行列輪講: 第 9 回 ガウス分布 3

松谷研究室

June 2, 2024

目次

このスライドの概要

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

目次

特性関数 (Characteristic Function)

特性関数

確率変数 X の特性関数 $\varphi_X(t)$ は, 次のように定義される.

$$\varphi_X(t) = \mathbb{E}\left[\exp(itX)\right]$$

X の確率密度を $p(x)$ とすれば, 特性関数は, 確率密度のフーリエ変換として表せる.

$$\varphi_X(t) = \int_{-\infty}^{\infty} \exp(itx)p(x)\,\mathrm{d}x$$

- i は虚数単位.
- 特性関数は, 確率分布に対して, 常に存在する.
  - モーメント母関数は, 存在しない場合がある (例えば, コーシー分布).
- 特性関数は, 確率分布と一対一で対応する.

特性関数 (Characteristic Function)

特性関数

確率変数 X のモーメント $\mathbb{E}\left[X^n\right]$ を求められる.

$$\mathbb{E}\left[X^n\right] = i^{-n} \left.\frac{\mathrm{d}^n}{\mathrm{d}t^n}\varphi_X(t)\right|_{t=0}$$

ガウス分布 $\mathcal{N}(x \mid \mu, \sigma^2)$ の特性関数は,

$$\mathbb{E}\left[\exp(itX)\right] = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)$$

特に, 標準正規分布 $\mathcal{N}(x \mid 0, 1)$ の特性関数は,

$$\mathbb{E}\left[\exp(itX)\right] = \exp\left(-\frac{t^2}{2}\right)$$

特性関数 (Characteristic Function)

- 確率変数 X の特性関数 $\varphi_X(t)$ は,

$$\begin{aligned}\varphi_X(t) = \mathbb{E}\left[\exp(itX)\right] &= \mathbb{E}\left[1 + itX + \frac{1}{2!}(itX)^2 + \frac{1}{3!}(itX)^3 + \cdots\right] \\ &= 1 + it\,\mathbb{E}\left[X\right] + \frac{1}{2!}(it)^2\,\mathbb{E}\left[X^2\right] + \frac{1}{3!}(it)^3\,\mathbb{E}\left[X^3\right] + \cdots\end{aligned}$$

- 例えば, 上式を t で $n = 3$ 回微分すれば,

$$\frac{\mathrm{d}^3}{\mathrm{d}t^3}\varphi_X(t) = i^3\,\mathbb{E}\left[X^3\right] + i^4 t\,\mathbb{E}\left[X^4\right] + \frac{1}{2!}i^5 t^2\,\mathbb{E}\left[X^5\right] + \cdots$$

- 上式に $t = 0$ を代入し, $i^{-3} = i$ を掛ければ,

$$i^{-3}\left.\frac{\mathrm{d}^3}{\mathrm{d}t^3}\varphi_X(t)\right|_{t=0} = \mathbb{E}\left[X^3\right]$$

特性関数 (Characteristic Function)

特性関数の性質

確率変数 X の線形変換 $Y = aX + b$ の特性関数は,

$$\varphi_Y(t) = \exp(ibt)\varphi_X(at)$$

2 つの独立な確率変数 X, Y があるとき,

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

和 $X + Y$ の特性関数は, X, Y の特性関数の積である.

$Y = aX + b$ の特性関数は, 次のように確認できる:

$$\begin{aligned}
\varphi_Y(t) &= \mathbb{E}\left[\exp(it(aX+b))\right] = \mathbb{E}\left[\exp(it(aX) + itb)\right] \\
&= \mathbb{E}\left[\exp(i(at)X)\exp(itb)\right] = \exp(itb)\,\mathbb{E}\left[\exp(i(at)X)\right] \\
&= \exp(itb)\varphi_X(at)
\end{aligned}$$

特性関数 (Characteristic Function)

- 2 つの独立な確率変数 X, Y があるとき，その和の特性関数は，

$$\begin{aligned}\varphi_{X+Y}(t) &= \mathbb{E}\left[\exp(it(X+Y))\right] = \mathbb{E}\left[\exp(itX)\exp(itY)\right] \\ &= \mathbb{E}\left[\exp(itX)\right]\mathbb{E}\left[\exp(itY)\right] = \varphi_X(t)\varphi_Y(t)\end{aligned}$$

ただし，X, Y は独立だから $(p(x,y) = p(x)p(y))$，

$$\begin{aligned}\mathbb{E}\left[\exp(it(X+Y))\right] &= \iint \exp(itx)\exp(ity)p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \exp(itx)p(x)\,\mathrm{d}x \int \exp(ity)p(y)\,\mathrm{d}y \\ &= \mathbb{E}\left[\exp(itx)\right]\mathbb{E}\left[\exp(ity)\right]\end{aligned}$$

特性関数 (Characteristic Function)

特性関数の性質

N 個の独立な確率変数 $X_1, \ldots, X_N$ があるとする.
係数 $a_1, \ldots, a_N$ による重み付き和 S_N を考える.

$$S_N = \sum_{i=1}^{N} a_i X_i$$

このような S_N の特性関数は,

$$\varphi_{S_N}(t) = \prod_{i=1}^{N} \varphi_{X_i}(a_i t) = \varphi_{X_1}(a_1 t) \cdots \varphi_{X_N}(a_N t)$$

中心極限定理 (Central Limit Theorem; CLT)

- 平均 μ, 分散 σ^2 の確率分布から, N 個のサンプル $X_1, \ldots, X_N$ を独立に得たとする ($\mathbb{E}\left[X_i\right] = \mu, \mathrm{Var}\left[X_i\right] = \sigma^2$).
- 言い換えると, 平均 μ, 分散 σ^2 の独立同分布に従う, N 個の確率変数 $X_1, \ldots, X_N$ があるとする.
- 標本平均 $\bar{\mu}$ と, 実際の平均 μ との誤差 Y を考える:

$$Y = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu) = \frac{\sqrt{N}}{\sigma} (\bar{\mu} - \mu), \quad \bar{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

- Y は, N 個の確率変数の和を, 平均 0, 分散 1 に標準化している.
- この確率変数 Y は, $N \to \infty$ の極限において, 平均 0, 分散 1 の正規分布 $\mathcal{N}(0, 1)$ に収束する (中心極限定理).

中心極限定理 (Central Limit Theorem; CLT)

- 補足: Y は, N 個の確率変数の和を, 平均 0, 分散 1 に標準化している.
- $\mathbb{E}\left[X_i\right] = \mu, \mathrm{Var}\left[X_i\right] = \sigma^2$ だから, Y の平均と分散は,

$$\mathbb{E}\left[Y\right] = \mathbb{E}\left[\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu)\right] = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(\mathbb{E}\left[X_i\right] - \mu) = 0$$

$$\mathrm{Var}\left[Y\right] = \mathrm{Var}\left[\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu)\right] = \frac{1}{\sigma^2 N}\sum_{i=1}^{N}\mathrm{Var}\left[X_i - \mu\right]$$

$$= \frac{1}{\sigma^2 N}\sum_{i=1}^{N}\mathrm{Var}\left[X_i\right] = \frac{1}{\sigma^2 N}\sigma^2 N = 1$$

- x, y が独立ならば, $\mathrm{Var}\left[x + y\right] = \mathrm{Var}\left[x\right] + \mathrm{Var}\left[y\right]$ (X_i は互いに独立)
- 性質: $\mathbb{E}\left[x + y\right] = \mathbb{E}\left[x\right] + \mathbb{E}\left[y\right], \mathbb{E}\left[x + a\right] = \mathbb{E}\left[x\right] + a$
- 性質: $\mathrm{Var}\left[ax\right] = a^2\,\mathrm{Var}\left[x\right], \mathrm{Var}\left[x + a\right] = \mathrm{Var}\left[x\right]$

中心極限定理 (Central Limit Theorem; CLT)

- 確率変数 Y の**特性関数** $\varphi_Y(t)$ を計算してみよう.

$$Y = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu) = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N} Z_i \quad (Z_i = X_i - \mu)$$

特性関数の性質から (Z_i は互いに独立),

$$\varphi_Y(t) = \mathbb{E}\left[\exp(itY)\right] = \prod_{i=1}^{N}\varphi_{Z_i}\left(\frac{t}{\sigma\sqrt{N}}\right) = \left(\varphi_Z\left(\frac{t}{\sigma\sqrt{N}}\right)\right)^N$$

$Z_1, \ldots, Z_N$ は同一の分布に従うので, 特性関数も同じ.
$\varphi_{Z_1} = \cdots = \varphi_{Z_N} = \varphi_Z$ とおいた.

中心極限定理 (Central Limit Theorem; CLT)

- 特性関数 φ_Z の, 2 次までのマクローリン展開を考える.

$$\varphi_Z\left(\frac{t}{\sigma\sqrt{N}}\right) = \varphi_Z(0) + \frac{t}{\sigma\sqrt{N}}\varphi'_Z(0) + \frac{1}{2!}\frac{t^2}{\sigma^2 N}\varphi''_Z(0) + O\left(N^{-\frac{3}{2}}\right)$$

- ここで, モーメントについての関係式を使うと,

$$i^{-n}\left.\frac{\mathrm{d}^n}{\mathrm{d}t^n}\varphi_X(t)\right|_{t=0} = \mathbb{E}\left[X^n\right] \quad \longrightarrow \quad \left.\frac{\mathrm{d}^n}{\mathrm{d}t^n}\varphi_X(t)\right|_{t=0} = i^n\,\mathbb{E}\left[X^n\right]$$

$\varphi_Z(0), \varphi'_Z(0), \varphi''_Z(0)$ は, $\mathbb{E}[Z] = 0, \mathrm{Var}[Z] = \sigma^2$ より,

$$\begin{aligned}
\varphi_Z(0) &= i^0\,\mathbb{E}\left[Z^0\right] = 1 \\
\varphi'_Z(0) &= i^1\,\mathbb{E}\left[Z^1\right] = i \cdot 0 = 0 \\
\varphi''_Z(0) &= i^2\,\mathbb{E}\left[Z^2\right] = -1 \cdot \left(\mathrm{Var}[Z] + \mathbb{E}[Z]^2\right) = -\sigma^2
\end{aligned}$$

中心極限定理 (Central Limit Theorem; CLT)

- 補足: $\mathbb{E}[X_i] = \mu, \mathrm{Var}[X_i] = \sigma^2$ ゆえ, $Z_i = X_i - \mu$ の平均と分散は,

$$\mathbb{E}[Z_i] = \mathbb{E}[X_i - \mu] = \mathbb{E}[X_i] - \mu = \mu - \mu = 0$$
$$\mathrm{Var}[Z_i] = \mathrm{Var}[X_i - \mu] = \mathrm{Var}[X_i] = \sigma^2$$

  $\mathbb{E}[x + a] = \mathbb{E}[x] + a,\ \mathrm{Var}[x + a] = \mathrm{Var}[x]$ の関係を使おう.

- $\varphi_Z(0) = 1, \varphi'_Z(0) = 0, \varphi''_Z(0) = -\sigma^2$ を代入すれば,

$$\begin{aligned}
\varphi_Z\left(\frac{t}{\sigma\sqrt{N}}\right) &= \varphi_Z(0) + \frac{t}{\sigma\sqrt{N}}\varphi'_Z(0) + \frac{1}{2!}\frac{t^2}{\sigma^2 N}\varphi''_Z(0) + O\left(N^{-\frac{3}{2}}\right) \\
&= 1 - \frac{1}{2}\frac{t^2}{\sigma^2 N}\sigma^2 + O\left(N^{-\frac{3}{2}}\right) \\
&= 1 - \frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)
\end{aligned}$$

中心極限定理 (Central Limit Theorem; CLT)

- 確率変数 Y の特性関数 $\varphi_Y(t)$ は,

$$\varphi_Y(t) = \left(\varphi_Z\left(\frac{t}{\sigma\sqrt{N}}\right)\right)^N = \left(1 - \frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)\right)^N$$

$N \to \infty$ の極限を考えると,

$$\varphi_Y(t) = \left(1 + \frac{-\frac{t^2}{2} + O\left(N^{-\frac{1}{2}}\right)}{N}\right)^N \to \exp\left(-\frac{t^2}{2}\right)$$

これは標準正規分布 $\mathcal{N}(0,1)$ の特性関数と一致する.

- $\exp$ に関する次の定義を用いた:

$$\exp(x) = \lim_{N\to\infty}\left(1 + \frac{x}{N}\right)^N$$

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 μ, 分散 σ^2 の独立同分布に従う, N 個の確率変数 $X_1, \ldots, X_N$ があるとする. これらの確率変数の和を標準化した, 新たな確率変数 Y を考える:

$$Y = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu) = \frac{\sqrt{N}}{\sigma} (\bar{\mu} - \mu), \quad \bar{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$N \to \infty$ の極限において, 確率変数 Y の分布は, 平均 0, 分散 1 の標準正規分布 $\mathcal{N}(0, 1)$ に収束する.

- 標準化: 平均 0, 分散 1 となっている ($\mathbb{E}[Y] = 0, \mathrm{Var}[Y] = 1$).
- X_i がどのような分布であっても, N を大きくすれば, Y は近似的に正規分布に従う.
- Y は, 標本平均 $\bar{\mu}$ と, 真の平均 μ (母平均) との誤差を表している.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

$N \to \infty$ の極限において，確率変数 Y の分布は，平均 0，分散 1 の標準正規分布 $\mathcal{N}(0, 1)$ に収束する．次のようにも表せる．

$$\Pr(Y \leq \alpha) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y$$

$\Pr(a \leq Y \leq b)$ は，確率変数 Y が $[a, b]$ の範囲をとる確率である．
Y の分布を $p(y)$ とすれば，

$$\Pr(a \leq Y \leq b) = \int_a^b p(y) \, \mathrm{d}y$$

特に，$a = -\infty$ であれば，

$$\Pr(Y \leq b) = \int_{-\infty}^b p(y) \, \mathrm{d}y$$

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 μ, 分散 σ^2 の独立同分布に従う, N 個の確率変数 $X_1, \ldots, X_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数を考える:

$$Y = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu), \quad Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu)$$

$$\mathbb{E}[Y] = 0, \mathrm{Var}[Y] = 1, \mathbb{E}[Z] = 0, \mathrm{Var}[Z] = \sigma^2$$

$N \to \infty$ の極限において,

- Y の分布は, 標準正規分布 $\mathcal{N}(0, 1)$ に収束する.
- Z の分布は, 正規分布 $\mathcal{N}(0, \sigma^2)$ に収束する.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 μ, 分散 σ^2 の独立同分布に従う, N 個の確率変数 $X_1, \ldots, X_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数を考える:

$$W = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i, \quad V = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$\mathbb{E}\left[W\right] = \sqrt{N}\mu, \mathrm{Var}\left[W\right] = \sigma^2, \mathbb{E}\left[V\right] = \mu, \mathrm{Var}\left[V\right] = \frac{\sigma^2}{N}$$

$N \to \infty$ の極限において,

- W の分布は, 正規分布 $\mathcal{N}(\sqrt{N}\mu, \sigma^2)$ に収束する.
- 標本平均 V の分布は, 正規分布 $\mathcal{N}(\mu, \frac{\sigma^2}{N})$ に収束する.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 μ, 分散 σ^2 の独立同分布に従う, N 個の確率変数 $X_1, \dots, X_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数を考える:

$$S = \sum_{i=1}^{N} X_i$$

$$\mathbb{E}[S] = N\mu, \mathrm{Var}[S] = N\sigma^2$$

$N \to \infty$ の極限において,

- 和 S の分布は, 正規分布 $\mathcal{N}(N\mu, N\sigma^2)$ に収束する.

モーメント母関数 (Moment-generating Function)

モーメント母関数 (再訪)

確率変数 X のモーメント母関数 $M_X(t)$ は, 次のように定義される.

$$M_X(t) = \mathbb{E}\left[\exp(tX)\right] = \mathbb{E}\left[1 + tX + \frac{1}{2!}t^2X^2 + \frac{1}{3!}t^3X^3 + \cdots\right]$$

t で n 回微分して, $t = 0$ を代入すると, $\mathbb{E}\left[X^n\right]$ が得られる.

$$\mathbb{E}\left[X^n\right] = M_X^{(n)}(t)\Big|_{t=0} = \frac{\partial^n}{\partial t^n}\mathbb{E}\left[\exp(tX)\right]\Big|_{t=0}$$

ガウス分布 $\mathcal{N}(x \mid \mu, \sigma^2)$ に対するモーメント母関数は,

$$M_X(t) = \mathbb{E}\left[\exp(tX)\right] = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

モーメント母関数と特性関数の関係

モーメント母関数と特性関数

確率変数 X のモーメント母関数 $M_X(t)$:

$$M_X(t) = \mathbb{E}\left[\exp(tX)\right]$$

確率変数 X の特性関数 $\varphi_X(t)$:

$$\varphi_X(t) = \mathbb{E}\left[\exp(itX)\right]$$

モーメント母関数と特性関数との関係:

$$\varphi_X(t) = M_X(it)$$

特性関数 (多次元)

特性関数 (多次元)

確率変数 $\mathbf{X}$ の特性関数 $\varphi_{\mathbf{X}}(\mathbf{t})$ は, 次のように定義される ($\mathbf{X}$ はベクトル).

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[\exp\left(i\mathbf{t}^\top \mathbf{X}\right)\right]$$

$\mathbf{X}$ の確率密度を $p(\mathbf{x})$ とすれば, 特性関数は, 確率密度のフーリエ変換として表せる.

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_{-\infty}^{\infty} \exp\left(i\mathbf{t}^\top \mathbf{x}\right) p(\mathbf{x})\, \mathrm{d}\mathbf{x}$$

特性関数 (多次元)

特性関数 (多次元)

多変量ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ の特性関数は,

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left(i\boldsymbol{\mu}^\top \mathbf{t} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}\right)$$

特に, 標準正規分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \mathbf{I})$ の特性関数は,

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left(-\frac{1}{2}\mathbf{t}^\top \mathbf{t}\right)$$

中心極限定理 (Central Limit Theorem; CLT)

- 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の確率分布から, N 個のサンプル $\mathbf{X}_1, \ldots, \mathbf{X}_N$ を独立に得たとする ($\mathbb{E}\left[\mathbf{X}_i\right] = \boldsymbol{\mu}, \mathrm{Var}\left[\mathbf{X}_i\right] = \boldsymbol{\Sigma}$).
- 言い換えると, 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の独立同分布に従う, N 個の確率変数 $\mathbf{X}_1, \ldots, \mathbf{X}_N$ があるとする.
- 標本平均 $\bar{\boldsymbol{\mu}}$ と, 実際の平均 $\boldsymbol{\mu}$ との誤差 $\mathbf{Y}$ を考える:

$$\mathbf{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \left(\mathbf{X}_i - \boldsymbol{\mu}\right) = \sqrt{N} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}\right), \quad \bar{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i$$

- 対角行列 $\boldsymbol{\Lambda}$, 直交行列 $\mathbf{U}$ は, 共分散の対角化 $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ により得られる.
- この確率変数 $\mathbf{Y}$ は, $N \to \infty$ の極限において, 平均 $\mathbf{0}$, 共分散 $\mathbf{I}$ の正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ に収束する (中心極限定理).

中心極限定理 (Central Limit Theorem; CLT)

- 補足: $\mathbf{Y}$ は, N 個の確率変数の和を, 平均 $\mathbf{0}$, 共分散 $\mathbf{I}$ に標準化する.
- $\mathbb{E}\left[\mathbf{X}_i\right] = \boldsymbol{\mu}, \mathrm{Var}\left[\mathbf{X}_i\right] = \boldsymbol{\Sigma}$ だから, $\mathbf{Y}$ の平均は,

$$\begin{aligned}\mathbb{E}\left[\mathbf{Y}\right] &= \mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{\top}\left(\mathbf{X}_i - \boldsymbol{\mu}\right)\right] \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{\top}\left(\boldsymbol{\mu} - \boldsymbol{\mu}\right) = \mathbf{0}\end{aligned}$$

- 性質: $\mathbb{E}\left[\mathbf{x} + \mathbf{y}\right] = \mathbb{E}\left[\mathbf{x}\right] + \mathbb{E}\left[\mathbf{y}\right]$, $\mathbb{E}\left[\mathbf{x} + \mathbf{a}\right] = \mathbb{E}\left[\mathbf{x}\right] + \mathbf{a}$

中心極限定理 (Central Limit Theorem; CLT)

- $\mathbb{E}[\mathbf{X}_i] = \boldsymbol{\mu}, \mathrm{Var}[\mathbf{X}_i] = \boldsymbol{\Sigma}$ だから，$\mathbf{Y}$ の共分散は，

$$
\begin{aligned}
\mathrm{Var}[\mathbf{Y}] &= \mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top(\mathbf{X}_i - \boldsymbol{\mu})\right] \\
&= \sum_{i=1}^{N}\frac{1}{N}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top \mathrm{Var}[\mathbf{X}_i - \boldsymbol{\mu}]\left(\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\right)^\top \\
&= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\left(\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\right)^\top = \mathbf{I}
\end{aligned}
$$

- $\mathbf{x}, \mathbf{y}$ が独立ならば，$\mathrm{Var}[\mathbf{x}+\mathbf{y}] = \mathrm{Var}[\mathbf{x}] + \mathrm{Var}[\mathbf{y}]$ ($\mathbf{X}_i$ は互いに独立)
- 性質: $\mathrm{Var}[\mathbf{A}\mathbf{x}] = \mathbf{A}\,\mathrm{Var}[\mathbf{x}]\,\mathbf{A}^\top$，$\mathrm{Var}[\mathbf{x}+\mathbf{a}] = \mathrm{Var}[\mathbf{x}]$
- 注意: $\mathrm{Var}[\mathbf{X}_i - \boldsymbol{\mu}] = \mathrm{Var}[\mathbf{X}_i] = \boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$

中心極限定理 (Central Limit Theorem; CLT)

- 次の特性関数 $\varphi_{\mathbf{Y}}(\mathbf{t}, u)$ を考えよう ($\mathbf{Y}$ の特性関数に u を足した).

$$\varphi_{\mathbf{Y}}(\mathbf{t}, u) = \mathbb{E}\left[\exp\left(iu\mathbf{t}^\top\mathbf{Y}\right)\right]$$

- 1 次元の変数 $\mathbf{t}^\top\mathbf{Y}$ についての特性関数とみなせる.

$$\begin{aligned}
\mathbf{t}^\top\mathbf{Y} &= \mathbf{t}^\top\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top(\mathbf{X}_i - \boldsymbol{\mu}) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{t}^\top\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\mathbf{X}_i - \mathbf{t}^\top\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\boldsymbol{\mu}\right)
\end{aligned}$$

- 1 次元の確率変数 $\mathbf{t}^\top\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\mathbf{X}_i$ の和をとって, $\mathbf{t}^\top\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\boldsymbol{\mu}$ だけ移動させ, さらに $\dfrac{1}{\sqrt{N}}$ でスケールしたもの.

中心極限定理 (Central Limit Theorem; CLT)

- 1 次元の確率変数 $\mathbf{t}^\top \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \mathbf{X}_i$ の和をとって, $\mathbf{t}^\top \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \boldsymbol{\mu}$ だけ移動させ, さらに $\dfrac{1}{\sqrt{N}}$ でスケールしたもの.

$$\mathbf{t}^\top \mathbf{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{t}^\top \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \mathbf{X}_i - \mathbf{t}^\top \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top \boldsymbol{\mu} \right)$$

- $\mathbb{E}\left[\mathbf{Y}\right] = \mathbf{0}, \mathrm{Var}\left[\mathbf{Y}\right] = \mathbf{I}$ だから,

$$\mathbb{E}\left[\mathbf{t}^\top \mathbf{Y}\right] = \mathbf{t}^\top \mathbb{E}\left[\mathbf{Y}\right] = 0$$
$$\mathrm{Var}\left[\mathbf{t}^\top \mathbf{Y}\right] = \mathbf{t}^\top \mathrm{Var}\left[\mathbf{Y}\right] \mathbf{t} = \mathbf{t}^\top \mathbf{t}$$

- 1 次元の場合の中心極限定理から, $N \to \infty$ の極限において, $\mathbf{t}^\top \mathbf{Y}$ は $\mathcal{N}(0, \mathbf{t}^\top \mathbf{t})$ に従う.

中心極限定理 (Central Limit Theorem; CLT)

- 1 次元の場合の中心極限定理から，$N \to \infty$ の極限において，$\mathbf{t}^\top \mathbf{Y}$ は $\mathcal{N}(0, \mathbf{t}^\top \mathbf{t})$ に従う．よって，$\mathbf{t}^\top \mathbf{Y}$ の特性関数 $\varphi_{\mathbf{Y}}(\mathbf{t}, u)$ は，

$$\lim_{N\to\infty} \varphi_{\mathbf{Y}}(\mathbf{t}, u) = \exp\left(-\frac{1}{2}\mathbf{t}^\top \mathbf{t} u^2\right)$$

- $u = 1$ とおけば，$\mathbf{Y}$ の特性関数 $\varphi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}\left[\exp\left(i\mathbf{t}^\top \mathbf{Y}\right)\right]$ は，

$$\lim_{N\to\infty} \varphi_{\mathbf{Y}}(\mathbf{t}) = \exp\left(-\frac{1}{2}\mathbf{t}^\top \mathbf{t}\right)$$

- これは，標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ の特性関数と対応する．
- 従って，確率変数 $\mathbf{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^\top (\mathbf{X}_i - \boldsymbol{\mu})$ は，$N \to \infty$ の極限において，標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ に収束する．

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の独立同分布に従う, N 個の確率変数 $\mathbf{X}_1, \ldots, \mathbf{X}_N$ があるとする. これらの確率変数の和を標準化した, 新たな確率変数 $\mathbf{Y}$ を考える:

$$\mathbf{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^{\top} \left(\mathbf{X}_i - \boldsymbol{\mu}\right) = \sqrt{N} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^{\top} \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}\right), \quad \bar{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i$$

$N \to \infty$ の極限において, 確率変数 $\mathbf{Y}$ の分布は, 標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ に収束する.

- 確率変数 $\mathbf{Y}$ は, 平均 $\mathbf{0}$, 分散 $\mathbf{I}$ となるように標準化されている.
- $\mathbf{X}_i$ は, どのような分布であってもよい.
- $\boldsymbol{\Lambda}, \mathbf{U}$ は, 共分散 $\boldsymbol{\Sigma}$ を対角化して得られる ($\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top}$).
- $\mathbf{U}$ は直交行列, $\boldsymbol{\Lambda}$ は $\boldsymbol{\Sigma}$ の固有値を並べた対角行列.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の独立同分布に従う, N 個の確率変数 $\mathbf{X}_1, \ldots, \mathbf{X}_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数 $\mathbf{Y}$ を考える:

$$\mathbf{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}^{\top} \left(\mathbf{X}_i - \boldsymbol{\mu} \right), \quad \mathbf{Z} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{X}_i - \boldsymbol{\mu} \right)$$

$$\mathbb{E}\left[\mathbf{Y}\right] = \mathbf{0}, \mathrm{Var}\left[\mathbf{Y}\right] = \mathbf{I}, \mathbb{E}\left[\mathbf{Z}\right] = \mathbf{0}, \mathrm{Var}\left[\mathbf{Z}\right] = \boldsymbol{\Sigma}$$

$N \to \infty$ の極限において,

- $\mathbf{Y}$ の分布は, 標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ に収束する.
- $\mathbf{Z}$ の分布は, 正規分布 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ に収束する.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の独立同分布に従う, N 個の確率変数 $\mathbf{X}_1, \ldots, \mathbf{X}_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数を考える:

$$\mathbf{W} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{X}_i, \quad \mathbf{V} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i$$

$$\mathbb{E}\left[\mathbf{W}\right] = \sqrt{N}\boldsymbol{\mu}, \mathrm{Var}\left[\mathbf{W}\right] = \boldsymbol{\Sigma}, \mathbb{E}\left[\mathbf{V}\right] = \boldsymbol{\mu}, \mathrm{Var}\left[\mathbf{V}\right] = \frac{\boldsymbol{\Sigma}}{N}$$

$N \to \infty$ の極限において,

- $\mathbf{W}$ の分布は, 正規分布 $\mathcal{N}(\sqrt{N}\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に収束する.
- 標本平均 $\mathbf{V}$ の分布は, 正規分布 $\mathcal{N}(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{N})$ に収束する.

中心極限定理 (Central Limit Theorem; CLT)

中心極限定理

平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の独立同分布に従う, N 個の確率変数 $\mathbf{X}_1, \ldots, \mathbf{X}_N$ があるとする. これらの確率変数の和を基にした, 新たな確率変数を考える:

$$\mathbf{S} = \sum_{i=1}^{N} \mathbf{X}_i$$

$$\mathbb{E}\left[\mathbf{S}\right] = N\boldsymbol{\mu}, \mathrm{Var}\left[\mathbf{S}\right] = N\boldsymbol{\Sigma}$$

$N \to \infty$ の極限において,

- 和 S の分布は, 正規分布 $\mathcal{N}(N\boldsymbol{\mu}, N\boldsymbol{\Sigma})$ に収束する.

目次

エントロピー (Entropy) (再掲)

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

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

条件付きエントロピー (Conditional Entropy) (再掲)

- 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

ガウス分布のエントロピー

- $\mathbf{x}$ (M 次元) に関するガウス分布 $p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ を考える:

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

- ガウス分布のエントロピー $\mathrm{H}[\mathbf{x}]$ を計算しよう.

$$\begin{aligned}
&\mathrm{H}[\mathbf{x}] = -\mathbb{E}[\ln p(\mathbf{x})] = -\int p(\mathbf{x}) \ln p(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= -\int p(\mathbf{x}) \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \frac{1}{2} \ln\left((2\pi)^M \det \boldsymbol{\Sigma} \right) \right) \mathrm{d}\mathbf{x} \\
&= \frac{1}{2} \ln\left((2\pi)^M \det \boldsymbol{\Sigma} \right) + \frac{1}{2} \int p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \, \mathrm{d}\mathbf{x} \\
&= \frac{1}{2} \ln\left((2\pi)^M \det \boldsymbol{\Sigma} \right) + \frac{1}{2} \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]
\end{aligned}$$

ガウス分布のエントロピー

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ のエントロピー $\mathrm{H}[\mathbf{x}]$:

$$\mathrm{H}[\mathbf{x}] = \frac{1}{2} \ln\left((2\pi)^M \det \boldsymbol{\Sigma}\right) + \frac{1}{2} \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

- 期待値 $\mathbb{E}[\cdot]$ はスカラー. $\mathbb{E}[\operatorname{tr} x] = \operatorname{tr} \mathbb{E}[x]$ なので,

$$\begin{aligned}
&\mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right] = \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]\right) \\
&= \mathbb{E}\left[\operatorname{tr}\left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)\right] = \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^\top\right)\right] \\
&= \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^\top\right]\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}\right) = \operatorname{tr} \mathbf{I} = M
\end{aligned}$$

- トレースの性質: $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{BCA}) = \operatorname{tr}(\mathbf{CAB})$
- $\mathbf{x}$ は M 次元なので, $\boldsymbol{\Sigma}$, $\mathbf{I}$ は M 次正方行列.

ガウス分布のエントロピー

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ のエントロピー $\mathrm{H}[\mathbf{x}]$:

$$
\begin{aligned}
\mathrm{H}[\mathbf{x}] &= \frac{1}{2}\ln\left((2\pi)^M \det \boldsymbol{\Sigma}\right) + \frac{1}{2}\mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right] \\
&= \frac{1}{2}\ln\left((2\pi)^M \det \boldsymbol{\Sigma}\right) + \frac{1}{2}M \\
&= \frac{1}{2}\ln\left((2\pi)^M \det \boldsymbol{\Sigma}\right) + \frac{1}{2}\ln e^M \\
&= \frac{1}{2}\ln\left((2\pi e)^M \det \boldsymbol{\Sigma}\right)
\end{aligned}
$$

- 共分散 Σ についての関数となる ($\boldsymbol{\mu}$ にはよらない).
- 共分散が大きくなれば, その分, エントロピーが増大する.

相互情報量 (Mutual Information) (再掲)

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

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

- $I(\mathrm{x}, \mathrm{y})$ 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 $(p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}))$, $I(\mathrm{x}, \mathrm{y}) = 0$.

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

- x, y が独立であれば, x (y) のことが分かっても, 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) (再掲)

- 相互情報量 $I\left(\mathbf{x},\mathbf{y}\right)$:

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

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

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

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

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

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

条件付きガウス分布と周辺ガウス分布 (再掲)

条件付きガウス分布

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

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

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

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

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

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

ガウス分布の相互情報量

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

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

- 周辺分布: $p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$, $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$.
- $\mathbf{x}, \mathbf{y}$ について, 次の相互情報量 $I(\mathbf{x}, \mathbf{y})$ を考える:

$$\begin{aligned}
I(\mathbf{x}, \mathbf{y}) &= \mathrm{H}[\mathbf{x}] + \mathrm{H}[\mathbf{y}] - \mathrm{H}[\mathbf{x}, \mathbf{y}] \\
&= \frac{1}{2} \ln\left((2\pi e)^M \det \boldsymbol{\Sigma}_{xx}\right) + \frac{1}{2} \ln\left((2\pi e)^N \det \boldsymbol{\Sigma}_{yy}\right) \\
&\quad - \frac{1}{2} \ln\left((2\pi e)^{M+N} \det \boldsymbol{\Sigma}\right) \\
&= -\frac{1}{2} \ln\left(\frac{\det \boldsymbol{\Sigma}}{\det \boldsymbol{\Sigma}_{xx} \det \boldsymbol{\Sigma}_{yy}}\right)
\end{aligned}$$

ガウス分布の相互情報量

- ブロック行列の行列式を考えれば (第 2 回), $\det \boldsymbol{\Sigma}$ は,

$$\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{xx} \det\left(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right)$$
$$\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{yy} \det\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)$$

- ブロック行列の行列式 (再掲):

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

- Weinstein-Aronszajn Identity (再掲):

$$\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA}) \quad (\mathbf{A}\text{ は } m \times n,\ \mathbf{B}\text{ は } n \times m\text{ 行列})$$

ガウス分布の相互情報量

- $\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{xx} \det\left(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right)$ を使えば,

$$
\begin{aligned}
I(\mathbf{x}, \mathbf{y}) &= -\frac{1}{2} \ln\left(\frac{\det \boldsymbol{\Sigma}_{xx} \det\left(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right)}{\det \boldsymbol{\Sigma}_{xx} \det \boldsymbol{\Sigma}_{yy}}\right) \\
&= -\frac{1}{2} \ln \det \boldsymbol{\Sigma}_{yy}^{-1} \det\left(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right) \\
&= -\frac{1}{2} \ln \det\left(\mathbf{I} - \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\right)
\end{aligned}
$$

- $\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{yy} \det\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)$ を使えば,

$$
\begin{aligned}
I(\mathbf{x}, \mathbf{y}) &= -\frac{1}{2} \ln\left(\frac{\det \boldsymbol{\Sigma}_{yy} \det\left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)}{\det \boldsymbol{\Sigma}_{xx} \det \boldsymbol{\Sigma}_{yy}}\right) \\
&= -\frac{1}{2} \ln \det\left(\mathbf{I} - \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)
\end{aligned}
$$

ガウス分布の相互情報量

- $\mathbf{x}, \mathbf{y}$ の同時分布がガウス分布であるとき，相互情報量は

$$\begin{aligned} I(\mathbf{x}, \mathbf{y}) &= -\frac{1}{2} \ln \det \left(\mathbf{I} - \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy} \right) \\ &= -\frac{1}{2} \ln \det \left(\mathbf{I} - \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx} \right) \end{aligned}$$

- $\mathbf{x}, \mathbf{y}$ が独立 ($\rightarrow$ 無相関) であったとする．
- このとき，共分散 $\boldsymbol{\Sigma}_{xy}, \boldsymbol{\Sigma}_{yx}$ は $\mathbf{0}$ になる：

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

- 従って，相互情報量も $I(\mathbf{x}, \mathbf{y}) = -\frac{1}{2} \ln \det \mathbf{I} = 0$ となる．

目次

不偏推定量 (Unbiased Estimate) (再掲)

- 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, 真のパラメータ $\boldsymbol{\theta}$ に一致する (偏りがない).

クラメール・ラオの下限 (再掲)

- 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 $\hat{\boldsymbol{\theta}}$ の共分散について, 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 (再掲)

- 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.

ガウス分布とクラメール・ラオの下限

- 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の M 次元ガウス分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ から, K 個のサンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ が独立に得られたとする:

$$\forall k \quad \mathbf{x}_{k,\mathrm{meas}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- K 個のサンプルは, 互いに独立に, 同一のガウス分布に従う.
- これら K 個の確率変数の同時分布もまた, ガウス分布になる.
- 変数をつなげて $\mathbf{x} = (\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}})$ とおくと,

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

平均は $\mathbf{A}\boldsymbol{\mu}$, 共分散は $\mathbf{B}$ である ($\mathbf{A}, \mathbf{B}$ は後述).

ガウス分布とクラメール・ラオの下限

- 補足: $\mathbf{x} = \left(\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}\right)$ の平均は $\mathbf{A}\boldsymbol{\mu}$:

$$\mathbb{E}\left[\mathbf{x}\right] = \begin{pmatrix} \mathbb{E}\left[\mathbf{x}_{1,\mathrm{meas}}\right] \\ \vdots \\ \mathbb{E}\left[\mathbf{x}_{K,\mathrm{meas}}\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{pmatrix} \boldsymbol{\mu} = \mathbf{A}\boldsymbol{\mu}$$

- $\mathbf{A}$ は, 単位行列を並べた $MK \times M$ 行列:

$$\mathbf{A} = \underbrace{\begin{pmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{pmatrix}^{\top}}_{K \text{ 個}}$$

ガウス分布とクラメール・ラオの下限

- 補足: $\mathbf{x} = (\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}})$ の共分散は B:

$$
\begin{aligned}
\mathrm{Var}\left[\mathbf{x}\right] &= \mathbb{E}\left[\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)^{\top}\right] \\
&= \begin{pmatrix} \mathrm{Cov}\left(\mathbf{x}_{1,\mathrm{meas}}, \mathbf{x}_{1,\mathrm{meas}}\right) & \cdots & \mathrm{Cov}\left(\mathbf{x}_{1,\mathrm{meas}}, \mathbf{x}_{K,\mathrm{meas}}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(\mathbf{x}_{K,\mathrm{meas}}, \mathbf{x}_{1,\mathrm{meas}}\right) & \cdots & \mathrm{Cov}\left(\mathbf{x}_{K,\mathrm{meas}}, \mathbf{x}_{K,\mathrm{meas}}\right) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{pmatrix} = \mathrm{diag}(\boldsymbol{\Sigma}, \ldots, \boldsymbol{\Sigma}) = \mathrm{B}
\end{aligned}
$$

- K 個のサンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ は, 互いに独立無相関.
- $k \neq l$ ならば $\mathrm{Cov}\left(\mathbf{x}_{k,\mathrm{meas}}, \mathbf{x}_{l,\mathrm{meas}}\right) = \mathbf{0}$.
- $\mathrm{Cov}\left(\mathbf{x}_{\mathrm{meas}}, \mathbf{x}_{\mathrm{meas}}\right) = \mathrm{Var}\left[\mathbf{x}_{\mathrm{meas}}\right] = \boldsymbol{\Sigma}$.
- B は, $\boldsymbol{\Sigma}$ を並べた $MK \times MK$ のブロック対角行列.

ガウス分布とクラメール・ラオの下限

- K 個の確率変数 (サンプル) の同時分布は，ガウス分布 $\mathcal{N}(\mathbf{A}\boldsymbol{\mu}, \mathbf{B})$:

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{MK}{2}} \sqrt{\det \mathbf{B}}} \exp\left(-\frac{1}{2} \left(\mathbf{x} - \mathbf{A}\boldsymbol{\mu}\right)^\top \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{A}\boldsymbol{\mu}\right) \right)$$

- 対数をとり，$\boldsymbol{\mu}$ で微分すると，

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\mu}} \left(\mathbf{x} - \mathbf{A}\boldsymbol{\mu}\right)^\top \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{A}\boldsymbol{\mu}\right) \\ &= \left(\mathbf{x} - \mathbf{A}\boldsymbol{\mu}\right)^\top \mathbf{B}^{-1} \mathbf{A} \end{aligned}$$

- スカラのベクトルによる微分 (再掲):

$$\frac{\partial \left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)^\top \mathbf{C} \left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)}{\partial \mathbf{x}} = -2 \left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)^\top \mathbf{C}\mathbf{A} \quad (\mathbf{C} \text{ は対称})$$

ガウス分布とクラメール・ラオの下限

- フィッシャー情報行列 $\mathbf{I}(\mathbf{x} \mid \boldsymbol{\mu})$ は,

$$
\begin{aligned}
\mathbf{I}(\mathbf{x} \mid \boldsymbol{\mu}) &= \mathbb{E}\left[\left(\frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}}\right)^{\top}\left(\frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}}\right)\right] \\
&= \mathbb{E}\left[\mathbf{A}^{\top}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)^{\top}\mathbf{B}^{-1}\mathbf{A}\right] \\
&= \mathbf{A}^{\top}\mathbf{B}^{-1}\,\mathbb{E}\left[\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)\left(\mathbf{x}-\mathbf{A}\boldsymbol{\mu}\right)^{\top}\right]\mathbf{B}^{-1}\mathbf{A} \\
&= \mathbf{A}^{\top}\mathbf{B}^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbf{A} = \mathbf{A}^{\top}\mathbf{B}^{-1}\mathbf{A} \\
&= \begin{pmatrix}\mathbf{I} & \cdots & \mathbf{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{I} \\ \vdots \\ \mathbf{I}\end{pmatrix} = K\boldsymbol{\Sigma}^{-1}
\end{aligned}
$$

- 逆行列 (Cramér–Rao 下限) は, $\mathbf{I}(\mathbf{x} \mid \boldsymbol{\mu})^{-1} = \frac{1}{K}\boldsymbol{\Sigma}$.

ガウス分布とクラメール・ラオの下限

- 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の M 次元ガウス分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ から, K 個のサンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ が独立に得られたとする:

$$\forall k \quad \mathbf{x}_{k,\mathrm{meas}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- これらのサンプルを基に, 平均 μ の不偏推定量 $\hat{\mu}$ を求めたとする.
- $\hat{\mu}$ の共分散について, 以下が成り立つ:

$$\mathrm{Var}\left[\hat{\boldsymbol{\mu}}\right] = \mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^\top\right] \geq \mathbf{I}^{-1}(\mathbf{x} \mid \boldsymbol{\mu}) = \frac{1}{K}\boldsymbol{\Sigma}$$

- 平均の不偏推定量のばらつきは, サンプル数 K を増やせば, 小さくなる.
- 不偏推定量の求め方については考慮しなかった.

ガウス分布とクラメール・ラオの下限

- 平均 $\boldsymbol{\mu}$, 共分散 $\boldsymbol{\Sigma}$ の M 次元ガウス分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ から, K 個のサンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ が独立に得られたとする:

$$\forall k \quad \mathbf{x}_{k,\mathrm{meas}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- これらのサンプルを基に, 平均 $\boldsymbol{\mu}$ の不偏推定量 $\hat{\boldsymbol{\mu}}$ を求めたとする.
- 平均に関しては, 標本平均が不偏推定量となる:

$$\hat{\boldsymbol{\mu}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}_{k,\mathrm{meas}}$$

- 次のように確かめられる:

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

ガウス分布とクラメール・ラオの下限

- 標本平均 $\hat{\mu}$ の共分散は,

$$
\begin{aligned}
\mathrm{Var}\left[\hat{\boldsymbol{\mu}}\right] &= \mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^\top\right] \\
&= \mathbb{E}\left[\left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_{k,\mathrm{meas}} - \boldsymbol{\mu}\right)\left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_{k,\mathrm{meas}} - \boldsymbol{\mu}\right)^\top\right] \\
&= \frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{E}\left[\left(\mathbf{x}_{k,\mathrm{meas}} - \boldsymbol{\mu}\right)\left(\mathbf{x}_{l,\mathrm{meas}} - \boldsymbol{\mu}\right)^\top\right] \\
&= \frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathrm{Cov}\left(\mathbf{x}_{k,\mathrm{meas}}, \mathbf{x}_{l,\mathrm{meas}}\right) = \frac{1}{K}\boldsymbol{\Sigma}
\end{aligned}
$$

- K 個のサンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ は, 互いに独立無相関.
- $k \neq l$ ならば $\mathrm{Cov}\left(\mathbf{x}_{k,\mathrm{meas}}, \mathbf{x}_{l,\mathrm{meas}}\right) = \mathbf{0}$.
- また, $\mathrm{Cov}\left(\mathbf{x}_{\mathrm{meas}}, \mathbf{x}_{\mathrm{meas}}\right) = \boldsymbol{\Sigma}$.

ガウス分布とクラメール・ラオの下限

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

$$\mathrm{Var}\left[\hat{\boldsymbol{\mu}}\right] = \mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^{\top}\right] \geq \mathbf{I}^{-1}(\mathbf{x} \mid \boldsymbol{\mu}) = \frac{1}{K}\boldsymbol{\Sigma}$$

- 標本平均 $\hat{\mu}$ の共分散は，

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

- 標本平均 $\hat{\mu}$ の共分散は，クラメール・ラオの下限と一致する．
- サンプル $\mathbf{x}_{1,\mathrm{meas}}, \ldots, \mathbf{x}_{K,\mathrm{meas}}$ から，ガウス分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ の平均パラメータを推定するとき，標本平均を使うのが一番良さそうである．

目次

高次元空間の体積と表面積

- D 次元の (ユークリッド) 空間を考える.
- 半径 r の D 次元**超球**とは, $x_1^2 + x_2^2 + \cdots + x_D^2 \leq r^2$ をみたす領域のことである (中心から D だけ離れた点の集合).
- 半径 $r = 1$ のとき, D 次元の**単位超球** (単位球) という.

- 半径 r, D 次元超球の**体積** $V_D(r)$ と**表面積** $S_D(r)$ を考えよう.
- 体積は r^D に比例するから, $V_D(r) = K_D r^D$ とかける.
- 係数 K_D は, D のみに依存する (r によらない).
- 表面積は, 体積を微分したものだから, $S_D(r) = DK_D r^{D-1}$ となる.

$$
\begin{aligned}
V_D(r) &= K_D r^D \\
S_D(r) &= DK_D r^{D-1}
\end{aligned}
$$

高次元空間の体積と表面積

- D 次元の単位超球の体積, 表面積を $V_D = V_D(1)$, $S_D = S_D(1)$ とおく.
- このとき, $V_D(r) = V_D r^D$, $S_D(r) = S_D r^{D-1}$ である.
- D 次元超球の表面積 $S_D(r)$ について, 次の関係が成り立つ (PRML 演習 1.18):

$$\prod_{i=1}^{D} \int_{-\infty}^{\infty} \exp\left(-x_i^2\right) \mathrm{d}x_i = \int_0^{\infty} \exp\left(-r^2\right) S_D(r) \, \mathrm{d}r$$
$$= S_D \int_0^{\infty} \exp\left(-r^2\right) r^{D-1} \, \mathrm{d}r$$

- この式の左辺は, ガウス積分の結果を使うと,

$$\prod_{i=1}^{D} \int_{-\infty}^{\infty} \exp\left(-x_i^2\right) \mathrm{d}x_i = \prod_{i=1}^{D} \sqrt{\pi} = \pi^{\frac{D}{2}}$$

高次元空間の体積と表面積

- 続いて，右辺を計算すると，

$$
\begin{aligned}
S_D \int_0^\infty \exp\left(-r^2\right) r^{D-1}\,\mathrm{d}r &= S_D \int_0^\infty \exp(-u) u^{\frac{D-1}{2}} \cdot \frac{1}{2} u^{-\frac{1}{2}}\,\mathrm{d}u \\
&= \frac{1}{2} S_D \int_0^\infty \exp(-u) u^{\frac{D}{2}-1}\,\mathrm{d}u \\
&= \frac{1}{2} S_D \Gamma\left(\frac{D}{2}\right)
\end{aligned}
$$

- $u = r^2$ で置換した $\left(\frac{\mathrm{d}r}{\mathrm{d}u} = \frac{1}{2} u^{-\frac{1}{2}}\right)$.
- ガンマ関数 $\Gamma(x)$ は，次のように定義される:

$$
\Gamma(x) = \int_0^\infty u^{x-1} \exp(-u)\,\mathrm{d}u
$$

高次元空間の体積と表面積

- 左辺と右辺の結果から,

$$\pi^{\frac{D}{2}} = \frac{1}{2} S_D \Gamma\left(\frac{D}{2}\right)$$

- 以上より, D 次元の単位超球の表面積 S_D は,

$$S_D = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$$

- また, 半径 r の D 次元超球の表面積 $S_D(r) = S_D r^{D-1}$ は,

$$S_D(r) = S_D r^{D-1} = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} r^{D-1}$$

高次元空間の体積と表面積

- 係数 K_D を用いて, D 次元超球の体積を $V_D(r) = K_D r^D$, 表面積を $S_D(r) = D K_D r^{D-1}$ と表記した.
- 表面積は $S_D(r) = \dfrac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} r^{D-1}$ であったから, 係数 K_D は, 次のように求まる:

$$K_D = \frac{1}{D}\frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} = \frac{\pi^{\frac{D}{2}}}{\frac{D}{2}\Gamma\left(\frac{D}{2}\right)} = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} \quad (\because \Gamma(x+1) = x\Gamma(x))$$

- 係数 K_D が求まったので, D 次元超球の体積 $V_D(r) = K_D r^D$ は,

$$V_D(r) = K_D r^D = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} r^D$$

高次元空間の体積と表面積

- D 次元超球の体積 $V_D(r)$ は, $S_D(s)$ を 0 から r まで積分しても求まる:

$$V_D(r) = \int_0^r S_D(s)\,\mathrm{d}s = \int_0^r S_D s^{D-1}\,\mathrm{d}s = S_D \frac{1}{D}\left[s^D\right]_0^r = \frac{S_D}{D} r^D$$

- $S_D = \dfrac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$ を代入すれば ($\Gamma(x+1) = x\Gamma(x)$ に注意),

$$V_D(r) = \frac{S_D}{D} r^D = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \frac{1}{D} r^D = \frac{\pi^{\frac{D}{2}}}{\frac{D}{2}\Gamma\left(\frac{D}{2}\right)} r^D = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} r^D$$

- $r = 1$ とすれば, D 次元の単位超球の体積 V_D は,

$$V_D = \frac{S_D}{D} = K_D = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)}$$

高次元空間の体積と表面積

- D 次元超球の体積 $V_D(r)$ と表面積 $S_D(r)$ は,

$$V_D(r) = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} r^D, \quad S_D(r) = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} r^{D-1}$$

- $D = 2$ のとき (円), 円の面積 $V_2(r)$ と円周 $S_2(r)$ が得られる:

$$V_2(r) = \frac{\pi^{\frac{2}{2}}}{\Gamma(1+1)} r^2 = \pi r^2, \quad S_2(r) = \frac{2\pi^{\frac{2}{2}}}{\Gamma(1)} r^{2-1} = 2\pi r$$

- $D = 3$ のとき (球), 球の体積 $V_3(r)$ と表面積 $S_3(r)$ が得られる:

$$V_3(r) = \frac{\pi^{\frac{3}{2}}}{\Gamma\left(\frac{3}{2}+1\right)} r^3 = \frac{4}{3}\pi r^3, \quad S_3(r) = \frac{2\pi^{\frac{3}{2}}}{\Gamma\left(\frac{3}{2}\right)} r^{3-1} = 4\pi r^2$$

- $\Gamma(x+1) = x\Gamma(x)$, $\Gamma(1) = 1$, $\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$ に注意.

高次元空間の体積と表面積

D 次元超球の体積と表面積

半径 r の D 次元超球の体積 $V_D(r)$ と表面積 $S_D(r)$ は,

$$V_D(r) = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} r^D, \quad S_D(r) = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} r^{D-1}$$

D 次元単位超球の体積 V_D と表面積 S_D は,

$$V_D = V_D(1) = \frac{S_D}{D} = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)}, \quad S_D = S_D(1) = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$$

特に $D = 2, 3$ のとき,

$$V_2(r) = \pi r^2, \quad S_2(r) = 2\pi r, \quad V_3(r) = \frac{4}{3}\pi r^3, \quad S_3(r) = 4\pi r^2$$

高次元空間の形状

- D 次元の単位超球について, 半径 $1-\varepsilon$ から 1 までの間の体積の割合を考える:

$$\frac{V_D(1)-V_D(1-\varepsilon)}{V_D(1)} = \frac{K_D - K_D(1-\varepsilon)^D}{K_D} = 1-(1-\varepsilon)^D$$

- D が大きくなると, 小さな ε に対しても, 上記の割合が 1 に近づく.
- 高次元では, 球のほとんどの体積は, 表面に近い薄皮に集中する.
- 超球の中身はスカスカ.

高次元空間の形状

- 半径 r の D 次元超球と，その超球を包む，一辺 $2r$ の D 次元超立方体の体積比は，

$$\frac{V_D(r)}{(2r)^D} = \frac{\pi^{\frac{D}{2}} r^D}{\Gamma\left(\frac{D}{2}+1\right)} \cdot \frac{1}{(2r)^D} = \frac{\pi^{\frac{D}{2}}}{2^D \Gamma\left(\frac{D}{2}+1\right)} = \frac{\pi^{\frac{D}{2}}}{D 2^{D-1} \Gamma\left(\frac{D}{2}\right)}$$

- 両者の中心は同じで，超立方体は，各面の中心で超球と接する．
- スターリング (Stirling) の公式を使って，$D \to \infty$ での比を調べてみよう:

$$\Gamma(x) \simeq \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x = (2\pi)^{\frac{1}{2}} e^{-x} x^{x-\frac{1}{2}} \quad (x \gg 1)$$

$$\Gamma(x+1) = x\Gamma(x) \simeq (2\pi)^{\frac{1}{2}} e^{-x} x^{x+\frac{1}{2}} \quad (x \gg 1)$$

高次元空間の形状

- スターリングの公式を使うと, $D \gg 1$ のとき,

$$\frac{V_D(r)}{(2r)^D} = \frac{\pi^{\frac{D}{2}}}{2^D \Gamma\left(\frac{D}{2}+1\right)} \simeq \frac{\pi^{\frac{D}{2}}}{2^D (2\pi)^{\frac{1}{2}} e^{-\frac{D}{2}} \left(\frac{D}{2}\right)^{\frac{D}{2}+\frac{1}{2}}}$$
$$= 2^{-D}(2\pi)^{-\frac{1}{2}}(e\pi)^{\frac{D}{2}} \left(\frac{2}{D}\right)^{\frac{D}{2}+\frac{1}{2}}$$
$$= \left(\frac{2}{\pi^2 e}\right)^{\frac{1}{2}} \left(\frac{e\pi}{2D}\right)^{\frac{D}{2}+\frac{1}{2}} \to 0 \quad (D \to \infty)$$

- D 次元超立方体の頂点数: 2^D
- 高次元では, 超立方体に内接する超球の体積は, 超立方体の体積と比べて, 非常に小さい.
- 高次元では, 超立方体の体積は, 非常に多数の頂点に集中する.

高次元空間の形状

- D 次元超立方体 (一辺 $2r$) を考える.
- 中心から, 各頂点までの距離: $\sqrt{Dr^2} = r\sqrt{D}$
- 中心から, 側面までの距離: r
- $D \to \infty$ のとき, 両者の比率 $\sqrt{D}$ は発散する.

- 以上の結果をまとめると, 高次元空間では,
- 超立方体に内接する超球の体積は, 超立方体の体積と比べて, 非常に小さい.
- 超立方体の体積は, 非常に多数の頂点に集中する.
- 超立方体では, 中心から側面と比べて, 頂点までの距離が非常に大きい.
- 従って, 超立方体は, ウニのような尖った形であり, その先端に体積が集中している.

目次

4. 高次元空間の形状
   - 高次元空間でのガウス分布の形状

高次元空間でのガウス分布の形状

- 次の D 次元ガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2\mathbf{I})$ を考える:

$$\begin{aligned}\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2\mathbf{I}) &= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^\top\mathbf{x}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{D} x_i^2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \prod_{i=1}^{D} \exp\left(-\frac{x_i^2}{2\sigma^2}\right)\end{aligned}$$

- D 次元超球の表面積 $S_D(r)$ について, 次の関係が成り立つ:

$$\begin{aligned}\prod_{i=1}^{D}\int_{-\infty}^{\infty} \exp(-x_i^2)\,\mathrm{d}x_i &= \int_0^{\infty} \exp(-r^2) S_D(r)\,\mathrm{d}r \\ &= S_D \int_0^{\infty} \exp(-r^2) r^{D-1}\,\mathrm{d}r\end{aligned}$$

高次元空間でのガウス分布の形状

- ガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})$ を $\mathbf{x}$ について積分すれば,

$$
\begin{aligned}
\int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I}) \, \mathrm{d}\mathbf{x} &= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \prod_{i=1}^{D} \int_{-\infty}^{\infty} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) \mathrm{d}x_i \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \prod_{i=1}^{D} \int_{-\infty}^{\infty} \exp(-y_i^2) \sqrt{2}\sigma \, \mathrm{d}y_i \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} (\sqrt{2}\sigma)^D \prod_{i=1}^{D} \int_{-\infty}^{\infty} \exp(-y_i^2) \, \mathrm{d}y_i \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} (\sqrt{2}\sigma)^D S_D \int_{0}^{\infty} \exp(-s^2) s^{D-1} \, \mathrm{d}s
\end{aligned}
$$

- $y_i = \frac{1}{\sqrt{2}\sigma} x_i$ の置換を行った $\left(\frac{\mathrm{d}x_i}{\mathrm{d}y_i} = \sqrt{2}\sigma\right)$.

高次元空間でのガウス分布の形状

- 引き続き計算を進めると,

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})\,\mathrm{d}\mathbf{x} &= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} (\sqrt{2}\sigma)^D S_D \int_0^{\infty} \exp\left(-s^2\right) s^{D-1}\,\mathrm{d}s \\ &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) r^{D-1}\,\mathrm{d}r \\ &= \int_0^{\infty} p(r)\,\mathrm{d}r = 1\end{aligned}$$

- $r = \sqrt{2}\sigma s$ の置換を行った $\left(\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{1}{\sqrt{2}\sigma}\right)$.
- $\mathbf{x}$ の半径 $r = \|\mathbf{x}\|$ についての確率密度 $p(r)$ が得られた.

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

高次元空間でのガウス分布の形状

- $\mathbf{x}$ についてのガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2\mathbf{I})$ から，$\mathbf{x}$ の半径 $r = \|\mathbf{x}\|$ についての確率密度 $p(r)$ が得られた．

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^\top\mathbf{x}\right)$$
$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- $p(r)$ の停留点を求めるため，$p(r)$ を r で微分すると，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r}p(r) &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}}\left((D-1)r^{D-2} - r^{D-1}\frac{r}{\sigma^2}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right) \\ &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}}\exp\left(-\frac{r^2}{2\sigma^2}\right)\left((D-1)r^{D-2} - \frac{r^D}{\sigma^2}\right)\end{aligned}$$

高次元空間でのガウス分布の形状

- $\dfrac{\mathrm{d}}{\mathrm{d}r}p(r) = 0$ とおいて, $p(r)$ の停留点 $\hat{r}$ を求めると,

$$(D-1)r^{D-2} - \frac{r^D}{\sigma^2} = 0 \quad \longrightarrow \quad \hat{r} = \sigma\sqrt{D-1}$$

- $D \gg 1$ とすれば, $\hat{r} \simeq \sigma\sqrt{D}$ である.
- 停留点 $\hat{r}$ の近傍 $\hat{r} + \varepsilon$ における, 確率密度 $p(\hat{r} + \varepsilon)$ を考えよう.

$$\begin{aligned}
p(\hat{r}+\varepsilon) &= \frac{S_D(\hat{r}+\varepsilon)^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{(\hat{r}+\varepsilon)^2}{2\sigma^2}\right) \\
&= \frac{S_D\hat{r}^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{\hat{r}^2}{2\sigma^2}\right) \left(1+\frac{\varepsilon}{\hat{r}}\right)^{D-1} \exp\left(-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right) \\
&= p(\hat{r}) \left(1+\frac{\varepsilon}{\hat{r}}\right)^{D-1} \exp\left(-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right)
\end{aligned}$$

高次元空間でのガウス分布の形状

- 引き続き計算を進めると，

$$
\begin{aligned}
p(\hat{r}+\varepsilon) &= p(\hat{r})\left(1+\frac{\varepsilon}{\hat{r}}\right)^{D-1}\exp\left(-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right) \\
&= p(\hat{r})\exp\left((D-1)\ln\left(1+\frac{\varepsilon}{\hat{r}}\right)-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right) \\
&\simeq p(\hat{r})\exp\left((D-1)\left(\frac{\varepsilon}{\hat{r}}-\frac{1}{2}\frac{\varepsilon^2}{\hat{r}^2}\right)-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right) \\
&= p(\hat{r})\exp\left(\frac{\hat{r}^2}{\sigma^2}\left(\frac{\varepsilon}{\hat{r}}-\frac{1}{2}\frac{\varepsilon^2}{\hat{r}^2}\right)-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right) \\
&= p(\hat{r})\exp\left(-\frac{\varepsilon^2}{\sigma^2}\right)
\end{aligned}
$$

- $\ln(1+x)\simeq x-\frac{1}{2}x^2$ と，$\hat{r}=\sigma\sqrt{D-1}$ を使った．

高次元空間でのガウス分布の形状

- D 次の確率変数 $\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})$ に従うとき，半径 $r = \|\mathbf{x}\|$ の確率密度は，

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- この確率密度 $p(r)$ は，$\hat{r} = \sigma\sqrt{D-1} \simeq \sigma\sqrt{D}$ において最大値をとる．
- $\hat{r}$ の近傍では，次が成り立つ:

$$p(\hat{r} + \varepsilon) \simeq p(\hat{r}) \exp\left(-\frac{\varepsilon^2}{\sigma^2}\right)$$

- $\hat{r}$ の周囲では，確率密度が指数的に減衰する．$D \gg 1$ のとき $\sigma \ll \hat{r}$ だから，ほとんどの確率密度は，大きな半径 $\hat{r}$ の近くの，薄皮に集中している．

高次元空間でのガウス分布の形状

- $\mathbf{x}$ についてのガウス分布に立ち返って，原点 $\mathbf{x} = \mathbf{0}$ と，半径 $\hat{r}$ 上のある点 $\mathbf{x}$ ($\|\mathbf{x}\| = \hat{r}$) における確率密度 $p(\mathbf{x})$ を比較すると，

$$
\begin{aligned}
p(\mathbf{0}) &= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \\
p(\mathbf{x}) &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^\top\mathbf{x}\right) = \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2}\hat{r}^2\right) \\
&\simeq \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{D}{2}\right) \quad (\because \hat{r} \simeq \sigma\sqrt{D})
\end{aligned}
$$

- 原点 $\mathbf{x} = \mathbf{0}$ での確率密度は，半径 $\hat{r}$ 上の点の $\exp\left(\frac{D}{2}\right)$ 倍ある.
- しかし，ほとんどの確率密度は，半径 $\hat{r}$ の近くの薄皮に集中している.

高次元空間でのガウス分布の形状

- 続いて，半径の期待値を考えよう．半径に関する確率密度 $p(r)$ は，

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- 半径の期待値 $\mathbb{E}[r]$ は，

$$\begin{aligned}\mathbb{E}[r] = \int_0^\infty r p(r)\,\mathrm{d}r &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty r r^{D-1} \exp\left(-\frac{r^2}{2\sigma^2}\right) \mathrm{d}r \\ &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty r^D \exp\left(-\frac{r^2}{2\sigma^2}\right) \mathrm{d}r\end{aligned}$$

高次元空間でのガウス分布の形状

- 積分を行うと,

$$
\begin{aligned}
\mathbb{E}[r] &= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty r^D \exp\left(-\frac{r^2}{2\sigma^2}\right) \mathrm{d}r \\
&= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty (2\sigma^2 s)^{\frac{D}{2}} \exp(-s)(2s)^{-\frac{1}{2}} \sigma \,\mathrm{d}s \\
&= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} (2\sigma^2)^{\frac{D}{2}} 2^{-\frac{1}{2}} \sigma \int_0^\infty s^{\frac{D}{2}-\frac{1}{2}} \exp(-s) \,\mathrm{d}s \\
&= \frac{S_D}{\pi^{\frac{D}{2}}} 2^{-\frac{1}{2}} \sigma \Gamma\left(\frac{D}{2} + \frac{1}{2}\right)
\end{aligned}
$$

- $s = \dfrac{r^2}{2\sigma^2}$ の置換を行った $\left(r = (2\sigma^2 s)^{\frac{1}{2}}, \ \dfrac{\mathrm{d}r}{\mathrm{d}s} = (2s)^{-\frac{1}{2}}\sigma\right)$.

高次元空間でのガウス分布の形状

- D 次元単位超球の表面積 $S_D = \dfrac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$ を代入すれば,

$$\mathbb{E}\left[r\right] = \frac{S_D}{\pi^{\frac{D}{2}}} 2^{-\frac{1}{2}} \sigma \Gamma\left(\frac{D}{2} + \frac{1}{2}\right) = \sqrt{2}\sigma \frac{\Gamma\left(\frac{D}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{D}{2}\right)}$$

- 以上をまとめると, D 次の確率変数 $\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2\mathbf{I})$ に従うとき, 半径 $r = \|\mathbf{x}\|$ の確率密度は,

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- 半径 $r = \|\mathbf{x}\|$ の期待値 $\mathbb{E}\left[r\right]$ は,

$$\mathbb{E}\left[r\right] = \sqrt{2}\sigma \frac{\Gamma\left(\frac{D}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{D}{2}\right)}$$

高次元空間でのガウス分布の形状

- Gautschi の不等式 (Gautschi's Inequality):

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s} \quad (x > 0, 0 < s < 1)$$

- 期待値 $\mathbb{E}[r] = \sqrt{2}\sigma \frac{\Gamma\left(\frac{D}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{D}{2}\right)}$ に適用すれば $(x = \frac{D}{2} - \frac{1}{2}, s = \frac{1}{2})$,

$$\sqrt{2}\sigma\left(\frac{D}{2}-\frac{1}{2}\right)^{1-\frac{1}{2}} < \mathbb{E}[r] < \sqrt{2}\sigma\left(\frac{D}{2}-\frac{1}{2}+1\right)^{1-\frac{1}{2}}$$
$$\longrightarrow \sigma\sqrt{D-1} < \mathbb{E}[r] < \sigma\sqrt{D+1}$$

- 実際には, 次の不等式が得られる:

$$\sigma\frac{D}{\sqrt{D+1}} < \mathbb{E}[r] < \sigma\sqrt{D}$$

高次元空間でのガウス分布の形状

- 期待値 $\mathbb{E}[r]$ を少し書き換えてみると,

$$\mathbb{E}[r] = \sqrt{2}\sigma \frac{\Gamma\left(\frac{D}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{D}{2}\right)} = \sqrt{2}\sigma \frac{\frac{2}{D+1}\Gamma\left(\frac{D+1}{2}+1\right)}{\frac{2}{D}\Gamma\left(\frac{D}{2}+1\right)} = \frac{\sqrt{2}\sigma D}{D+1}\frac{\Gamma\left(\frac{D+1}{2}\right)}{\Gamma\left(\frac{D}{2}+1\right)}$$

- Gautschi の不等式を適用すれば $(x = \dfrac{D}{2}+\dfrac{1}{2},\ s = \dfrac{1}{2})$,

$$\frac{\sqrt{2}\sigma D}{D+1}\left(\frac{D+1}{2}\right)^{\frac{1}{2}} < \mathbb{E}[r] \quad \longrightarrow \quad \sigma\frac{D}{\sqrt{D+1}} < \mathbb{E}[r]$$

- これは $\sigma\sqrt{D-1}$ よりも良い下界となる.

高次元空間でのガウス分布の形状

- 続いて，半径の分散を考えよう．2 次モーメント $\mathbb{E}\left[r^2\right]$ は，

$$
\begin{aligned}
\mathbb{E}\left[r^2\right] &= \int_0^\infty r^2 p(r)\,\mathrm{d}r = \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty r^{D+1} \exp\left(-\frac{r^2}{2\sigma^2}\right) \mathrm{d}r \\
&= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} \int_0^\infty (2\sigma^2 s)^{\frac{D}{2}+\frac{1}{2}} \exp(-s)(2s)^{-\frac{1}{2}}\sigma\,\mathrm{d}s \\
&= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} (2\sigma^2)^{\frac{D}{2}+\frac{1}{2}} 2^{-\frac{1}{2}}\sigma \int_0^\infty s^{\frac{D}{2}} \exp(-s)\,\mathrm{d}s \\
&= \frac{S_D}{(2\pi\sigma^2)^{\frac{D}{2}}} (2\sigma^2)^{\frac{D}{2}+\frac{1}{2}} 2^{-\frac{1}{2}}\sigma\Gamma\left(\frac{D}{2}+1\right)
\end{aligned}
$$

- $s = \dfrac{r^2}{2\sigma^2}$ の置換を行った ($r = (2\sigma^2 s)^{\frac{1}{2}}$, $\dfrac{\mathrm{d}r}{\mathrm{d}s} = (2s)^{-\frac{1}{2}}\sigma$).

高次元空間でのガウス分布の形状

- D 次元単位超球の表面積 $S_D = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$ を代入すれば，

$$\mathbb{E}\left[r^2\right] = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \frac{(2\sigma^2)^{\frac{D}{2}+\frac{1}{2}} 2^{-\frac{1}{2}} \sigma}{(2\pi\sigma^2)^{\frac{D}{2}}} \Gamma\left(\frac{D}{2}+1\right)$$
$$= 2\sigma^2 \frac{\Gamma\left(\frac{D}{2}+1\right)}{\Gamma\left(\frac{D}{2}\right)} = 2\sigma^2 \frac{\frac{D}{2}\Gamma\left(\frac{D}{2}\right)}{\Gamma\left(\frac{D}{2}\right)} = \sigma^2 D$$

- 分散 $\mathrm{Var}\left[r\right] = \mathbb{E}\left[r^2\right] - \mathbb{E}\left[r\right]^2$ は 0 以上だから，

$$\mathrm{Var}\left[r\right] = \sigma^2 D - \mathbb{E}\left[r\right]^2 \geq 0 \quad \longrightarrow \quad \mathbb{E}\left[r\right] \leq \sigma\sqrt{D}$$

高次元空間でのガウス分布の形状

- D 次の確率変数 $\mathbf{x}$ がガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})$ に従うとき，半径 $r = \|\mathbf{x}\|$ の期待値は，

$$\mathbb{E}\left[r\right] = \sqrt{2}\sigma \frac{\Gamma\left(\frac{D}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{D}{2}\right)}, \quad \sigma \frac{D}{\sqrt{D+1}} < \mathbb{E}\left[r\right] \leq \sigma\sqrt{D}$$

- 高次元 $(D \gg 1)$ であれば，$\mathbb{E}\left[r\right] \simeq \sigma\sqrt{D}$ とできる．
- 先ほどみたように，半径についての確率密度 $p(r)$ は，$r = \sigma\sqrt{D-1} \simeq \sigma\sqrt{D}$ において最大値をとる．
- $p(r)$ のほとんどの確率密度は，半径 $\sigma\sqrt{D}$ の超球面上に集中している．

- 高次元空間でのガウス分布は，半径 $\sigma\sqrt{D}$ の超球面上での一様分布とみなせる．

高次元空間でのガウス分布の形状

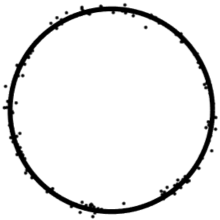

Figure 3.6 A Gaussian point cloud in two dimensions (left) and its intuitive visualization in high dimensions (right). In high dimensions, the standard normal distribution is very close to the uniform distribution on the sphere of radius $\sqrt{n}$.

図 1: 高次元空間におけるガウス分布 (右側)[1]

[1] https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf

レイリー分布 (Rayleigh Distribution)

- $\mathbf{x}$ が D 次元のガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})$ に従うとき，半径 r は次の確率分布に従う：

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- $D = 2$ としたときの分布を，**レイリー分布**という：

$$p(r) = \frac{S_2 r^{2-1}}{(2\pi\sigma)^{\frac{2}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right) = \frac{2\pi r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

- 単位円の円周 $S_2 = 2\pi$ を用いた．
- 2 次元のガウス分布 $\mathcal{N}(\mathbf{x} \mid \mathbf{0}, \sigma^2 \mathbf{I})$ の変数変換 (極座標表示) によっても得られる (練習問題)．