

行列輪講: 第 6 回 確率分布, ガウス積分

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 17, 2023

目次

① 概要

② 確率分布

③ ガウス積分

目次

- 1 概要
- 2 確率分布
- 3 ガウス積分

このスライドの概要

- 確率分布と、ガウス積分について確認する
 - 確率密度関数, ベイズの定理, モーメント
 - エントロピー, KL ダイバージェンス, 相互情報量
 - 不偏推定量, クラメール・ラオ (Cramér–Rao) の下限
 - 積分の変数変換, 偶関数と奇関数の積分, ガウス積分
- 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

目次

① 概要

② 確率分布

③ ガウス積分

確率密度関数 (Probability Density Function; PDF)

- 確率変数 x が, ある確率密度関数 $p(x)$ に従うとする.
- $p(x)$ は, $[a, b]$ の範囲で定義されるとする.
- $p(x)$ は, 2 つの条件を満たす:

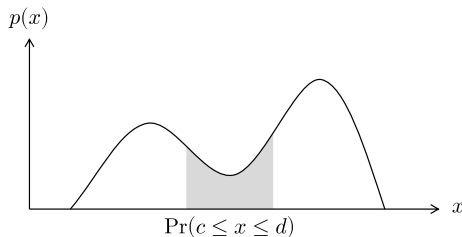
$$p(x) \geq 0, \quad \int_a^b p(x) dx = 1$$

確率密度関数と確率

- 確率密度と確率は別物である.
- x が, $[c, d]$ の範囲をとる**確率**を, $\Pr(c \leq x \leq d)$ とする.

$$\Pr(c \leq x \leq d) = \int_c^d p(x) dx$$

- $\Pr(c \leq x \leq d)$ は, 以下の図で, 影がかかった部分である.



条件付き確率分布 (Conditional Probability)

- 確率変数 $x \in [a, b]$, $y \in [r, s]$ を考える.
- y に条件付けられた, x の確率密度関数 $p(x | y)$ を考える.
- $p(x | y)$ を, **条件付き確率密度関数** という.
- $p(x | y)$ も, 2 つの条件を満たす:

$$p(x | y) \geq 0, \quad \int_a^b p(x | y) dx = 1$$

同時分布 (Joint Probability)

- 複数の確率変数 x_1, x_2, \dots, x_N を考える ($x_i \in [a_i, b_i]$).
- これらの確率密度関数 $p(x_1, x_2, \dots, x_N)$ を考える.
- $p(x_1, x_2, \dots, x_N)$ を, **同時確率密度関数**という.
- 確率変数をまとめて, $\mathbf{x} = (x_1, x_2, \dots, x_N)$ とかく.
- $p(\mathbf{x})$ も, 2 つの条件を満たす:

$$p(\mathbf{x}) = p(x_1, x_2, \dots, x_N) \geq 0$$
$$\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}) d\mathbf{x} = \int_{a_N}^{b_N} \cdots \int_{a_1}^{b_1} p(x_1, \dots, x_N) dx_1 \cdots dx_N = 1$$

- 積分の範囲は $\mathbf{a} = (a_1, a_2, \dots, a_N)$, $\mathbf{b} = (b_1, b_2, \dots, b_N)$ である.
- 確率変数の範囲については, 以後は明記しない.

加法定理 (Sum Rule), 周辺化 (Marginalization), 乗法定理

- 確率変数 \mathbf{x}, \mathbf{y} を考える.
- 同時分布 $p(\mathbf{x}, \mathbf{y})$ と, **周辺分布** $p(\mathbf{x})$ について, 次が成り立つ.
- 加法定理**, **周辺化**, **積分消去**とよばれる.

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

- 同時分布 $p(\mathbf{x}, \mathbf{y})$, 条件付き確率分布 $p(\mathbf{x} | \mathbf{y})$, $p(\mathbf{y} | \mathbf{x})$, 周辺分布 $p(\mathbf{x})$, $p(\mathbf{y})$ について, 次が成り立つ.
- 乗法定理**とよばれる.

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

- 同時分布は, 周辺分布と, 条件付き確率分布に**分解**できる.

独立 (Independent)

- 確率変数 x, y を考える.
- x, y が**独立**であるとする.
- y は x の確率分布に (x は y の確率分布に) 影響を与えない.
- 条件付き確率分布 $p(x | y)$ は, $p(x)$ に等しい.
 - y が与えられても, x についての確率分布は変化しない.
- 同様に, $p(y | x)$ は, $p(y)$ に等しい.
- 同時分布 $p(x, y)$ を, 個々の確率分布 $p(x), p(y)$ に分解できる.

$$\begin{aligned} p(x, y) &= p(x | y)p(y) = p(x)p(y) \quad (\because p(x | y) = p(x)) \\ &= p(y | x)p(x) = p(x)p(y) \quad (\because p(y | x) = p(y)) \end{aligned}$$

- $p(x, y) = p(x)p(y)$ であれば, x, y は独立である.

ベイズの定理 (Bayes' Theorem)

- 確率の乗法定理から、**ベイズの定理**が得られる:

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

- 分母は周辺化としてかける.

$$\begin{aligned} p(\mathbf{y}) &= p(\mathbf{y}) \underbrace{\int p(\mathbf{x} | \mathbf{y}) d\mathbf{x}}_{=1} = \int p(\mathbf{x} | \mathbf{y})p(\mathbf{y}) d\mathbf{x} \\ &= \int p(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \int p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) d\mathbf{x} \end{aligned}$$

- ベイズの定理は、次のようになる.

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) d\mathbf{x}}$$

ベイズの定理 (Bayes' Theorem)

ベイズの定理

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) d\mathbf{x}}$$

分母は, $p(\mathbf{x} \mid \mathbf{y})$ が確率分布になるための正規化 (\mathbf{x} による積分が 1).

ベイズの定理 (Bayes' Theorem)

- ベイズの定理:

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) d\mathbf{x}}$$

- センサデータ \mathbf{y} を手掛かりに, 変数 \mathbf{x} を推定したい.

- 例えば, \mathbf{y} は GPS のデータ, \mathbf{x} は GPS の位置.

1 \mathbf{x} に関する仮説を, **事前分布** (Prior) $p(\mathbf{x})$ として決める.

2 センサのモデルを, $p(\mathbf{y} \mid \mathbf{x})$ として決める.

- 変数 \mathbf{x} のもとで, どのようなセンサデータ \mathbf{y} が得られるのか?

3 新たなセンサデータ \mathbf{y} を得たら, $p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})$ を計算し, 正規化する.

4 \mathbf{y} を取り込んだ, \mathbf{x} の**事後分布** (Posterior) $p(\mathbf{x} \mid \mathbf{y})$ が得られる.

ベイズの定理 (Bayes' Theorem)

- 2 つの変数 y_1, y_2 から, 変数 x の事後分布 $p(x | y_1, y_2)$ を推定したい.

$$p(x | y_1, y_2) = \frac{p(y_1, y_2 | x)p(x)}{p(y_1, y_2)} \quad (\because \text{ベイズの定理})$$

- x のもとで y_1, y_2 が互いに独立なら, 以下が成り立つ:

$$p(y_1, y_2 | x) = p(y_1 | x)p(y_2 | x) = \frac{p(x | y_1)p(y_1)}{p(x)} \frac{p(x | y_2)p(y_2)}{p(x)}$$

- 上記を代入すれば, 次が得られる.

$$p(x | y_1, y_2) = \eta p(x | y_1)p(x | y_2)$$

- $\eta = \frac{p(y_1)p(y_2)}{p(y_1, y_2)p(x)}$ は正規化項である.

モーメント, 平均, 分散, 歪度, 尖度

- 確率分布 $p(x)$ の形状ではなく, モーメントだけを扱うことは多々ある.
- $\mathbb{E}[f(x)]$ を, 関数 $f(x)$ の, 確率分布 $p(x)$ のもとでの**期待値**とする.

$$\mathbb{E}[f(x)] = \int f(x)p(x) dx$$

- 1 次モーメント: **平均** $\mu = \mathbb{E}[x]$ (Mean)
- 2 次モーメント: **分散** $\sigma^2 = \text{Var}[x] = \mathbb{E}[(x - \mu)^2]$ (Variance)
- 3 次モーメント: $\mathbb{E}[(x - \mu)^3]$
- 4 次モーメント: $\mathbb{E}[(x - \mu)^4]$

$$\mu = \mathbb{E}[x] = \int xp(x) dx$$

$$\sigma^2 = \text{Var}[x] = \mathbb{E}[(x - \mu)^2] = \int (x - \mu)^2 p(x) dx$$

- 確率分布 $p(x)$ の歪度 (Skewness; わいと):

$$\mathbb{E} \left[\left(\frac{x - \mu}{\sigma} \right)^3 \right] = \frac{\mathbb{E} [(x - \mu)^3]}{\sigma^3} = \frac{\mathbb{E} [(x - \mu)^3]}{\mathbb{E} [(x - \mu)^2]^{\frac{3}{2}}}$$

- 確率分布 $p(x)$ の尖度 (Kurtosis; せんど):

$$\mathbb{E} \left[\left(\frac{x - \mu}{\sigma} \right)^4 \right] = \frac{\mathbb{E} [(x - \mu)^4]}{\sigma^4} = \frac{\mathbb{E} [(x - \mu)^4]}{\mathbb{E} [(x - \mu)^2]^2}$$

- 尖度は, $\mathbb{E} \left[\left(\frac{x - \mu}{\sigma} \right)^4 \right] - 3$ とすることもある.

- ガウス分布の尖度を 0 とするため.

モーメント, 平均, 共分散 (多変数の場合)

- 行列関数 $\mathbf{F}(\mathbf{x})$ の, 確率分布 $p(\mathbf{x})$ のもとでの期待値:

$$\mathbb{E}[\mathbf{F}(\mathbf{x})] = \int \mathbf{F}(\mathbf{x})p(\mathbf{x}) d\mathbf{x}$$

- $\mathbb{E}[\mathbf{F}(\mathbf{x})]$ の (i, j) 成分は, \mathbf{F} の (i, j) 成分 f_{ij} を用いて,

$$(\mathbb{E}[\mathbf{F}(\mathbf{x})])_{ij} = \mathbb{E}[f_{ij}(\mathbf{x})] = \int f_{ij}(\mathbf{x})p(\mathbf{x}) d\mathbf{x}$$

- 平均 $\boldsymbol{\mu}$ と共分散 $\boldsymbol{\Sigma}$ (Covariance) は, 次のようにかける.

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \int \mathbf{x}p(\mathbf{x}) d\mathbf{x}$$

$$\boldsymbol{\Sigma} = \text{Var}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\right] = \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top p(\mathbf{x}) d\mathbf{x}$$

モーメント, 平均, 共分散 (多変数の場合)

- 2 つの変数 \mathbf{x}, \mathbf{y} について, 次が成り立つ.

$$\mathbb{E}[\mathbf{x} + \mathbf{y}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{y}]$$

- n 個の変数 $\mathbf{x}_1, \dots, \mathbf{x}_n$ について, 次が成り立つ.

$$\mathbb{E}\left[\sum_{i=1}^n \mathbf{x}_i\right] = \sum_{i=1}^n \mathbb{E}[\mathbf{x}_i]$$

- 和の期待値と, 期待値の和が等しい.
- スカラの場合にも, 同様に成立する.

$$\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y], \quad \mathbb{E}\left[\sum_{i=1}^n x_i\right] = \sum_{i=1}^n \mathbb{E}[x_i]$$

モーメント, 平均, 共分散 (多変数の場合)

- 変数 \mathbf{x} と, ある定数 \mathbf{A} について, 次が成り立つ.

$$\mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A} \mathbb{E}[\mathbf{x}], \quad \mathbb{E}[\mathbf{x}\mathbf{A}] = \mathbb{E}[\mathbf{x}] \mathbf{A}$$

- 変数 \mathbf{x} と, ある定数 \mathbf{a} について, 次が成り立つ.

$$\mathbb{E}[\mathbf{x} + \mathbf{a}] = \mathbb{E}[\mathbf{x}] + \mathbf{a}$$

- スカラの場合にも, 同様に成立する.

$$\mathbb{E}[ax] = a \mathbb{E}[x], \quad \mathbb{E}[x + a] = \mathbb{E}[x] + a$$

モーメント, 平均, 共分散 (多変数の場合)

- 確率分布 $p(\mathbf{x})$ の共分散が $\Sigma = \text{Var}[\mathbf{x}]$ であるとする.
- $\mathbf{y} = \mathbf{A}\mathbf{x}$ としたとき, $p(\mathbf{y})$ の共分散 $\text{Var}[\mathbf{y}]$ は,

$$\begin{aligned}\text{Var}[\mathbf{y}] &= \mathbb{E} \left[(\mathbf{y} - \mathbb{E}[\mathbf{y}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] \\ &= \text{(練習問題)} \\ &= \mathbf{A}\Sigma\mathbf{A}^\top\end{aligned}$$

- ただし, $\mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A} \mathbb{E}[\mathbf{x}]$.
- 同様に, $p(x)$ の分散が σ^2 であれば, $y = ax$ の分散は $a^2\sigma^2$.

モーメント, 平均, 共分散 (多変数の場合)

- 確率分布 $p(\mathbf{x})$ の共分散が $\Sigma = \text{Var}[\mathbf{x}]$ であるとする.
- $\mathbf{y} = \mathbf{x} + \mathbf{a}$ としたとき (\mathbf{a} は定数), $p(\mathbf{y})$ の共分散 $\text{Var}[\mathbf{y}]$ は,

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

- ただし, $\mathbb{E}[\mathbf{x} + \mathbf{a}] = \mathbb{E}[\mathbf{x}] + \mathbf{a}$.
- 同様に, $p(x)$ の分散が σ^2 であれば, $y = x + a$ の分散も σ^2 .

モーメント, 平均, 共分散 (多変数の場合)

- 確率分布 $p(\mathbf{x})$ の共分散 $\Sigma = \text{Var}[\mathbf{x}]$ について, 次が成り立つ.

$$\Sigma = \text{Var}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^\top = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

- スカラーの場合は, 次のようになる.

$$\sigma^2 = \text{Var}[x] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2 = \mathbb{E}[x^2] - \mu^2$$

- 練習問題.

独立 (Independent), 無相関 (Uncorrelated)

- 以下が成り立つとき, \mathbf{x}, \mathbf{y} は**独立**である:

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

- 以下が成り立つとき, \mathbf{x}, \mathbf{y} は**無相関**である:

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

- 独立 \Rightarrow 無相関は成り立つが, 無相関 \Rightarrow 独立は**成り立たない**.
- 独立 \Rightarrow 無相関は, **練習問題**.

相互共分散 (Cross-covariance)

- \mathbf{x} と \mathbf{y} との間の**相互共分散** $\text{Cov}(\mathbf{x}, \mathbf{y})$ は, 次のようにかける.

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

- 左辺を展開すれば, 次のように示せる.

$$\begin{aligned} & \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top] \\ &= \mathbb{E}[\mathbf{x}\mathbf{y}^\top] - \mathbb{E}[\mathbf{x}\mathbb{E}[\mathbf{y}]^\top] - \mathbb{E}[\mathbb{E}[\mathbf{x}]\mathbf{y}^\top] + \mathbb{E}[\mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top] \\ &= \mathbb{E}[\mathbf{x}\mathbf{y}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top + \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top \end{aligned}$$

- \mathbf{x}, \mathbf{y} が無相関 ($\mathbb{E}[\mathbf{x}\mathbf{y}^\top] = \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top$) であれば, 相互共分散は 0.
- 独立 \Rightarrow 無相関であるから, 独立であれば, 相互共分散は 0.

相互共分散 (Cross-covariance)

- $\text{Cov}(\mathbf{x}, \mathbf{y}) = \text{Cov}(\mathbf{y}, \mathbf{x})^\top$ が成り立つ:

$$\begin{aligned}\text{Cov}(\mathbf{x}, \mathbf{y}) &= \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] \\ &= \mathbb{E} \left[(\mathbf{y} - \mathbb{E}[\mathbf{y}]) (\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top \right]^\top = \text{Cov}(\mathbf{y}, \mathbf{x})^\top\end{aligned}$$

- $\text{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \text{Cov}(\mathbf{x}, \mathbf{z}) + \text{Cov}(\mathbf{y}, \mathbf{z})$ も成り立つ:

$$\begin{aligned}\text{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{z}) &= \mathbb{E} \left[(\mathbf{x} + \mathbf{y} - \mathbb{E}[\mathbf{x} + \mathbf{y}]) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top \right] \\ &= \mathbb{E} \left[((\mathbf{x} - \mathbb{E}[\mathbf{x}]) + (\mathbf{y} - \mathbb{E}[\mathbf{y}])) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top \right] \\ &= \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top \right] + \mathbb{E} \left[(\mathbf{y} - \mathbb{E}[\mathbf{y}]) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top \right] \\ &= \text{Cov}(\mathbf{x}, \mathbf{z}) + \text{Cov}(\mathbf{y}, \mathbf{z})\end{aligned}$$

- ただし, $\mathbb{E}[\mathbf{x} + \mathbf{y}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{y}]$.

相互共分散 (Cross-covariance)

- $\text{Cov}(\mathbf{Ax}, \mathbf{By}) = \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}^\top$ も成り立つ:

$$\begin{aligned}\text{Cov}(\mathbf{Ax}, \mathbf{By}) &= \mathbb{E} \left[(\mathbf{Ax} - \mathbb{E}[\mathbf{Ax}]) (\mathbf{By} - \mathbb{E}[\mathbf{By}])^\top \right] \\ &= \mathbb{E} \left[\mathbf{A} (\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{B} (\mathbf{y} - \mathbb{E}[\mathbf{y}]))^\top \right] \\ &= \mathbb{E} \left[\mathbf{A} (\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \mathbf{B}^\top \right] \\ &= \mathbf{A} \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] \mathbf{B}^\top = \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}^\top\end{aligned}$$

- $\text{Cov}(\mathbf{x} + \mathbf{a}, \mathbf{y} + \mathbf{b}) = \text{Cov}(\mathbf{x}, \mathbf{y})$ も成り立つ:

$$\begin{aligned}\text{Cov}(\mathbf{x} + \mathbf{a}, \mathbf{y} + \mathbf{b}) &= \mathbb{E} [(\mathbf{x} + \mathbf{a} - \mathbb{E}[\mathbf{x} + \mathbf{a}]) (\mathbf{y} + \mathbf{b} - \mathbb{E}[\mathbf{y} + \mathbf{b}])] \\ &= \mathbb{E} [(\mathbf{x} + \mathbf{a} - \mathbb{E}[\mathbf{x}] - \mathbf{a}) (\mathbf{y} + \mathbf{b} - \mathbb{E}[\mathbf{y}] - \mathbf{b})] \\ &= \mathbb{E} [(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top] = \text{Cov}(\mathbf{x}, \mathbf{y})\end{aligned}$$

共分散と相互共分散

- \mathbf{x}, \mathbf{y} の共分散を $\text{Var}[\mathbf{x}], \text{Var}[\mathbf{y}]$, 相互共分散を $\text{Cov}(\mathbf{x}, \mathbf{y})$ とする.
- $\mathbf{z} = \mathbf{x} + \mathbf{y}$ としたとき, $p(\mathbf{z})$ の共分散 $\text{Var}[\mathbf{z}]$ は,

$$\begin{aligned} & \mathbb{E} \left[(\mathbf{z} - \mathbb{E}[\mathbf{z}]) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top \right] \\ &= \mathbb{E} \left[(\mathbf{x} + \mathbf{y} - \mathbb{E}[\mathbf{x} + \mathbf{y}]) (\mathbf{x} + \mathbf{y} - \mathbb{E}[\mathbf{x} + \mathbf{y}])^\top \right] \\ &= \mathbb{E} \left[((\mathbf{x} - \mathbb{E}[\mathbf{x}]) + (\mathbf{y} - \mathbb{E}[\mathbf{y}])) ((\mathbf{x} - \mathbb{E}[\mathbf{x}]) + (\mathbf{y} - \mathbb{E}[\mathbf{y}]))^\top \right] \\ &= \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top \right] + \mathbb{E} \left[(\mathbf{y} - \mathbb{E}[\mathbf{y}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] \\ &\quad + \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] + \mathbb{E} \left[(\mathbf{y} - \mathbb{E}[\mathbf{y}]) (\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top \right] \\ &= \text{Var}[\mathbf{x}] + \text{Var}[\mathbf{y}] + \text{Cov}(\mathbf{x}, \mathbf{y}) + \text{Cov}(\mathbf{y}, \mathbf{x}) \end{aligned}$$

- 同様に, $z = x + y$ の分散は, $\text{Var}[z] = \text{Var}[x] + \text{Var}[y] + 2 \text{Cov}(x, y)$.

平均, 共分散, 相互共分散 (1 変数の場合)

平均, 共分散, 相互共分散 (1 変数の場合)

$$\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y]$$

$$\mathbb{E}[x + a] = \mathbb{E}[x] + a$$

$$\mathbb{E}[ax] = a \mathbb{E}[x]$$

$$\text{Var}[x] = \mathbb{E}[(x - \mathbb{E}[x])^2]$$

$$\text{Var}[ax] = a^2 \text{Var}[x]$$

$$\text{Var}[x + a] = \text{Var}[x]$$

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

平均, 共分散, 相互共分散 (1 変数の場合)

平均, 共分散, 相互共分散 (1 変数の場合)

$$\text{Cov}(x, y) = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$$

$$\text{Cov}(x, y) = \mathbb{E}[xy] - \mathbb{E}[x] \mathbb{E}[y]$$

$$\text{Cov}(x + y, z) = \text{Cov}(x, z) + \text{Cov}(y, z)$$

$$\text{Cov}(ax, by) = ab \text{Cov}(x, y)$$

$$\text{Cov}(x + a, y + b) = \text{Cov}(x, y)$$

$$\text{Var}[x + y] = \text{Var}[x] + \text{Var}[y] + 2 \text{Cov}(x, y)$$

平均, 共分散, 相互共分散 (1 変数の場合)

無相関 (Uncorrelated)

x, y が無相関であれば, $\mathbb{E}[xy] = \mathbb{E}[x] \mathbb{E}[y]$.

独立であれば ($p(x, y) = p(x)p(y)$), 無相関.

$$\text{Cov}(x, y) = 0$$

$$\text{Var}[x + y] = \text{Var}[x] + \text{Var}[y]$$

平均, 共分散, 相互共分散 (多変数の場合)

平均, 共分散, 相互共分散 (多変数の場合)

$$\mathbb{E}[\mathbf{x} + \mathbf{y}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{y}]$$

$$\mathbb{E}[\mathbf{x} + \mathbf{a}] = \mathbb{E}[\mathbf{x}] + \mathbf{a}$$

$$\mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A} \mathbb{E}[\mathbf{x}]$$

$$\mathbb{E}[\mathbf{x}\mathbf{A}] = \mathbb{E}[\mathbf{x}] \mathbf{A}$$

$$\text{Var}[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top]$$

$$\text{Var}[\mathbf{A}\mathbf{x}] = \mathbf{A} \text{Var}[\mathbf{x}] \mathbf{A}^\top$$

$$\text{Var}[\mathbf{x} + \mathbf{a}] = \text{Var}[\mathbf{x}]$$

$$\text{Var}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}] \mathbb{E}[\mathbf{x}]^\top$$

平均, 共分散, 相互共分散 (多変数の場合)

平均, 共分散, 相互共分散 (多変数の場合)

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

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

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

$$\text{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \text{Cov}(\mathbf{x}, \mathbf{z}) + \text{Cov}(\mathbf{y}, \mathbf{z})$$

$$\text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) = \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}^\top$$

$$\text{Cov}(\mathbf{x} + \mathbf{a}, \mathbf{y} + \mathbf{b}) = \text{Cov}(\mathbf{x}, \mathbf{y})$$

$$\text{Var}[\mathbf{x} + \mathbf{y}] = \text{Var}[\mathbf{x}] + \text{Var}[\mathbf{y}] + \text{Cov}(\mathbf{x}, \mathbf{y}) + \text{Cov}(\mathbf{y}, \mathbf{x})$$

平均, 共分散, 相互共分散 (多変数の場合)

無相関 (Uncorrelated)

\mathbf{x}, \mathbf{y} が無相関であれば, $\mathbb{E} [\mathbf{x}\mathbf{y}^\top] = \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{y}]^\top$.

独立であれば ($p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$), 無相関.

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$$

$$\text{Var} [\mathbf{x} + \mathbf{y}] = \text{Var} [\mathbf{x}] + \text{Var} [\mathbf{y}]$$

平均, 共分散, 相互共分散の関係 (発展)

二次形式

$$\mathbb{E} \left[(\mathbf{Ax} + \mathbf{b})^\top (\mathbf{Cx} + \mathbf{d}) \right] = \text{tr}(\mathbf{A} \text{Var}[\mathbf{x}] \mathbf{C}^\top) + (\mathbf{A} \mathbb{E}[\mathbf{x}] + \mathbf{b})^\top (\mathbf{C} \mathbb{E}[\mathbf{x}] + \mathbf{d})$$

- これまでにみてきた公式を駆使すれば, 導出できる.
- 上式を基に, 次が直ちに得られる:

$$\mathbb{E} \left[\mathbf{x}^\top \mathbf{x} \right] = \text{tr}(\text{Var}[\mathbf{x}]) + \mathbb{E}[\mathbf{x}]^\top \mathbb{E}[\mathbf{x}]$$

$$\mathbb{E} \left[\mathbf{x}^\top \mathbf{Cx} \right] = \text{tr}(\mathbf{C} \text{Var}[\mathbf{x}]) + \mathbb{E}[\mathbf{x}]^\top \mathbf{C} \mathbb{E}[\mathbf{x}]$$

$$\mathbb{E} \left[(\mathbf{Ax})^\top (\mathbf{Ax}) \right] = \text{tr}(\mathbf{A} \text{Var}[\mathbf{x}] \mathbf{A}^\top) + (\mathbf{A} \mathbb{E}[\mathbf{x}])^\top (\mathbf{A} \mathbb{E}[\mathbf{x}])$$

$$\mathbb{E} \left[(\mathbf{x} + \mathbf{a})^\top (\mathbf{x} + \mathbf{a}) \right] = \text{tr}(\text{Var}[\mathbf{x}]) + (\mathbb{E}[\mathbf{x}] + \mathbf{a})^\top (\mathbb{E}[\mathbf{x}] + \mathbf{a})$$

平均, 共分散, 相互共分散の関係 (発展)

- $\mathbb{E}[\cdot]$ の中身がスカラーであることを考えて, 式を変形すると,

$$\begin{aligned} & \mathbb{E} \left[(\mathbf{Ax} + \mathbf{b})^\top (\mathbf{Cx} + \mathbf{d}) \right] \\ &= \mathbb{E} \left[\text{tr} \left((\mathbf{Cx} + \mathbf{d}) (\mathbf{Ax} + \mathbf{b})^\top \right) \right] = \text{tr} \left(\mathbb{E} \left[(\mathbf{Cx} + \mathbf{d}) (\mathbf{Ax} + \mathbf{b})^\top \right] \right) \\ &= \text{tr} \left(\text{Cov}(\mathbf{Cx} + \mathbf{d}, \mathbf{Ax} + \mathbf{b}) + \mathbb{E}[\mathbf{Cx} + \mathbf{d}] \mathbb{E}[\mathbf{Ax} + \mathbf{b}]^\top \right) \\ &= \text{tr}(\text{Cov}(\mathbf{Cx} + \mathbf{d}, \mathbf{Ax} + \mathbf{b})) + \text{tr} \left(\mathbb{E}[\mathbf{Cx} + \mathbf{d}] \mathbb{E}[\mathbf{Ax} + \mathbf{b}]^\top \right) \\ &= \text{tr}(\text{Cov}(\mathbf{Cx} + \mathbf{d}, \mathbf{Ax} + \mathbf{b})) + \mathbb{E}[\mathbf{Ax} + \mathbf{b}]^\top \mathbb{E}[\mathbf{Cx} + \mathbf{d}] \\ &= \text{tr}(\text{Cov}(\mathbf{Cx} + \mathbf{d}, \mathbf{Ax} + \mathbf{b})) + (\mathbf{A} \mathbb{E}[\mathbf{x}] + \mathbf{b})^\top (\mathbf{C} \mathbb{E}[\mathbf{x}] + \mathbf{d}) \end{aligned}$$

- ただし, 以下の関係を用いた:

$$\mathbb{E}[\text{tr}(\mathbf{A})] = \text{tr}(\mathbb{E}[\mathbf{A}]), \quad \text{tr}(\mathbf{ab}^\top) = \mathbf{b}^\top \mathbf{a}$$

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

平均, 共分散, 相互共分散の関係 (発展)

- 続いて, $\text{tr}(\text{Cov}(\mathbf{C}\mathbf{x} + \mathbf{d}, \mathbf{A}\mathbf{x} + \mathbf{b})) = \text{tr}(\mathbf{A} \text{Var}[\mathbf{x}] \mathbf{C}^\top)$ を示す.

$$\begin{aligned}\text{tr}(\text{Cov}(\mathbf{C}\mathbf{x} + \mathbf{d}, \mathbf{A}\mathbf{x} + \mathbf{b})) &= \text{tr}(\text{Cov}(\mathbf{C}\mathbf{x}, \mathbf{A}\mathbf{x})) \\ &= \text{tr}(\mathbf{C} \text{Cov}(\mathbf{x}, \mathbf{x}) \mathbf{A}^\top) \\ &= \text{tr}(\mathbf{C} \text{Var}[\mathbf{x}] \mathbf{A}^\top) = \text{tr}(\mathbf{A} \text{Var}[\mathbf{x}] \mathbf{C}^\top)\end{aligned}$$

- ただし, 以下の関係を用いた:

$$\begin{aligned}\text{Cov}(\mathbf{x} + \mathbf{a}, \mathbf{y} + \mathbf{b}) &= \text{Cov}(\mathbf{x}, \mathbf{y}) \\ \text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) &= \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}^\top \\ \text{Cov}(\mathbf{x}, \mathbf{x}) &= \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top] = \text{Var}[\mathbf{x}] \\ \text{tr}(\mathbf{A}^\top) &= \text{tr}(\mathbf{A}), \quad \text{Var}[\mathbf{x}]^\top = \text{Var}[\mathbf{x}]\end{aligned}$$

二次関数

$$\mathbb{E} \left[(\mathbf{Ax} + \mathbf{b}) (\mathbf{Cx} + \mathbf{d})^\top \right] = \mathbf{A} \text{Var} [\mathbf{x}] \mathbf{C}^\top + (\mathbf{A} \mathbb{E} [\mathbf{x}] + \mathbf{b}) (\mathbf{C} \mathbb{E} [\mathbf{x}] + \mathbf{d})^\top$$

- 上式を基に, 次が直ちに得られる:

$$\mathbb{E} [\mathbf{xx}^\top] = \text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top$$

$$\mathbb{E} \left[(\mathbf{Ax}) (\mathbf{Ax})^\top \right] = \mathbf{A} \left(\text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top \right) \mathbf{A}^\top$$

$$\mathbb{E} \left[(\mathbf{x} + \mathbf{a}) (\mathbf{x} + \mathbf{a})^\top \right] = \text{Var} [\mathbf{x}] + (\mathbb{E} [\mathbf{x}] + \mathbf{a}) (\mathbb{E} [\mathbf{x}] + \mathbf{a})^\top$$

平均, 共分散, 相互共分散の関係 (発展)

- 次のように変形できる:

$$\begin{aligned}\mathbb{E} \left[(\mathbf{Ax} + \mathbf{b}) (\mathbf{Cx} + \mathbf{d})^\top \right] \\ &= \textcolor{red}{(練習問題)} \\ &= \mathbf{A} \operatorname{Var} [\mathbf{x}] \mathbf{C}^\top + (\mathbf{A} \mathbb{E} [\mathbf{x}] + \mathbf{b}) (\mathbf{C} \mathbb{E} [\mathbf{x}] + \mathbf{d})^\top\end{aligned}$$

- ただし, 以下の関係を用いた:

$$\begin{aligned}\mathbb{E} [\mathbf{xy}^\top] &= \operatorname{Cov} (\mathbf{x}, \mathbf{y}) + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{y}]^\top \\ \operatorname{Cov} (\mathbf{x} + \mathbf{a}, \mathbf{y} + \mathbf{b}) &= \operatorname{Cov} (\mathbf{x}, \mathbf{y}), \quad \operatorname{Cov} (\mathbf{Ax}, \mathbf{By}) = \mathbf{A} \operatorname{Cov} (\mathbf{x}, \mathbf{y}) \mathbf{B}^\top \\ \operatorname{Cov} (\mathbf{x}, \mathbf{x}) &= \mathbb{E} \left[(\mathbf{x} - \mathbb{E} [\mathbf{x}]) (\mathbf{x} - \mathbb{E} [\mathbf{x}])^\top \right] = \operatorname{Var} [\mathbf{x}]\end{aligned}$$

平均, 共分散, 相互共分散の関係 (発展)

二次関数

$$\begin{aligned}\mathbb{E} [\mathbf{x} \mathbf{a}^\top \mathbf{x}] &= \left(\text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top \right) \mathbf{a} \\ \mathbb{E} [\mathbf{x}^\top \mathbf{a} \mathbf{x}] &= \mathbf{a}^\top \left(\text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top \right)\end{aligned}$$

- 次のように確認できる:

$$\begin{aligned}\mathbb{E} [\mathbf{x} (\mathbf{a}^\top \mathbf{x})] &= \mathbb{E} [\mathbf{x} \mathbf{x}^\top \mathbf{a}] = \mathbb{E} [\mathbf{x} \mathbf{x}^\top] \mathbf{a} \\ &= \left(\text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top \right) \mathbf{a} \\ \mathbb{E} [(\mathbf{x}^\top \mathbf{a}) \mathbf{x}] &= \mathbb{E} [\mathbf{a}^\top \mathbf{x} \mathbf{x}^\top] = \mathbf{a}^\top \mathbb{E} [\mathbf{x} \mathbf{x}^\top] \\ &= \mathbf{a}^\top \left(\text{Var} [\mathbf{x}] + \mathbb{E} [\mathbf{x}] \mathbb{E} [\mathbf{x}]^\top \right)\end{aligned}$$

イェンセンの不等式 (Jensen's Inequality)

- 下に凸な関数 $f(\mathbf{x})$ を考える.
- $p(\mathbf{x})$ を, $\int p(\mathbf{x}) d\mathbf{x} = 1$ をみたす関数とする.
- $f(\mathbf{x}), p(\mathbf{x})$ について, **イェンセンの不等式**が成り立つ:

$$\int f(\mathbf{x})p(\mathbf{x}) d\mathbf{x} \geq f\left(\int \mathbf{x}p(\mathbf{x}) d\mathbf{x}\right)$$

- $p(\mathbf{x})$ を確率分布とすれば, 期待値 $\mathbb{E}[\cdot]$ を用いて, 次のようにかける:

$$\mathbb{E}[f(\mathbf{x})] \geq f(\mathbb{E}[\mathbf{x}])$$

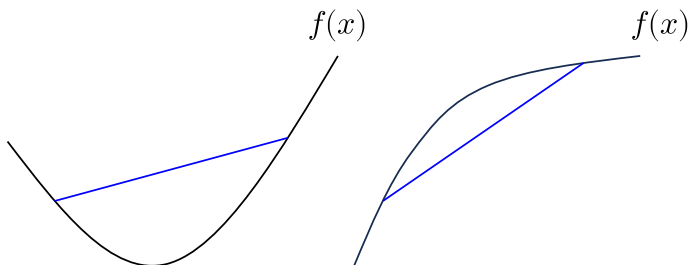
- $f(\mathbf{x})$ が上に凸であれば, 次が成り立つ (不等号が逆になる):

$$\int f(\mathbf{x})p(\mathbf{x}) d\mathbf{x} \leq f\left(\int \mathbf{x}p(\mathbf{x}) d\mathbf{x}\right)$$

$$\mathbb{E}[f(\mathbf{x})] \leq f(\mathbb{E}[\mathbf{x}])$$

イェンセンの不等式 (Jensen's Inequality)

- 下に凸であれば, 任意の 2 点を結ぶ直線が, 上側にある (凸関数, 左).
 - x^2 , $x \ln x$, $\frac{1}{x}$ ($x > 0$), $\tan x$ ($0 < x < \frac{\pi}{2}$) など
- 上に凸であれば, 任意の 2 点を結ぶ直線が, 下側にある (凹関数, 右).
 - $\ln x$, $\tan x$ ($-\frac{\pi}{2} < x < 0$) など



エントロピー (Entropy)

- 確率分布 $p(\mathbf{x})$ について, 次の量 $H[\mathbf{x}]$ を考える.

$$H[\mathbf{x}] = -\mathbb{E}[\ln p(\mathbf{x})] = -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$$

- $H[\mathbf{x}]$ を, **エントロピー**, **シャノン情報量**, **平均情報量**とよぶ.
- 直感的には, \mathbf{x} を, どのくらい予測しづらいのかを表す.
 - \mathbf{x} が離散変数であれば, 一様分布のときにエントロピーが最大である.
 - \mathbf{x} が連続変数であれば, 正規分布 (ガウス分布) のときに最大.
 - この証明には, 変分計算が必要である.

条件付きエントロピー (Conditional Entropy)

- 同時分布 $p(\mathbf{x}, \mathbf{y})$ のエントロピー $H[\mathbf{x}, \mathbf{y}]$ を考える.

$$H[\mathbf{x}, \mathbf{y}] = -\mathbb{E}[\ln p(\mathbf{x}, \mathbf{y})] = -\iint p(\mathbf{x}, \mathbf{y}) \ln p(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}$$

- $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} | \mathbf{x})p(\mathbf{x})$ のように分解できるので, 代入すれば

$$\begin{aligned} H[\mathbf{x}, \mathbf{y}] &= -\iint p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) (\ln p(\mathbf{y} | \mathbf{x}) + \ln p(\mathbf{x})) \, d\mathbf{x} \, d\mathbf{y} \\ &= -\iint p(\mathbf{x}, \mathbf{y}) \ln p(\mathbf{y} | \mathbf{x}) \, d\mathbf{x} \, d\mathbf{y} \\ &\quad - \int p(\mathbf{x}) \ln p(\mathbf{x}) \, d\mathbf{x} \underbrace{\int p(\mathbf{y} | \mathbf{x}) \, d\mathbf{y}}_{=1} \\ &= H[\mathbf{y} | \mathbf{x}] + H[\mathbf{x}] \end{aligned}$$

条件付きエントロピー (Conditional Entropy)

- 条件付き確率分布 $p(y | x)$ について, $H[y | x]$ を次のように定める.

$$H[y | x] = -\mathbb{E}[\ln p(y | x)] = -\iint p(x, y) \ln p(y | x) dx dy$$

- $H[y | x]$ を, **条件付きエントロピー**とよぶ.
- 次が成り立つ:

$$H[x, y] = H[y | x] + H[x] = H[x | y] + H[y]$$

- $H[x]$: x の予測しづらさ
- $H[x, y]$: x と y の予測しづらさ
- $H[y | x]$: x が既知であるときの, y の予測しづらさ

カルバック-ライブラーダイバージェンス (Kullback-Leibler Divergence)

- 2つの確率分布 $p(\mathbf{x}), q(\mathbf{x})$ があるとき, 次の量 $\text{KL}(p \parallel q)$ を考える.

$$\text{KL}(p \parallel q) = -\mathbb{E} \left[\ln \frac{q(\mathbf{x})}{p(\mathbf{x})} \right] = - \int p(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$$

- $\text{KL}(p \parallel q)$ を, **カルバック-ライブラーダイバージェンス**とよぶ.
- 長いので, **KL ダイバージェンス**ともよぶ.
- 確率分布 $p(\mathbf{x})$ と $q(\mathbf{x})$ との距離のような概念である.
- $\text{KL}(p \parallel q) \neq \text{KL}(q \parallel p)$ であるから, 厳密な距離ではない.

カルバック-ライブラーダイバージェンス (Kullback-Leibler Divergence)

- カルバック-ライブラーダイバージェンス:

$$\text{KL}(p \parallel q) = -\mathbb{E} \left[\ln \frac{q(\mathbf{x})}{p(\mathbf{x})} \right] = - \int p(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$$

- $\text{KL}(p \parallel q) \geq 0$ (練習問題).
- $\forall \mathbf{x} \ p(\mathbf{x}) = q(\mathbf{x})$ のときのみ 0.

相互情報量 (Mutual Information)

- 確率変数 x, y について, 次の量 $I(x, y)$ を考える.

$$I(x, y) = \mathbb{E} \left[\ln \frac{p(x, y)}{p(x)p(y)} \right] = \iint p(x, y) \ln \frac{p(x, y)}{p(x)p(y)} dx dy$$

- $I(x, y)$ を, **相互情報量**とよぶ.
- $x(y)$ を得たとき, $y(x)$ についての情報がどのくらい増えるか.
- x, y が独立であれば ($p(x, y) = p(x)p(y)$), $I(x, y) = 0$.

$$I(x, y) = \iint p(x, y) \ln \frac{p(x)p(y)}{p(x)p(y)} dx dy = \iint p(x, y) \ln 1 dx dy = 0$$

- x, y が独立であれば, $x(y)$ のことが分かってても, $y(x)$ については何の情報ももたらさない.

相互情報量 (Mutual Information)

- 相互情報量 $I(\mathbf{x}, \mathbf{y})$:

$$I(\mathbf{x}, \mathbf{y}) = \mathbb{E} \left[\ln \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right] = \iint p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x} d\mathbf{y}$$

- エントロピー $H[\mathbf{x}]$, $H[\mathbf{y}]$, $H[\mathbf{x}, \mathbf{y}]$ を使って, 次のようにかける.

$$I(\mathbf{x}, \mathbf{y}) = H[\mathbf{x}] + H[\mathbf{y}] - H[\mathbf{x}, \mathbf{y}]$$

- $H[\mathbf{x}, \mathbf{y}] = H[\mathbf{y} | \mathbf{x}] + H[\mathbf{x}] = H[\mathbf{x} | \mathbf{y}] + H[\mathbf{y}]$ であるから,

$$I(\mathbf{x}, \mathbf{y}) = H[\mathbf{y}] - H[\mathbf{y} | \mathbf{x}] = H[\mathbf{x}] - H[\mathbf{x} | \mathbf{y}]$$

- $H[\mathbf{y}]$: \mathbf{y} の予測しづらさ
- $H[\mathbf{y} | \mathbf{x}]$: \mathbf{x} が既知であるときの, \mathbf{y} の予測しづらさ
- $I(\mathbf{x}, \mathbf{y})$: \mathbf{x} が分かったとき, \mathbf{y} がどのくらい予測しやすくなるか

相互情報量 (Mutual Information)

- 相互情報量 $I(\mathbf{x}, \mathbf{y})$:

$$I(\mathbf{x}, \mathbf{y}) = \mathbb{E} \left[\ln \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right] = \iint p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x} d\mathbf{y}$$

- カルバック-ライブラーダイバージェンスを用いると,

$$I(\mathbf{x}, \mathbf{y}) = -\mathbb{E} \left[\ln \frac{p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x}, \mathbf{y})} \right] = \text{KL}(p(\mathbf{x}, \mathbf{y}) \parallel p(\mathbf{x})p(\mathbf{y}))$$

- KL ダイバージェンスの性質 (≥ 0) から, $I(\mathbf{x}, \mathbf{y}) \geq 0$.
- $\forall \mathbf{x}, \mathbf{y} \ p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ のとき (独立であるとき) に限って 0.
- 確率変数 \mathbf{x}, \mathbf{y} が, 独立に近いかどうかを表している.

不偏推定量 (Unbiased Estimate)

- パラメータ θ をもつ \mathbf{x} の確率分布 $p(\mathbf{x} \mid \theta)$ を考える.
- この確率分布から, 観測データが得られたとする.

$$\mathbf{x}_{\text{meas}} \leftarrow p(\mathbf{x} \mid \theta)$$

- 観測データ \mathbf{x}_{meas} を基に, パラメータ θ を推定したい.
- パラメータ θ の推定量 $\hat{\theta}$ が, **不偏推定量**であるとき, 以下が成り立つ:

$$\mathbb{E}[\hat{\theta}] = \theta \quad \longrightarrow \quad \mathbb{E}[\hat{\theta} - \theta] = 0$$

- 推定量の期待値 $\mathbb{E}[\hat{\theta}]$ が, 真のパラメータ θ に一致する (偏りが無い).

不偏推定量 (Unbiased Estimate)

- 例えば, 平均 μ と共分散 Σ をもつ \mathbf{x} の確率分布 $p(\mathbf{x} \mid \mu, \Sigma)$ から, N 個の観測データが得られたとする:

$$\{\mathbf{x}_{1,\text{meas}}, \dots, \mathbf{x}_{N,\text{meas}}\} \leftarrow p(\mathbf{x} \mid \mu, \Sigma)$$
$$\mathbb{E}[\mathbf{x}_{\text{meas}}] = \mu, \quad \mathbb{E}\left[(\mathbf{x}_{\text{meas}} - \mu)(\mathbf{x}_{\text{meas}} - \mu)^\top\right] = \Sigma$$

- 観測データは, **互いに独立**であるとする.
- 観測データ $\{\mathbf{x}_{i,\text{meas}}\}$ を基に, 分布の平均 μ と共分散 Σ を推定したい.

不偏推定量 (Unbiased Estimate)

- 平均 μ と共分散 Σ をもつ \mathbf{x} の確率分布 $p(\mathbf{x} \mid \mu, \Sigma)$ から, N 個の観測データが得られたとする:

$$\{\mathbf{x}_{1,\text{meas}}, \dots, \mathbf{x}_{N,\text{meas}}\} \leftarrow p(\mathbf{x} \mid \mu, \Sigma)$$

- 観測データ $\{\mathbf{x}_{i,\text{meas}}\}$ を基に, 分布の平均 μ と共分散 Σ を推定したい.
- 平均と共分散の**不偏推定量**は, 次のようになる:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{i,\text{meas}}, \quad \hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \hat{\mu})(\mathbf{x}_{i,\text{meas}} - \hat{\mu})^\top$$

- 共分散については, 分母が N の代わりに $N-1$ となることに注意.
- $\hat{\mu}$ は, **標本平均**という.
- $\hat{\Sigma}$ は, **不偏分散**という (分母が N のときは, **標本分散**).

不偏推定量 (Unbiased Estimate)

- 不偏分散は、分母が N の代わりに $N - 1$ になる (少し大きく補正):

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \hat{\boldsymbol{\mu}}) (\mathbf{x}_{i,\text{meas}} - \hat{\boldsymbol{\mu}})^\top$$

- 共分散の計算に、真の平均 $\boldsymbol{\mu}$ ではなく、**観測データから計算された** 標本平均 $\hat{\boldsymbol{\mu}}$ を使っているから。
 - 観測データ $\mathbf{x}_{i,\text{meas}}$ と、標本平均 $\hat{\boldsymbol{\mu}}$ の間には相関がある。
 - 差分 $\mathbf{x}_{i,\text{meas}} - \hat{\boldsymbol{\mu}}$ は、真の平均から計算された、本当の差分 $\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu}$ よりは小さくなりがちである。
- $\hat{\boldsymbol{\mu}}$ が平均の不偏推定量であることは、次のように分かる:

$$\mathbb{E}[\hat{\boldsymbol{\mu}}] = \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N \mathbf{x}_{i,\text{meas}}\right] = \frac{1}{N} \sum_{i=1}^N \underbrace{\mathbb{E}[\mathbf{x}_{i,\text{meas}}]}_{=\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\mu} = \boldsymbol{\mu}$$

不偏推定量 (Unbiased Estimate)

- また, $\Sigma = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$ を用いて,

$$\begin{aligned}\mathbb{E}[\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top] &= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \mathbf{x}_{i,\text{meas}}\right)\left(\frac{1}{N}\sum_{j=1}^N \mathbf{x}_{j,\text{meas}}\right)^\top\right] \\&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}[\mathbf{x}_{i,\text{meas}}\mathbf{x}_{j,\text{meas}}^\top] \\&= \frac{1}{N^2}\sum_{i=1}^N \mathbb{E}[\mathbf{x}_{i,\text{meas}}\mathbf{x}_{i,\text{meas}}^\top] + \frac{1}{N^2}\sum_{i \neq j} \mathbb{E}[\mathbf{x}_{i,\text{meas}}\mathbf{x}_{j,\text{meas}}^\top] \\&= \frac{1}{N^2} \cdot N(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top) + \frac{1}{N^2} \cdot N(N-1)\boldsymbol{\mu}\boldsymbol{\mu}^\top = \frac{1}{N}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top\end{aligned}$$

- 観測データは互いに独立だから, 無相関であって, $i \neq j$ のとき

$$\mathbb{E}[\mathbf{x}_{i,\text{meas}}\mathbf{x}_{j,\text{meas}}^\top] = \mathbb{E}[\mathbf{x}_{i,\text{meas}}]\mathbb{E}[\mathbf{x}_{j,\text{meas}}]^\top = \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

不偏推定量 (Unbiased Estimate)

- $\hat{\mu}$ の共分散 $\mathbb{E} \left[(\hat{\mu} - \mathbb{E} [\hat{\mu}]) (\hat{\mu} - \mathbb{E} [\hat{\mu}])^\top \right]$ は,

$$\begin{aligned} \mathbb{E} \left[(\hat{\mu} - \mathbb{E} [\hat{\mu}]) (\hat{\mu} - \mathbb{E} [\hat{\mu}])^\top \right] &= \mathbb{E} \left[(\hat{\mu} - \mu) (\hat{\mu} - \mu)^\top \right] \\ &= \mathbb{E} [\hat{\mu} \hat{\mu}^\top] - \mathbb{E} [\hat{\mu}] \mu^\top - \mu \mathbb{E} [\hat{\mu}]^\top + \mu \mu^\top \\ &= \frac{1}{N} \Sigma + \mu \mu^\top - \mu \mu^\top - \mu \mu^\top + \mu \mu^\top \\ &= \frac{1}{N} \Sigma \end{aligned}$$

- $\mathbb{E} [\hat{\mu} \hat{\mu}^\top] = \frac{1}{N} \Sigma + \mu \mu^\top$, $\mathbb{E} [\hat{\mu}] = \mu$ を用いた.
- 観測データが増えれば ($N \rightarrow \infty$), 上記の共分散は 0 に近づく.
- 言い換えると, 平均の不偏推定量 $\hat{\mu}$ は, $\mathbb{E} [\hat{\mu}] = \mu$ に近づいてゆく.

不偏推定量 (Unbiased Estimate)

- $\hat{\Sigma}$ が共分散の不偏推定量であることも、次のように分かる:

$$\begin{aligned}\mathbb{E}[\hat{\Sigma}] &= \mathbb{E}\left[\frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \hat{\boldsymbol{\mu}})(\mathbf{x}_{i,\text{meas}} - \hat{\boldsymbol{\mu}})^\top\right] \\&= \mathbb{E}\left[\frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu} + \boldsymbol{\mu} - \hat{\boldsymbol{\mu}})(\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu} + \boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top\right] \\&= \frac{1}{N-1} \left(\sum_{i=1}^N \mathbb{E}\left[(\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})(\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})^\top\right] \right. \\&\quad \left. + \mathbb{E}\left[\sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top\right] \right. \\&\quad \left. + \mathbb{E}\left[\sum_{i=1}^N (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})(\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})^\top\right] + \sum_{i=1}^N \mathbb{E}\left[(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top\right] \right)\end{aligned}$$

不偏推定量 (Unbiased Estimate)

- 各項は,

$$\mathbb{E} \left[(\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu}) (\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})^\top \right] = \boldsymbol{\Sigma}$$

$$\mathbb{E} \left[(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \right] = \frac{1}{N} \boldsymbol{\Sigma}$$

$$\mathbb{E} \left[\sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \right] = \mathbb{E} \left[N (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \right] = -\boldsymbol{\Sigma}$$

$$\mathbb{E} \left[\sum_{i=1}^N (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (\mathbf{x}_{i,\text{meas}} - \boldsymbol{\mu})^\top \right] = \mathbb{E} \left[N (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top \right] = -\boldsymbol{\Sigma}$$

- $\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{i,\text{meas}}$ であるから, $\sum_{i=1}^N \mathbf{x}_{i,\text{meas}} = N \hat{\boldsymbol{\mu}}$.

- これらを, $\mathbb{E} \left[\hat{\boldsymbol{\Sigma}} \right]$ の式に代入する.

不偏推定量 (Unbiased Estimate)

- $\mathbb{E} [\hat{\Sigma}]$ を計算すると,

$$\begin{aligned}\mathbb{E} [\hat{\Sigma}] &= \frac{1}{N-1} \left(\sum_{i=1}^N \Sigma - \Sigma - \Sigma + \sum_{i=1}^N \frac{1}{N} \Sigma \right) \\ &= \frac{1}{N-1} (N\Sigma - \Sigma - \Sigma + \Sigma) = \frac{1}{N-1} (N-1)\Sigma = \Sigma\end{aligned}$$

- 以上より, 平均と共分散の**不偏推定量**は, 次のようになる:

$$\begin{aligned}\hat{\mu} &= \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{i,\text{meas}}, & \hat{\Sigma} &= \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,\text{meas}} - \hat{\mu})(\mathbf{x}_{i,\text{meas}} - \hat{\mu})^\top \\ \mathbb{E} [\hat{\mu}] &= \mu, & \mathbb{E} [\hat{\Sigma}] &= \Sigma \\ \mathbb{E} [\hat{\mu}\hat{\mu}^\top] &= \frac{1}{N}\Sigma + \mu\mu^\top, & \mathbb{E} [(\hat{\mu} - \mu)(\hat{\mu} - \mu)^\top] &= \frac{1}{N}\Sigma\end{aligned}$$

クラメール・ラオの下限 (Cramér–Rao Lower Bound)

- パラメータ θ をもつ \mathbf{x} の確率分布 $p(\mathbf{x} | \theta)$ を考える.
- この確率分布から, 観測データが得られたとする.

$$\mathbf{x}_{\text{meas}} \leftarrow p(\mathbf{x} | \theta)$$

- 観測データ \mathbf{x}_{meas} を基に, パラメータ θ の不偏推定量 $\hat{\theta}$ を計算する.

$$\mathbb{E} [\hat{\theta} - \theta] = 0$$

- 不偏推定量 $\hat{\theta}$ の共分散について, 以下が成り立つ:

$$\mathbb{E} \left[(\hat{\theta} - \theta) (\hat{\theta} - \theta)^{\top} \right] \geq \mathbf{I}^{-1}(\mathbf{x} | \theta)$$

- これを, **クラメール・ラオの下限**という.

クラメル・ラオの下限 (Cramér–Rao Lower Bound)

- 確率分布 $p(\mathbf{x} \mid \boldsymbol{\theta})$ のパラメータ $\boldsymbol{\theta}$ の不偏推定量 $\hat{\boldsymbol{\theta}}$ について,

$$\mathbb{E} \left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^\top \right] \geq \mathbf{I}^{-1}(\mathbf{x} \mid \boldsymbol{\theta})$$

- $\mathbf{I}(\mathbf{x} \mid \boldsymbol{\theta})$ は, **フィッシャー情報行列**とよぶ.

$$\mathbf{I}(\mathbf{x} \mid \boldsymbol{\theta}) = \mathbb{E} \left[\left(\frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \left(\frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right]$$

- 不等号 $\mathbf{A} \geq \mathbf{B}$ は, 行列 $\mathbf{A} - \mathbf{B}$ が半正定値になることを意味する.

$$\forall \mathbf{x} \neq \mathbf{0} \quad \mathbf{x}^\top (\mathbf{A} - \mathbf{B}) \mathbf{x} \geq 0 \quad \longrightarrow \quad \mathbf{A} - \mathbf{B} \geq 0$$

- 観測データを使ってパラメータを推定するとき, その精度には限度がある.

目次

- ① 概要
- ② 確率分布
- ③ ガウス積分

微分のレイアウト (再掲)

- 以下の2つのレイアウトに大別される (x, y を縦ベクトルとする).
- **分子レイアウト** (Numerator Layout)
 - $\frac{\partial y}{\partial x}$ は, 縦ベクトル, $\frac{\partial y}{\partial \mathbf{x}}$ は, 横ベクトル
- **分母レイアウト** (Denominator Layout)
 - $\frac{\partial y}{\partial x}$ は, 横ベクトル, $\frac{\partial y}{\partial \mathbf{x}}$ は, 縦ベクトル

スカラによる微分 (再掲)

分子レイアウト

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{pmatrix}$$

分母レイアウト

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{pmatrix} \equiv \frac{\partial \mathbf{y}^\top}{\partial x}$$

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{pmatrix}$$

まれ

ベクトルによる微分 (再掲)

分子レイアウト

$$\frac{\partial y}{\partial \mathbf{x}} = \left(\frac{\partial y}{\partial x_1} \quad \cdots \quad \frac{\partial y}{\partial x_n} \right) \equiv \frac{\partial y}{\partial \mathbf{x}^\top}$$

分母レイアウト

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

$$\equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

$$\equiv \frac{\partial \mathbf{y}^\top}{\partial \mathbf{x}}$$

行列による微分 (再掲)

分子レイアウト

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1n}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{pmatrix}$$
$$\equiv \frac{\partial y}{\partial \mathbf{X}^\top}$$

分母レイアウト

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{pmatrix}$$
$$\equiv \frac{\partial y}{\partial \mathbf{X}}$$

ヤコビ行列, ヤコビアン (Jacobi Matrix, Jacobian)

- n, m 次変数 $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$, $\mathbf{y} = (y_1, y_2, \dots, y_m)^\top$ を考える.
- $\frac{\partial y_i}{\partial x_j}$ を (i, j) 成分とした $m \times n$ 行列 \mathbf{J} を, ヤコビ行列という.

$$\mathbf{J} = \frac{\partial(y_1, y_2, \dots, y_m)}{\partial(x_1, x_2, \dots, x_n)} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

- 分子レイアウトにおける $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top}$ と同じ.
- ヤコビ行列の行列式 $\det \mathbf{J}$ を, ヤコビアンとよぶ.

積分の変数変換

- n, m 次元変数 $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$, $\mathbf{y} = (y_1, y_2, \dots, y_m)^\top$ を考える.
- 関数 $f(\mathbf{y})$ があり, 次の積分を行いたいとする:

$$\int f(\mathbf{y}) \, d\mathbf{y}$$

- $\mathbf{y} = \mathbf{g}(\mathbf{x})$, $m \times n$ 行列を $\mathbf{G} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ とおく.
- ヤコビアン $\det \mathbf{G}$ を使うと, 次のように変数変換できる:

$$\int f(\mathbf{y}) \, d\mathbf{y} \quad \longrightarrow \quad \int f(\mathbf{g}(\mathbf{x})) |\det \mathbf{G}| \, d\mathbf{x}$$

- $d\mathbf{y} = |\det \mathbf{G}| \, d\mathbf{x}$ のように関連付けられる.
- ヤコビアンは, 変数変換による微小な体積 $d\mathbf{x}$, $d\mathbf{y}$ の変化率を表す.

積分の変数変換 (スカラーの場合)

- 関数 $f(y)$ があり, 次の計算を行いたいとする:

$$\int f(y) \, dy$$

- $y = g(x)$ とすると, 次のように変数変換できる:

$$\int f(y) \, dy \quad \longrightarrow \quad \int f(g(x)) \frac{dy}{dx} \, dx$$

偶関数の積分

- 関数 $f(x)$ が偶関数であれば, $f(-x) = f(x)$ となる.
- $-a$ から a までの積分は,

$$\int_{-a}^a f(x) dx = \int_{-a}^0 f(x) dx + \int_0^a f(x) dx$$

- 第一項において $x = -y$ とすると, $\frac{dx}{dy} = -1$, 積分範囲は $[-a, 0]$ から $[a, 0]$ になるので,

$$\begin{aligned}\int_{-a}^a f(x) dx &= \int_a^0 \textcolor{red}{f}(-y) \frac{dx}{dy} dy + \int_0^a f(x) dx \\ &= \int_a^0 \textcolor{red}{f}(y)(-1) dy + \int_0^a f(x) dx \\ &= \int_0^a f(y) dy + \int_0^a f(x) dx = 2 \int_0^a f(x) dx\end{aligned}$$

奇関数の積分

- 関数 $f(x)$ が奇関数であれば, $f(-x) = -f(x)$ となる.
- $-a$ から a までの積分は,

$$\int_{-a}^a f(x) dx = \int_{-a}^0 f(x) dx + \int_0^a f(x) dx$$

- 第一項において $x = -y$ とすると, $\frac{dx}{dy} = -1$, 積分範囲は $[-a, 0]$ から $[a, 0]$ になるので,

$$\begin{aligned}\int_{-a}^a f(x) dx &= \int_a^0 \textcolor{red}{f}(-y) \frac{dx}{dy} dy + \int_0^a f(x) dx \\ &= \int_a^0 \textcolor{red}{-f}(y) (-1) dy + \int_0^a f(x) dx \\ &= - \int_0^a f(y) dy + \int_0^a f(x) dx = 0\end{aligned}$$

偶関数, 奇関数の積分

偶関数, 奇関数の積分

$f(x)$ が偶関数であれば, $[-a, a]$ の積分は, $[0, a]$ の積分の 2 倍:

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx$$

$f(x)$ が奇関数であれば, $[-a, a]$ の積分は 0:

$$\int_{-a}^a f(x) dx = 0$$

- 以下も成り立つ:
- (奇関数) \times (奇関数) = (偶関数)
- (奇関数) \times (偶関数) = (奇関数)
- (偶関数) \times (偶関数) = (偶関数)

偶関数, 奇関数の積分 (多変数関数)

- n 次変数 $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ を考える.
- $\mathbf{a} = (a_1, a_2, \dots, a_n)^\top$ として, $-\mathbf{a}$ から \mathbf{a} までの $f(\mathbf{x})$ の積分は,

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x}) \, d\mathbf{x} = \int_{-a_1}^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \cdots dx_n$$

x_1 に関する積分を, $[-a_1, 0]$, $[0, a_1]$ の 2 つの範囲に分割すると,

$$\begin{aligned} \int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x}) \, d\mathbf{x} &= \int_{-a_1}^0 \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \cdots dx_n \\ &\quad + \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \cdots dx_n \end{aligned}$$

偶関数, 奇関数の積分 (多変数関数)

- 第一項について $x_i = -y_i$ とすると, $\frac{dx_i}{dy_i} = -1$ であり, 積分範囲の正負も入れ替わるから,

$$\begin{aligned} & \int_{-a_1}^0 \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \\ &= \int_{a_1}^0 \int_{a_2}^{-a_2} \cdots \int_{a_n}^{-a_n} f(-y_1, -y_2, \dots, -y_n) (-1)^n dy_1 dy_2 \cdots dy_n \\ &= \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(-y_1, -y_2, \dots, -y_n) dy_1 dy_2 \cdots dy_n \\ &= \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(-\mathbf{y}) d\mathbf{y} \end{aligned}$$

- 積分範囲を交換すると, 符号が反転する.
- 上記の例では, n 個の積分範囲を交換したので, $(-1)^n$ が消える.

偶関数, 奇関数の積分 (多変数関数)

- この結果を使って, 第一項を置き換えれば,

$$\begin{aligned}\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x}) \, d\mathbf{x} &= \int_{-a_1}^0 \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \cdots dx_n \\ &\quad + \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \cdots dx_n \\ &= \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(-\mathbf{y}) \, d\mathbf{y} + \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(\mathbf{x}) \, d\mathbf{x}\end{aligned}$$

- 偶関数なら $f(-\mathbf{y}) = f(\mathbf{y})$, 奇関数なら $f(-\mathbf{y}) = -f(\mathbf{y})$ だから,

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x}) \, d\mathbf{x} = \begin{cases} 2 \int_0^{a_1} \int_{-a_2}^{a_2} \cdots \int_{-a_n}^{a_n} f(\mathbf{x}) \, d\mathbf{x} & (f(\mathbf{x}) \text{ が偶関数}) \\ 0 & (f(\mathbf{x}) \text{ が奇関数}) \end{cases}$$

ガウス積分

ガウス積分の基本形

$$\int_0^{\infty} \exp(-x^2) dx = \frac{\sqrt{\pi}}{2}$$

上記の積分を I とおくと, I^2 は

$$I^2 = \int_0^{\infty} \int_0^{\infty} \exp(-(x^2 + y^2)) dx dy$$

$x = r \cos \theta$, $y = r \sin \theta$ のように変数変換すると, ヤコビアンは

$$\begin{aligned} \left| \det \frac{\partial(x, y)}{\partial(r, \theta)} \right| &= \left| \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \right| \\ &= |r (\cos^2 \theta + \sin^2 \theta)| = |r| \end{aligned}$$

ガウス積分

$x = r \cos \theta$, $y = r \sin \theta$, $x, y \in [0, \infty]$ であるから, $r \in [0, \infty)$, $\theta \in [0, \frac{\pi}{2}]$.

ヤコビアンは $|r| = r$ であるから, $dx dy = r dr d\theta$.

以上より, I^2 は

$$\begin{aligned} I^2 &= \int_0^\infty \int_0^\infty \exp(-(x^2 + y^2)) \, dx \, dy \\ &= \int_0^\infty \int_0^{\frac{\pi}{2}} \exp(-r^2) \, r \, dr \, d\theta = \int_0^{\frac{\pi}{2}} d\theta \int_0^\infty \exp(-r^2) \, r \, dr \\ &= \frac{\pi}{2} \left[-\frac{1}{2} \exp(-r^2) \right]_0^\infty = \frac{\pi}{4} \end{aligned}$$

であるから, $I = \frac{\sqrt{\pi}}{2}$.

ガウス積分

ガウス積分の基本形

$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}$$

上記の積分を I とおくと, I^2 は

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-(x^2 + y^2)) dx dy$$

$x = r \cos \theta$, $y = r \sin \theta$ のように変数変換すると, ヤコビアンは r . よって, $dx dy = r dr d\theta$. $x, y \in (-\infty, \infty)$ であるから, $r \in [0, \infty]$, $\theta \in [0, 2\pi]$.

$$\begin{aligned} I^2 &= \int_0^{\infty} \int_0^{2\pi} \exp(-r^2) r dr d\theta = \int_0^{2\pi} d\theta \int_0^{\infty} \exp(-r^2) r dr \\ &= 2\pi \left[-\frac{1}{2} \exp(-r^2) \right]_0^{\infty} = \pi \quad \longrightarrow \quad I = \sqrt{\pi} \end{aligned}$$

ガウス積分

ガウス積分の基本形

$$\int_0^{\infty} \exp(-x^2) dx = \frac{\sqrt{\pi}}{2}$$
$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}$$

$\exp(-x^2)$ は偶関数 ($f(x) = f(-x)$) であるから, $-\infty$ から ∞ までの積分値は, 0 から ∞ までの積分値の倍となる.

ガウス積分

ガウス積分の基本形

$$\int_0^{\infty} \exp(-ax^2) dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} \quad (a > 0)$$

$$\int_{-\infty}^{\infty} \exp(-ax^2) dx = \sqrt{\frac{\pi}{a}} \quad (a > 0)$$

$y = \sqrt{ax}$ とすると, $x = \frac{y}{\sqrt{a}}, \frac{dx}{dy} = \frac{1}{\sqrt{a}}$ だから,

$$\int_0^{\infty} \exp(-ax^2) dx = \int_0^{\infty} \exp(-y^2) \frac{dx}{dy} dy = \frac{1}{\sqrt{a}} \int_0^{\infty} \exp(-y^2) dy = \frac{1}{\sqrt{a}} \frac{\sqrt{\pi}}{2}$$

$\exp(-ax^2)$ は偶関数だから, 積分範囲を $[0, \infty)$ から $(-\infty, \infty)$ に広げると, 積分値は倍となる.

ガウス積分

ガウス積分の基本形

$$\int_0^{\infty} x \exp(-ax^2) dx = \frac{1}{2a} \quad (a > 0)$$
$$\int_{-\infty}^{\infty} x \exp(-ax^2) dx = 0 \quad (a > 0)$$

次のように計算できる:

$$\int_0^{\infty} x \exp(-ax^2) dx = \left[-\frac{1}{2a} \exp(-ax^2) \right]_0^{\infty} = \frac{1}{2a}$$

$x \exp(-ax^2)$ は奇関数だから, $-\infty$ から ∞ までの積分は 0.

ガウス積分の漸化式

$$I_n = \int_0^{\infty} x^n \exp(-ax^2) dx \text{ とおくと } (a > 0),$$

$$I_{n+2} = \frac{n+1}{2a} I_n$$

次のように部分積分すれば ($\lim_{x \rightarrow \infty} x^n \exp(-x) = 0$ に注意):

$$\begin{aligned} I_n &= \int_0^{\infty} x^n \exp(-ax^2) dx \\ &= \left[\frac{x^{n+1}}{n+1} \exp(-ax^2) \right]_0^{\infty} + \frac{2a}{n+1} \int_0^{\infty} x^{n+2} \exp(-ax^2) dx \\ &= \frac{2a}{n+1} \int_0^{\infty} x^{n+2} \exp(-ax^2) dx = \frac{2a}{n+1} I_{n+2} \end{aligned}$$

ガウス積分

ガウス積分の漸化式

$$I_n = \int_0^{\infty} x^n \exp(-ax^2) dx \text{ とおくと } (a > 0),$$

$$I_{n+2} = -\frac{\partial}{\partial a} I_n$$

次のように計算できる:

$$\begin{aligned} \frac{\partial}{\partial a} I_n &= \frac{\partial}{\partial a} \int_0^{\infty} x^n \exp(-ax^2) dx \\ &= \int_0^{\infty} \frac{\partial}{\partial a} x^n \exp(-ax^2) dx \\ &= -\int_0^{\infty} x^{n+2} \exp(-ax^2) dx = -I_{n+2} \end{aligned}$$

微分と積分の交換については条件がある (上の場合は可能).

ガウス積分

ガウス積分 (2 乗)

$$\int_0^{\infty} x^2 \exp(-ax^2) dx = \frac{1}{4} \sqrt{\frac{\pi}{a^3}} \quad (a > 0)$$

$$\int_{-\infty}^{\infty} x^2 \exp(-ax^2) dx = \frac{1}{2} \sqrt{\frac{\pi}{a^3}} \quad (a > 0)$$

$$I_n = \int_0^{\infty} x^n \exp(-ax^2) dx \text{ とすると, } I_{n+2} = \frac{n+1}{2a} I_n \text{ である.}$$

$$I_0 = \int_0^{\infty} \exp(-ax^2) dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} \text{ だから,}$$

$$I_2 = \int_0^{\infty} x^2 \exp(-ax^2) dx = \frac{0+1}{2a} I_0 = \frac{1}{2a} \frac{1}{2} \sqrt{\frac{\pi}{a}} = \frac{1}{4} \sqrt{\frac{\pi}{a^3}}$$

$x^2 \exp(-ax^2)$ は偶関数なので, $(-\infty, \infty)$ の積分は, $[0, \infty)$ の積分の 2 倍.

ガウス積分

ガウス積分 (3 乗)

$$\int_0^{\infty} x^3 \exp(-ax^2) dx = \frac{1}{2a^2} \quad (a > 0)$$

$$\int_{-\infty}^{\infty} x^3 \exp(-ax^2) dx = 0 \quad (a > 0)$$

$$I_n = \int_0^{\infty} x^n \exp(-ax^2) dx \text{ とすると, } I_{n+2} = \frac{n+1}{2a} I_n \text{ である.}$$

$$I_1 = \int_0^{\infty} x \exp(-ax^2) dx = \frac{1}{2a} \text{ だから,}$$

$$I_3 = \int_0^{\infty} x^3 \exp(-ax^2) dx = \frac{1+1}{2a} I_1 = \frac{1}{a} \frac{1}{2a} = \frac{1}{2a^2}$$

$x^3 \exp(-ax^2)$ は奇関数だから, $-\infty$ から ∞ までの積分は 0.

ガウス積分

ガウス積分 (二次関数の一般形)

$$\int_{-\infty}^{\infty} \exp(-ax^2 + bx + c) dx = \exp\left(\frac{b^2}{4a} + c\right) \sqrt{\frac{\pi}{a}}$$

$-ax^2 + bx + c = -a\left(x - \frac{b}{2a}\right)^2 + \frac{b^2}{4a} + c$ と平方完成できる。

$y = x - \frac{b}{2a}$ とすると, $\frac{\partial x}{\partial y} = 1$, 積分範囲は $(-\infty, \infty)$ だから,

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(-ax^2 + bx + c) dx &= \exp\left(\frac{b^2}{4a} + c\right) \int_{-\infty}^{\infty} \exp(-ay^2) \frac{\partial x}{\partial y} dy \\ &= \exp\left(\frac{b^2}{4a} + c\right) \int_{-\infty}^{\infty} \exp(-ay^2) dy = \exp\left(\frac{b^2}{4a} + c\right) \sqrt{\frac{\pi}{a}} \end{aligned}$$