# 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

August 14, 2023

# 目次

# 目次

# 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

# 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:

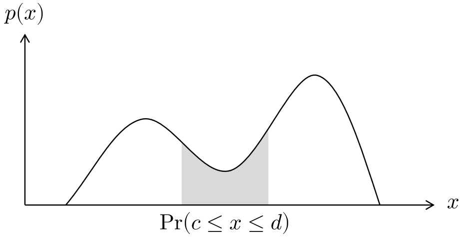
$$p(x) \geq 0, \quad \int_a^b p(x)\, \mathrm{d}x = 1$$

# 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.

$$\Pr(c \leq x \leq d) = \int_c^d p(x)\,\mathrm{d}x$$

- $\Pr(c \leq x \leq d)$ 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 $p(x \mid y)$ を考える.
- $p(x \mid y)$ を, **条件付き確率密度関数**という.
- $p(x \mid y)$ も, 2 つの条件を満たす:

$$p(x \mid y) \geq 0, \quad \int_a^b p(x \mid y)\,\mathrm{d}x = 1$$

# 同時分布 (Joint Probability)

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

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

- 積分の範囲は $\mathbf{a} = (a_1, a_2, \ldots, a_N)$, $\mathbf{b} = (b_1, b_2, \ldots, b_N)$ 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.

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

- 同時分布 $p(\mathbf{x}, \mathbf{y})$, 条件付き確率分布 $p(\mathbf{x} \mid \mathbf{y})$, $p(\mathbf{y} \mid \mathbf{x})$, 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.

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

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

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

- $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ であれば, $\mathbf{x}, \mathbf{y}$ 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:

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

- 分母は周辺化としてかける.

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

- ベイズの定理は, 次のようになる.

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

# ベイズの定理 (Bayes' Theorem)

## ベイズの定理

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

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:

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

- 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 $\mathbf{y}_1$, $\mathbf{y}_2$ から, 変数 $\mathbf{x}$ の事後分布 $p(\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2)$ を推定したい.

$$p(\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y}_1, \mathbf{y}_2)} \quad (\because \text{ベイズの定理})$$

- $\mathbf{x}$ のもとで $\mathbf{y}_1, \mathbf{y}_2$ が互いに独立なら, 以下が成り立つ:

$$p(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{x}) = p(\mathbf{y}_1 \mid \mathbf{x}) p(\mathbf{y}_2 \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{y}_1) p(\mathbf{y}_1)}{p(\mathbf{x})} \frac{p(\mathbf{x} \mid \mathbf{y}_2) p(\mathbf{y}_2)}{p(\mathbf{x})}$$

- 上記を代入すれば, 次が得られる.

$$p(\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2) = \eta \; p(\mathbf{x} \mid \mathbf{y}_1) p(\mathbf{x} \mid \mathbf{y}_2)$$

- $\eta = \dfrac{p(\mathbf{y}_1) p(\mathbf{y}_2)}{p(\mathbf{y}_1, \mathbf{y}_2) p(\mathbf{x})}$ 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.
- $\mathbb{E}\left[f(x)\right]$ を，関数 $f(x)$ の，確率分布 $p(x)$ のもとでの**期待値**とする.

$$\mathbb{E}\left[f(x)\right] = \int f(x)p(x)\,\mathrm{d}x$$

- 1 次モーメント: **平均** $\mu = \mathbb{E}\left[x\right]$ (Mean)
- 2 次モーメント: **分散** $\sigma^2 = \mathbb{E}\left[(x-\mu)^2\right]$ (Variance)
- 3 次モーメント: **歪度** $\mathbb{E}\left[(x-\mu)^3\right]$ (Skewness; わいど)
- 4 次モーメント: **尖度** $\mathbb{E}\left[(x-\mu)^4\right]$ (Kurtosis; せんど)

$$\mu = \mathbb{E}\left[x\right] = \int x p(x)\,\mathrm{d}x$$

$$\sigma^2 = \mathbb{E}\left[(x-\mu)^2\right] = \int (x-\mu)^2 p(x)\,\mathrm{d}x$$

# モーメント, 平均, 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:

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

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

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

- 平均 $\boldsymbol{\mu}$ と分散 $\boldsymbol{\Sigma}$ は, 次のようにかける.

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

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

# モーメント, 平均, 分散 (多変数の場合)

- 確率分布 $p(\mathbf{x})$ の分散 $\boldsymbol{\Sigma}$ について, 次が成り立つ.

$$\boldsymbol{\Sigma} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{x}]^\top = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

- スカラの場合は, 次のようになる.

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

- 以下のように示せる.

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

# 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:

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

- 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, 次のように示せる.

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}\mathbf{y}^\top\right] &= \iint \mathbf{x}\mathbf{y}^\top p(\mathbf{x}, \mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} = \iint \mathbf{x}\mathbf{y}^\top p(\mathbf{x})p(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \quad (\because \text{独立}) \\ &= \int \mathbf{x}p(\mathbf{x})\,\mathrm{d}\mathbf{x} \int \mathbf{y}^\top p(\mathbf{y})\,\mathrm{d}\mathbf{y} = \mathbb{E}\left[\mathbf{x}\right]\mathbb{E}\left[\mathbf{y}\right]^\top\end{aligned}$$

# 共分散 (Covariance)

- $\mathbf{x}, \mathbf{y}$ について，共分散は次のようにかける.

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

- 左辺を展開すれば，次のように示せる.

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

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

# イェンセンの不等式 (Jensen's Inequality)

- 上に凸な関数 $f(\mathbf{x})$ を考える.
- $p(\mathbf{x})$ を, $\int p(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ 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:

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

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

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

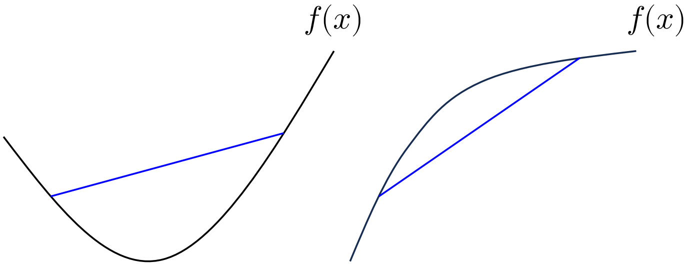
- $f(\mathbf{x})$ が下に凸であれば, 次が成り立つ (不等号が逆になる):

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

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

# 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).
  - $x^2$, $x\ln x$, $\dfrac{1}{x}$ $(x>0)$, $\tan x$ $(0<x<\dfrac{\pi}{2})$ 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).
  - $\ln x$, $\tan x$ $(-\dfrac{\pi}{2}<x<0)$ など



# エントロピー (Entropy)

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

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

- $\mathrm{H}[\mathbf{x}]$ 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 $\mathrm{H}[\mathbf{x}, \mathbf{y}]$ を考える.

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

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

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

# 条件付きエントロピー (Conditional Entropy)

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

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

- $\mathrm{H}[\mathbf{y} \mid \mathbf{x}]$ を, 条件付きエントロピーとよぶ.
- 次が成り立つ:

$$\mathrm{H}[\mathbf{x}, \mathbf{y}] = \mathrm{H}[\mathbf{y} \mid \mathbf{x}] + \mathrm{H}[\mathbf{x}] = \mathrm{H}[\mathbf{x} \mid \mathbf{y}] + \mathrm{H}[\mathbf{y}]$$

- $\mathrm{H}[\mathbf{x}]$: $\mathbf{x}$ の予測しづらさ
- $\mathrm{H}[\mathbf{x}, \mathbf{y}]$: $\mathbf{x}$ と $\mathbf{y}$ の予測しづらさ
- $\mathrm{H}[\mathbf{y} \mid \mathbf{x}]$: $\mathbf{x}$ が既知であるときの, $\mathbf{y}$ 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 $\mathrm{KL}\left(p \parallel q\right)$ を考える.

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

- $\mathrm{KL}\left(p \parallel q\right)$ 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.
- $\mathrm{KL}\left(p \parallel q\right) \neq \mathrm{KL}\left(q \parallel p\right)$ 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:

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

- $\ln$ は下に凸だから，イェンセンの不等式より $(\mathbb{E}\left[f(x)\right] \leq f(\mathbb{E}\left[x\right]))$

$$\begin{aligned} \mathrm{KL}\left(p \parallel q\right) &= -\mathbb{E}\left[\ln \frac{q(\mathbf{x})}{p(\mathbf{x})}\right] \geq -\ln \mathbb{E}\left[\frac{q(\mathbf{x})}{p(\mathbf{x})}\right] \\ &= -\ln \int p(\mathbf{x}) \frac{q(\mathbf{x})}{p(\mathbf{x})}\, \mathrm{d}\mathbf{x} = -\ln \int q(\mathbf{x})\, \mathrm{d}\mathbf{x} = -\ln 1 = 0 \end{aligned}$$

- 以上より，$\mathrm{KL}\left(p \parallel q\right) \geq 0$.
- $\forall \mathbf{x}\ p(\mathbf{x}) = q(\mathbf{x})$ のときのみ $0$.

# 相互情報量 (Mutual Information)

- 確率変数 $\mathbf{x}, \mathbf{y}$ について，次の量 $I(\mathbf{x}, \mathbf{y})$ を考える．

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

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

$$I(\mathbf{x}, \mathbf{y}) = \iint p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} = \iint p(\mathbf{x}, \mathbf{y}) \ln 1 \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} = 0$$

- $\mathbf{x}, \mathbf{y}$ が独立であれば，$\mathbf{x}$ $(\mathbf{y})$ のことが分かっても，$\mathbf{y}$ $(\mathbf{x})$ 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 $I(\mathbf{x}, \mathbf{y})$:

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

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

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

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

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

- $\mathrm{H}[\mathbf{y}]$: $\mathbf{y}$ の予測しづらさ
- $\mathrm{H}[\mathbf{y} \mid \mathbf{x}]$: $\mathbf{x}$ 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 $I(\mathbf{x}, \mathbf{y})$:

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

- カルバック–ライブラーダイバージェンスを用いると,

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

- KL ダイバージェンスの性質 ($\geq 0$) 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 $\boldsymbol{\theta}$ をもつ $\mathbf{x}$ の確率分布 $p(\mathbf{x} \mid \boldsymbol{\theta})$ 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.

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

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

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} \quad \longrightarrow \quad \mathbb{E}\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \mathbf{0}$$

- 推定量の期待値 $\mathbb{E}\left[\hat{\boldsymbol{\theta}}\right]$ が, パラメータ $\boldsymbol{\theta}$ に一致する (偏りがない).

# 不偏推定量 (Unbiased Estimate)

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

$$\{\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{N,\mathrm{meas}}\} \leftarrow p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

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

# 不偏推定量 (Unbiased Estimate)

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

$$\{\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{N,\mathrm{meas}}\} \leftarrow p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- 観測データ $\{\mathbf{x}_{i,\mathrm{meas}}\}$ を基に, 分布の平均 $\boldsymbol{\mu}$ と共分散 $\boldsymbol{\Sigma}$ 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:

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

- 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.
- $\hat{\boldsymbol{\mu}}$ は, 標本平均という.
- $\hat{\boldsymbol{\Sigma}}$ 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):

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

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

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

# 不偏推定量 (Unbiased Estimate)

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

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

- 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

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

# 不偏推定量 (Unbiased Estimate)

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

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

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

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

# 不偏推定量 (Unbiased Estimate)

- 各項は,

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

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

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

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

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

- $\mathbb{E}\left[\hat{\boldsymbol{\Sigma}}\right]$ を計算すると,

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

- 以上より, 平均と共分散の不偏推定量は, 次のようになる:

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

# クラメール・ラオの下限 (Cramér–Rao Lower Bound)

- パラメータ $\boldsymbol{\theta}$ をもつ $\mathbf{x}$ の確率分布 $p(\mathbf{x} \mid \boldsymbol{\theta})$ 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.

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

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

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \mathbf{0}$$

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

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

- これを, クラメール・ラオの下限という.

# クラメール・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

# 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 ($\mathbf{x}, \mathbf{y}$ を縦ベクトルとする).
- 分子レイアウト (Numerator Layout)
  - $\dfrac{\partial \mathbf{y}}{\partial x}$ は, 縦ベクトル, $\dfrac{\partial y}{\partial \mathbf{x}}$ は, 横ベクトル
- 分母レイアウト (Denominator Layout)
  - $\dfrac{\partial \mathbf{y}}{\partial x}$ は, 横ベクトル, $\dfrac{\partial y}{\partial \mathbf{x}}$ は, 縦ベクトル

# スカラによる微分 (再掲)

| 分子レイアウト | 分母レイアウト |
|---|---|
| $\dfrac{\partial y}{\partial x}$ | $\dfrac{\partial y}{\partial x}$ |
| $\dfrac{\partial \mathbf{y}}{\partial x} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x} \\ \vdots \\ \dfrac{\partial y_m}{\partial x} \end{pmatrix}$ | $\dfrac{\partial \mathbf{y}}{\partial x} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x} & \cdots & \dfrac{\partial y_m}{\partial x} \end{pmatrix} \equiv \dfrac{\partial \mathbf{y}}{\partial x}^{\top}$ |
| $\dfrac{\partial \mathbf{Y}}{\partial x} = \begin{pmatrix} \dfrac{\partial y_{11}}{\partial x} & \cdots & \dfrac{\partial y_{1n}}{\partial x} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_{m1}}{\partial x} & \cdots & \dfrac{\partial y_{mn}}{\partial x} \end{pmatrix}$ | まれ |

# ベクトルによる微分 (再掲)

| 分子レイアウト | 分母レイアウト |
|---|---|
| $\frac{\partial y}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} & \cdots & \frac{\partial y}{\partial x_n} \end{pmatrix} \equiv \frac{\partial y}{\partial \mathbf{x}^\top}$ | $\frac{\partial y}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$ |
| $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top}$ | $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \equiv \frac{\partial \mathbf{y}^\top}{\partial \mathbf{x}}$ |

# 行列による微分 (再掲)

| 分子レイアウト | 分母レイアウト |
| --- | --- |
| $\dfrac{\partial y}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial y}{\partial x_{11}} & \cdots & \dfrac{\partial y}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y}{\partial x_{1n}} & \cdots & \dfrac{\partial y}{\partial x_{mn}} \end{pmatrix} \equiv \dfrac{\partial y}{\partial \mathbf{X}^\top}$ | $\dfrac{\partial y}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial y}{\partial x_{11}} & \cdots & \dfrac{\partial y}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y}{\partial x_{m1}} & \cdots & \dfrac{\partial y}{\partial x_{mn}} \end{pmatrix} \equiv \dfrac{\partial y}{\partial \mathbf{X}}$ |

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

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

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

- 分子レイアウトにおける $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}} \equiv \dfrac{\partial \mathbf{y}}{\partial \mathbf{x}^\top}$ 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

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

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

- $\mathbf{y} = \mathbf{g}(\mathbf{x})$, $m \times n$ 行列を $\mathbf{G} = \dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$ 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:

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

- $\mathrm{d}\mathbf{y} = |\det \mathbf{G}|\,\mathrm{d}\mathbf{x}$ 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 $\mathrm{d}\mathbf{x}$, $\mathrm{d}\mathbf{y}$ の変化率を表す.

# 積分の変数変換 (スカラの場合)

- 関数 $f(y)$ があり, 次の計算を行いたいとする:

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

- $y = g(x)$ とすると, 次のように変数変換できる:

$$\int f(y)\,\mathrm{d}y \quad \longrightarrow \quad \int f(g(x))\frac{\mathrm{d}y}{\mathrm{d}x}\,\mathrm{d}x$$

# 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，

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = \int_{-a}^{0} f(x)\,\mathrm{d}x + \int_{0}^{a} f(x)\,\mathrm{d}x$$

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

$$\begin{aligned}\int_{-a}^{a} f(x)\,\mathrm{d}x &= \int_{a}^{0} f(-y)\frac{\mathrm{d}x}{\mathrm{d}y}\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x \\ &= \int_{a}^{0} f(y)(-1)\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x \\ &= \int_{0}^{a} f(y)\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x = 2\int_{0}^{a} f(x)\,\mathrm{d}x\end{aligned}$$

# 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，

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = \int_{-a}^{0} f(x)\,\mathrm{d}x + \int_{0}^{a} f(x)\,\mathrm{d}x$$

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

$$\begin{aligned}\int_{-a}^{a} f(x)\,\mathrm{d}x &= \int_{a}^{0} f(-y)\frac{\mathrm{d}x}{\mathrm{d}y}\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x \\ &= \int_{a}^{0} -f(y)(-1)\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x \\ &= \int_{a}^{0} (-f(y))\,\mathrm{d}y + \int_{0}^{a} f(x)\,\mathrm{d}x = 0\end{aligned}$$

# 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:

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = 2\int_{0}^{a} f(x)\,\mathrm{d}x$$

$f(x)$ が奇関数であれば，$[-a, a]$ の積分は 0:

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = 0$$

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

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{-\mathbf{a}}^{\mathbf{0}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

- 上式の積分は, 次と同じことである:

$$\begin{aligned}
&\int_{-a_1}^{a_1}\int_{-a_2}^{a_2}\cdots\int_{-a_n}^{a_n} f(x_1, x_2, \ldots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n \\
&= \int_{-a_1}^{0}\int_{-a_2}^{0}\cdots\int_{-a_n}^{0} f(x_1, x_2, \ldots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n \\
&\quad + \int_{0}^{a_1}\int_{0}^{a_2}\cdots\int_{0}^{a_n} f(x_1, x_2, \ldots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n
\end{aligned}$$

# 偶関数，奇関数の積分 (多変数関数)

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

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{-\mathbf{a}}^{\mathbf{0}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

- 第一項において $\mathbf{x} = -\mathbf{y}$ とすると，ヤコビ行列は $\dfrac{\partial \mathbf{x}}{\partial \mathbf{y}} = -\mathbf{I}$，ヤコビアンは $\det(-\mathbf{I}) = (-1)^n$．積分範囲は $[-\mathbf{a}, \mathbf{0}]$ から $[\mathbf{a}, \mathbf{0}]$ になるので，

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathbf{a}}^{\mathbf{0}} f(-\mathbf{y})(-1)^n\,\mathrm{d}\mathbf{y} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

- $\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{I}$，$\det(c\mathbf{A}) = c^n \det \mathbf{A}$ を用いた．

# 偶関数, 奇関数の積分 (多変数関数)

- $\mathbf{a} = (a_1, a_2, \ldots, a_n)^\top$ として, $-\mathbf{a}$ から $\mathbf{a}$ までの $f(\mathbf{x})$ の積分は,

$$\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbf{a}}^{\mathbf{0}} f(-\mathbf{y})(-1)^n\,\mathrm{d}\mathbf{y} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad (\mathbf{y} = -\mathbf{x})$$

- 第一項は, 次と同じことである:

$$\begin{aligned}
\int_{\mathbf{a}}^{\mathbf{0}} f(-\mathbf{y})(-1)^n\,\mathrm{d}\mathbf{y} &= \int_{a_1}^{0}\int_{a_2}^{0}\cdots\int_{a_n}^{0} f(-\mathbf{y})(-1)^n\,\mathrm{d}y_1\,\mathrm{d}y_2\cdots\mathrm{d}y_n \\
&= \int_{0}^{a_1}\int_{0}^{a_2}\cdots\int_{0}^{a_n} f(-\mathbf{y})\,\mathrm{d}y_1\,\mathrm{d}y_2\cdots\mathrm{d}y_n \\
&= \int_{\mathbf{0}}^{\mathbf{a}} f(-\mathbf{y})\,\mathrm{d}\mathbf{y}
\end{aligned}$$

- 積分範囲を入れ替えると, 符号が反転することに注意 (上では $n$ 回入れ替える).

# 偶関数，奇関数の積分 (多変数関数)

- $\mathbf{a} = (a_1, a_2, \ldots, a_n)^\top$ として，$-\mathbf{a}$ から $\mathbf{a}$ までの $f(\mathbf{x})$ の積分は，

$$\begin{aligned}
\int_{-\mathbf{a}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \int_{-\mathbf{a}}^{\mathbf{0}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbf{a}}^{\mathbf{0}} f(-\mathbf{y})(-1)^n\,\mathrm{d}\mathbf{y} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad (\mathbf{y} = -\mathbf{x}) \\
&= \int_{\mathbf{0}}^{\mathbf{a}} f(-\mathbf{y})\,\mathrm{d}\mathbf{y} + \int_{\mathbf{0}}^{\mathbf{a}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}
\end{aligned}$$

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

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

# ガウス積分

## ガウス積分の基本形

$$\int_0^{\infty} \exp\left(-x^2\right) \mathrm{d}x = \frac{\sqrt{\pi}}{2}$$

上記の積分を $I$ とおくと, $I^2$ は

$$I^2 = \int_0^{\infty} \int_0^{\infty} \exp\left(-\left(x^2 + y^2\right)\right) \mathrm{d}x \, \mathrm{d}y$$

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

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

# ガウス積分

$x = r\cos\theta$, $y = r\sin\theta$, $x, y \in [0, \infty]$ であるから, $r \in [0, \infty)$, $\theta \in [0, \dfrac{\pi}{2}]$.
ヤコビアンは $|r| = r$ であるから, $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$.
以上より, $I^2$ は

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

であるから, $I = \dfrac{\sqrt{\pi}}{2}$.

# ガウス積分

## ガウス積分の基本形

$$\int_{-\infty}^{\infty} \exp\left(-x^2\right) \mathrm{d}x = \sqrt{\pi}$$

上記の積分を $I$ とおくと, $I^2$ は

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\left(x^2+y^2\right)\right) \mathrm{d}x\,\mathrm{d}y$$

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

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

# ガウス積分

## ガウス積分の基本形

$$\int_0^\infty \exp\left(-x^2\right) \mathrm{d}x = \frac{\sqrt{\pi}}{2}$$
$$\int_{-\infty}^\infty \exp\left(-x^2\right) \mathrm{d}x = \sqrt{\pi}$$

$\exp\left(-x^2\right)$ は偶関数 $(f(x) = f(-x))$ であるから，$-\infty$ から $\infty$ までの積分値は，$0$ から $\infty$ 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

$$\int_0^\infty \exp\left(-ax^2\right)\mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{a}} \quad (a>0)$$
$$\int_{-\infty}^\infty \exp\left(-ax^2\right)\mathrm{d}x = \sqrt{\frac{\pi}{a}} \quad (a>0)$$

$y=\sqrt{a}x$ とすると, $x=\dfrac{y}{\sqrt{a}}$, $\dfrac{\mathrm{d}x}{\mathrm{d}y}=\dfrac{1}{\sqrt{a}}$ だから,

$$\int_0^\infty \exp\left(-ax^2\right)\mathrm{d}x = \int_0^\infty \exp\left(-y^2\right)\frac{\mathrm{d}x}{\mathrm{d}y}\mathrm{d}y = \frac{1}{\sqrt{a}}\int_0^\infty \exp\left(-y^2\right)\mathrm{d}y = \frac{1}{\sqrt{a}}\frac{\sqrt{\pi}}{2}$$

$\exp\left(-ax^2\right)$ 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

$$\int_0^\infty x \exp\left(-ax^2\right) \mathrm{d}x = \frac{1}{2a} \quad (a > 0)$$

$$\int_{-\infty}^\infty x \exp\left(-ax^2\right) \mathrm{d}x = 0 \quad (a > 0)$$

次のように計算できる:

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

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

# ガウス積分

## ガウス積分の漸化式

$I_n = \int_0^\infty x^n \exp\left(-ax^2\right) \mathrm{d}x$ 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 ($\lim_{x\to\infty} x^n \exp(-x) = 0$ に注意):

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

# ガウス積分

## ガウス積分の漸化式

$I_n = \int_0^\infty x^n \exp\left(-ax^2\right) \mathrm{d}x$ 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:

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

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)

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

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

$I_0 = \int_0^\infty \exp\left(-ax^2\right) \mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{a}}$ だから，

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

$x^2 \exp\left(-ax^2\right)$ 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)

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

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

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

$I_1 = \int_0^\infty x \exp\left(-ax^2\right) \mathrm{d}x = \frac{1}{2a}$ だから，

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

$x^3 \exp\left(-ax^2\right)$ は奇関数だから，$-\infty$ から $\infty$ 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)

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

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

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

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