{b} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$$

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

- 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) から確認できる:
- \mathbf{x} のモーメント母関数は,

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

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

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

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

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

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

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

ガウス分布の線形変換

- この式を確かめよう:

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

ガウス分布の線形変換

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

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

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

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

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.

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

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:

$$\mathbf{A}\boldsymbol{\mu} + \mathbf{b} = \mu_i$$

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top = (\boldsymbol{\Sigma})_{ii} = \text{Cov}(x_i, x_i) = \text{Var}[x_i] = \sigma_i^2$$

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

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

- 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, $p(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{R})$ になる (注意: \mathbf{x} は定数扱い).
- $p(\mathbf{y} | \mathbf{x})$ 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

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

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

- 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.
- $p(\mathbf{y} | \mathbf{x})$ 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, $p(\mathbf{x} | \mathbf{y})$ が得られる:

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{\int_{-\infty}^{\infty} p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) d\mathbf{x}}$$

- $p(\mathbf{x} | \mathbf{y})$ 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

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

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

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

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

- $\exp(\cdot)$ 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,

$$\begin{aligned} & (y - (Ax + b))^T R^{-1} (y - (Ax + b)) + (x - \mu_x)^T \Sigma_{xx}^{-1} (x - \mu_x) \\ &= y^T R^{-1} y - y^T R^{-1} Ax - x^T A^T R^{-1} y + x^T (A^T R^{-1} A + \Sigma_{xx}^{-1}) x \\ &\quad - (y - Ax)^T R^{-1} b - b^T R^{-1} (y - Ax) - \mu_x^T \Sigma_{xx}^{-1} x - x^T \Sigma_{xx}^{-1} \mu_x \\ &\quad + b^T R^{-1} b + \mu_x^T \Sigma_{xx}^{-1} \mu_x \\ &= \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} A^T R^{-1} A + \Sigma_{xx}^{-1} & -A^T R^{-1} \\ -R^{-1} A & R^{-1} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &\quad - \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} -A^T R^{-1} b + \Sigma_{xx}^{-1} \mu_x \\ R^{-1} b \end{pmatrix} \\ &\quad - (-b^T R^{-1} A + \mu_x^T \Sigma_{xx}^{-1} \quad b^T R^{-1}) \begin{pmatrix} x \\ y \end{pmatrix} + C \end{aligned}$$

- x, 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} , Σ_{xx} は対称行列だから、その逆行列も対称. よって,

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

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

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

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} のシュア補行列.

- この式を逆方向に使うと,

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

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,

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

ガウス分布の線形変換

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

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

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

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

ガウス分布の線形変換

- また,

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

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

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

ガウス分布の線形変換

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

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

- $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} | \mathbf{x})p(\mathbf{x})$ 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 $p(\mathbf{x} | \mathbf{y})$, $p(\mathbf{y})$ に分解できる:

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

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

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

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

- ただし、

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

ガウス分布の線形変換

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

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xx} \mathbf{A}^\top (\mathbf{R} + \mathbf{A} \Sigma_{xx} \mathbf{A}^\top)^{-1} \mathbf{A} \Sigma_{xx} = (\Sigma_{xx}^{-1} + \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A})^{-1}$$

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

- Woodbury の公式を用いた:

$$\begin{aligned} (\mathbf{D} + \mathbf{CAB})^{-1} &= \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{C} (\mathbf{A}^{-1} + \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} \mathbf{B} \mathbf{D}^{-1} \\ \mathbf{AB} (\mathbf{D} + \mathbf{CAB})^{-1} &= (\mathbf{A}^{-1} + \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} \mathbf{B} \mathbf{D}^{-1} \end{aligned}$$

ガウス分布の線形変換

ガウス分布の線形変換

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

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

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

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

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

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

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

ガウス分布の非線形変換 (線形化)

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

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

- $\mathbf{g}(\mathbf{x})$ 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, $p(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R})$ になる (注意: \mathbf{x} は定数扱い).
- $p(\mathbf{y} | \mathbf{x})$ 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)

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

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

- 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.
- $p(\mathbf{y} | \mathbf{x})$ 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, $p(\mathbf{x} | \mathbf{y})$ が得られる:

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{\int_{-\infty}^{\infty} p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) d\mathbf{x}}$$

- $p(\mathbf{x} | \mathbf{y})$ 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)

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

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

- 一般の $\mathbf{g}(\mathbf{x})$ に対して, $p(\mathbf{y} | \mathbf{x})$, $p(\mathbf{y})$ 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:

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

- $p(\mathbf{y} | \mathbf{x}) \approx \mathcal{N}(\boldsymbol{\mu}_y + \mathbf{G}(\mathbf{x} - \boldsymbol{\mu}_x), \mathbf{R})$ 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, $p(\mathbf{y} | \mathbf{x})$ が次のようになる:

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

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

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

- 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.

$$\mathbb{E}[\mathbf{z}] = \mathbb{E}[\mathbf{x} + \mathbf{y}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{y}] = \boldsymbol{\mu}_x + \boldsymbol{\mu}_y$$

$$\begin{aligned}\text{Var}[\mathbf{z}] &= \text{Var}[\mathbf{x} + \mathbf{y}] \\ &= \text{Var}[\mathbf{x}] + \text{Var}[\mathbf{y}] + \text{Cov}(\mathbf{x}, \mathbf{y}) + \text{Cov}(\mathbf{y}, \mathbf{x}) \\ &= \text{Var}[\mathbf{x}] + \text{Var}[\mathbf{y}] = \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{yy}\end{aligned}$$

- \mathbf{x} , \mathbf{y} は**独立 (無相関)** なので, $\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, $\text{Cov}(\mathbf{y}, \mathbf{x}) = \mathbf{0}$.

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:

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

ガウス分布の和

- 重み付き和の共分散は,

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

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

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),

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

ガウス分布の積

- $\Sigma^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}$ (対称), $\Sigma^{-1}\mu = \Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2$ とおいて,

$$\begin{aligned} & \mathbf{x}^\top (\Sigma_1^{-1} + \Sigma_2^{-1}) \mathbf{x} - \mathbf{x}^\top (\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2) \\ & \quad - (\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)^\top \mathbf{x} + \mu_1^\top \Sigma_1^{-1}\mu_1 + \mu_2^\top \Sigma_2^{-1}\mu_2 \\ & = \mathbf{x}^\top \Sigma^{-1} \mathbf{x} - \mathbf{x}^\top \Sigma^{-1} \mu - \mu^\top \Sigma^{-1} \mathbf{x} + \mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_2^\top \Sigma_2^{-1} \mu_2 \\ & = (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) - \mu^\top \mu + \mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_2^\top \Sigma_2^{-1} \mu_2 \\ & = (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) + \text{Const.} \end{aligned}$$

- \mathbf{x} によらない定数項をまとめて, Const. とした.
- よって, 2つのガウス分布 $\mathcal{N}(\mu_1, \Sigma_1)$, $\mathcal{N}(\mu_2, \Sigma_2)$ の積は,

$$\mathcal{N}(\mu_1, \Sigma_1) \mathcal{N}(\mu_2, \Sigma_2) \propto \exp\left((\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu)\right) \propto \mathcal{N}(\mu, \Sigma)$$

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 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \dots, \mathcal{N}(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)$ 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 ($-\frac{1}{2}$ 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

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

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

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

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

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 $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \dots, \mathcal{N}(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)$ 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、

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

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):

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

- ただし,

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

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認しよう.
- 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 0 とおく:

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

- 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,

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

- 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, $-\frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \boldsymbol{\Sigma}}} \boldsymbol{\Sigma}^{-1}$.
- $\boldsymbol{\Sigma}^{-1}$ 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) \implies (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) \iff (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:

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

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, $\mathbb{E}[\mathbf{x}\mathbf{y}^\top] = \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top$ だから,

$$\Sigma_{xy} = \text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top = \mathbf{0}$$

$$\Sigma_{yx} = \text{Cov}(\mathbf{y}, \mathbf{x}) = (\text{Cov}(\mathbf{x}, \mathbf{y}))^\top = \mathbf{0}$$

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

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

- 上記の議論を逆に進めることで, (独立) \implies (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

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

- 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

$$\mathbb{E}[\mathbf{y}] = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top (\boldsymbol{\mu} - \boldsymbol{\mu}) = \mathbf{0}$$

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

- 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.