正規分布

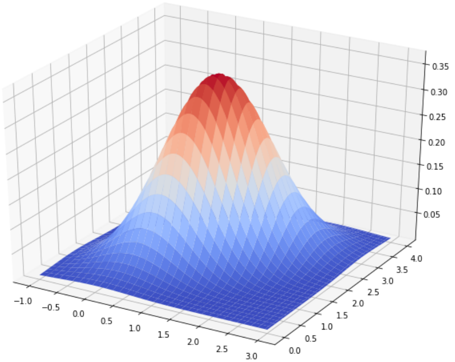
正田 備也

masada@rikkyo.ac.jp

Contents

正規分布 normal distribution

- ガウス分布 Gaussian distribution とも呼ばれる
- 連続量をモデル化するときによく使われる
  - 下図は 2 変量正規分布の確率密度関数の例



単変量正規分布 univariate normal distribution

- 単変量正規分布は、$(-\infty, \infty)$ 上に定義される
- 単変量正規分布のパラメータは、平均 μ と標準偏差 σ
  - 平均 μ、標準偏差 σ の単変量正規分布を、以下 $\mathcal{N}(\mu, \sigma^2)$ と書く
  - 確率変数 x が $\mathcal{N}(\mu, \sigma^2)$ に従うことを、以下 $x \sim \mathcal{N}(\mu, \sigma^2)$ と書く
- 単変量正規分布 $\mathcal{N}(\mu, \sigma^2)$ の確率密度関数 (pdf) は

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \tag{1}$$

  - 「;」は、その右側にある μ と σ が自由パラメータ（我々がその値を指定してはじめて分布が定まるパラメータ）であることを意味する

参考：ガウス積分

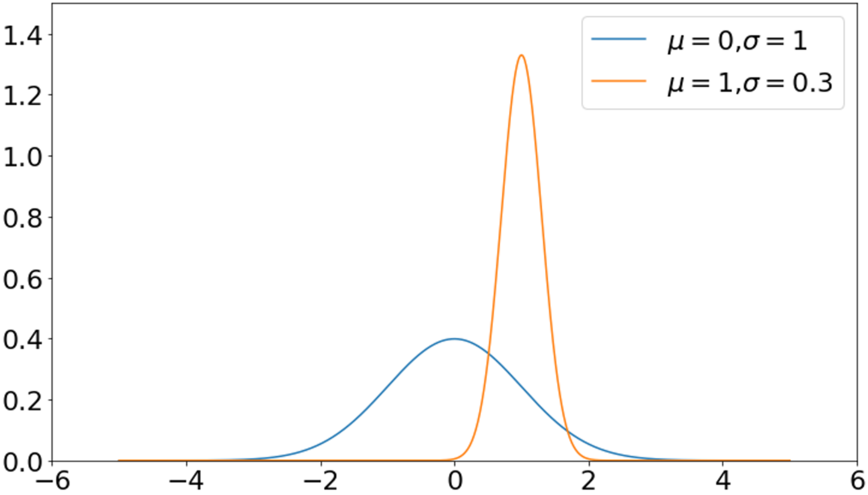
- $\sqrt{2\pi\sigma^2}$ って、何？
  - $-\infty$ から ∞ までの範囲で積分すると 1 になるようにするには、この値で割っておかないといけない、という意味
  - ひとことで言えば、規格化定数 (normalizing constant)
- つまり $\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}$ が成り立つ、ということ
  - こういう積分を、ガウス積分と呼ぶ
  - ガウス積分の公式 $\int_{-\infty}^{\infty} e^{-a(x+b)^2} dx = \sqrt{\frac{\pi}{a}}$

密度関数についての注意事項（再）

$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2}$$

- 式 (2) の x に特定の値を代入して得られる $p(x;\mu,\sigma)$ の値は、確率としての意味を持たない
  - そもそも、密度関数の値は 1 を超えることがいくらでもある
- 密度関数を特定の範囲で積分すると、確率とみなすことのできる値が得られる
  - 当然、全範囲で積分すると 1 になる（つまり $\int_{-\infty}^{\infty} p(x;\mu,\sigma)dx = 1$）

単変量正規分布の密度関数の例



標準正規分布 standard normal distribution

- 平均が 0 で標準偏差が 1 の単変数正規分布 $\mathcal{N}(0,1)$ のことを、標準正規分布と呼ぶ
- 標準正規分布に従う確率変数を x とする
- このとき、$y=\sigma x+\mu$ は、平均が μ で標準偏差 σ の正規分布 $\mathcal{N}(\mu,\sigma^2)$ にしたがう
  - つまり、 $x\sim\mathcal{N}(0,1)$ ならば $\sigma x+\mu\sim\mathcal{N}(\mu,\sigma^2)$
  - 一般に、正規分布にしたがう確率変数のアフィン変換はまた、正規分布にしたがう

正規分布に独立にしたがう確率変数の和

- 確率変数 x と y が、それぞれ独立に正規分布 $\mathcal{N}(\mu_x, \sigma_x^2)$ と $\mathcal{N}(\mu_y, \sigma_y^2)$ にしたがうとする
- このとき、$x + y$ も正規分布にしたがうことを、以下、示す
  - $\mathcal{N}(\mu_x, \sigma_x^2)$ と $\mathcal{N}(\mu_y, \sigma_y^2)$ の密度関数を、それぞれ $p(x; \mu_x, \sigma_x)$ と $p(y; \mu_y, \sigma_y)$ と書く
  - x と y は独立なので、同時分布の確率密度関数 $p(x, y)$ は $p(x; \mu_x, \sigma_x)p(y; \mu_y, \sigma_y)$ と積で書ける
  - $z = x + y$ とおくと、$y = z - x$
  - y を $z - x$ で置き換えて、x について積分する

cf. https://faculty.math.illinois.edu/~r-ash/Stat/StatLec1-5.pdf の Sec. 3.11

$$
\begin{aligned}
&\int_{-\infty}^{\infty} p(x;\mu_x,\sigma_x)p(z-x;\mu_y,\sigma_y)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{\{(z-x)-\mu_y\}^2}{2\sigma_y^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_y^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{\{(z-x)-\mu_y\}^2}{2\sigma_y^2}\right) dx
\end{aligned}
\tag{3}
$$

ここで、$\exp()$ の中身を x について平方完成すると

$$
\begin{aligned}
&-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{\{(z-x)-\mu_y\}^2}{2\sigma_y^2} = -\frac{\sigma_y^2(x-\mu_x)^2 + \sigma_x^2\{x-(z-\mu_y)\}^2}{2\sigma_x^2\sigma_y^2} \\
&= -\frac{(\sigma_x^2+\sigma_y^2)x^2 - 2\{\sigma_y^2\mu_x + \sigma_x^2(z-\mu_y)\}x + \sigma_y^2\mu_x^2 + \sigma_x^2(z-\mu_y)^2}{2\sigma_x^2\sigma_y^2}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
&-\frac{(\sigma_x^2+\sigma_y^2)x^2-2\{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)\}x+\sigma_y^2\mu_x^2+\sigma_x^2(z-\mu_y)^2}{2\sigma_x^2\sigma_y^2} \\
&=-\frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left\{x^2-2\frac{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)}{\sigma_x^2+\sigma_y^2}x\right\}-\frac{\sigma_y^2\mu_x^2+\sigma_x^2(z-\mu_y)^2}{2\sigma_x^2\sigma_y^2} \\
&=-\frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left\{x-\frac{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)}{\sigma_x^2+\sigma_y^2}\right\}^2+\frac{\{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)\}^2}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)}-\frac{\sigma_y^2\mu_x^2+\sigma_x^2(z-\mu_y)^2}{2\sigma_x^2\sigma_y^2}
\end{aligned}
\tag{5}
$$

x を含まない後半の 2 つの項に注目して z について平方完成すると

$$
\begin{aligned}
&\frac{\{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)\}^2}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)}-\frac{\sigma_y^2\mu_x^2+\sigma_x^2(z-\mu_y)^2}{2\sigma_x^2\sigma_y^2} \\
&=\frac{\sigma_y^4\mu_x^2+2\sigma_x^2\sigma_y^2\mu_x(z-\mu_y)+\sigma_x^4(z-\mu_y)^2-(\sigma_x^2+\sigma_y^2)\{\sigma_y^2\mu_x^2+\sigma_x^2(z-\mu_y)^2\}}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)} \\
&=\frac{2\mu_x(z-\mu_y)-\mu_x^2-(z-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}=-\frac{1}{2(\sigma_x^2+\sigma_y^2)}\{z-(\mu_x+\mu_y)\}^2
\end{aligned}
\tag{6}
$$

元に戻ると

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma_x^2}}\frac{1}{\sqrt{2\pi\sigma_y^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}-\frac{\{(z-x)-\mu_y\}^2}{2\sigma_y^2}\right)dx \\
&=\frac{1}{\sqrt{2\pi\sigma_x^2}}\frac{1}{\sqrt{2\pi\sigma_y^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left\{x-\frac{\sigma_y^2\mu_x+\sigma_x^2(z-\mu_y)}{\sigma_x^2+\sigma_y^2}\right\}^2-\frac{\{z-(\mu_x+\mu_y)\}^2}{2(\sigma_x^2+\sigma_y^2)}\right]dx \\
&=\frac{1}{\sqrt{2\pi\sigma_x^2}}\frac{1}{\sqrt{2\pi\sigma_y^2}}\exp\left[-\frac{\{z-(\mu_x+\mu_y)\}^2}{2(\sigma_x^2+\sigma_y^2)}\right]\sqrt{\pi\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}} \\
&=\frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}}\exp\left[-\frac{\{z-(\mu_x+\mu_y)\}^2}{2(\sigma_x^2+\sigma_y^2)}\right] \qquad (7)
\end{aligned}
$$

ただし、途中でガウス積分 $\int_{-\infty}^{\infty} e^{-a(x+b)^2}=\sqrt{\frac{\pi}{a}}$ を使った。
以上より、$z\sim\mathcal{N}(\mu_x+\mu_y,\sigma_x^2+\sigma_y^2)$ となることを示せた。

参考：変数変換

http://nlp.dse.ibaraki.ac.jp/~shinnou/siryou/toukei-kentei/1-stat-var-trans.pdf

- 1次元の場合
  - 確率変数 X の確率密度関数を $f_X(x)$ とする
  - X を $Y = g(X)$ と変数変換するとする
  - 確率変数 Y の密度関数 $f_Y(y)$ は $f_Y(y) = f_X(x)\frac{dx}{dy}$ となる
- 2次元の場合
  - 確率変数 X と Y の同時密度関数を $f_{X,Y}(x,y)$ とする
  - X と Y を $(U,V) = (g_1(X,Y), g_2(X,Y))$ と変数変換するとする
  - U と V の同時密度関数は $f_{U,V}(u,v) = f_{X,Y}(x,y)\|J\|$ となる
    - J はヤコビアン $J = \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial u} & \frac{\partial g_1^{-1}}{\partial v} \\ \frac{\partial g_2^{-1}}{\partial u} & \frac{\partial g_2^{-1}}{\partial v} \end{bmatrix}$
    - $\|J\|$ はヤコビアン J の行列式の絶対値

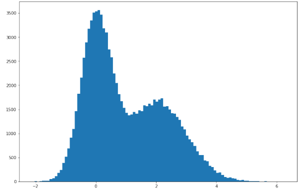
（$U = X, V = X + Y$ の場合が、確率変数の和の場合に相当する）

確率変数の和がしたがう確率分布

- 正規分布の密度関数は、ラクダのコブのような形をしている
- ところで、たった今、異なる正規分布にしたがう独立な２つの確率変数の和が正規分布にしたがうことを示した
- しかし、独立な２つの確率変数が、いずれもコブ状の密度関数をもつ分布にしたがうなら、それら確率変数の和がしたがう確率分布の密度関数には、２つのコブがあるのでは？
- …これはよくある勘違い。確率変数の足し算を考えることと、密度関数の足し算を考えることとは、全く別のこと
- では、２つのコブがある分布はどのような場合に作られる？

混合分布

- 以下のようにして生成される観測データはどのような分布にしたがうか？
  - まずコインを投げる
  - 表が出たら $\mathcal{N}(\mu_x, \sigma_x^2)$ から値を生成
  - 裏が出たら $\mathcal{N}(\mu_y, \sigma_y^2)$ から値を生成



- このようにして生成される実数値がしたがう分布の密度関数には、コブが2つある
- このような分布を混合分布というが、詳細はまたいずれ

正規分布の再生性

- 再生性 reproductive property
  - 同じ分布だがパラメータが違うだけの分布に従う確率変数の和もまた、それら 2 つの分布とパラメータが違うだけの同じ分布に従うとき、この分布族は再生性を持つ、という。
- M 個の確率変数 $x_1, \ldots, x_M$ が、それぞれ独立に正規分布 $\mathcal{N}(\mu_1, \sigma_1^2), \ldots, \mathcal{N}(\mu_M, \sigma_M^2)$ にしたがうとする
- このとき、$\sum_{m=1}^{M} a_m x_m$ は、以下の正規分布にしたがう

$$\mathcal{N}\left(\sum_{m=1}^{M} a_m \mu_m, \sum_{m=1}^{M} a_m^2 \sigma_m^2\right) \tag{8}$$

Contents

単変量正規分布に従う観測データの尤度

- 与えられている観測データを $\mathcal{D} = \{x_1, \ldots, x_N\}$ とする
  - 各 x_i は、$-\infty < x_i < \infty$ を満たす実数値とする
- 各 x_i は同じ正規分布 $\mathcal{N}(\mu, \sigma^2)$ に独立にしたがうものとして、この観測データをモデル化することにする
- このときデータセット $\mathcal{D}$ の尤度は、以下のような μ と σ の関数になる

$$p(\mathcal{D}; \mu, \sigma) = \prod_{i=1}^{N} p(x_i; \mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x_i - \mu)^2}{2\sigma^2} \right) \tag{9}$$

単変量正規分布の最尤推定

- 観測データ $\mathcal{D}$ の対数尤度は以下のようになる

$$\ln p(\mathcal{D}; \mu, \sigma) = -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2} \tag{10}$$

- この対数尤度を最大化する μ と σ を、以下、求める

$$\frac{\partial \ln p(\mathcal{D}; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{N} \frac{x_i - \mu}{\sigma^2} \tag{11}$$

$\frac{\partial \ln p(\mathcal{D};\mu,\sigma)}{\partial \mu} = 0$ より $\mu = \frac{\sum_i x_i}{N} = \bar{x}$

$$\frac{\partial \ln p(\mathcal{D}; \mu, \sigma)}{\partial \sigma} = -\frac{N}{\sigma} + \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma^3} \tag{12}$$

$\frac{\partial \ln p(\mathcal{D};\mu,\sigma)}{\partial \sigma} = 0$ より $\sigma^2 = \frac{\sum_i (x_i-\mu)^2}{N} = \frac{\sum_i (x_i-\bar{x})^2}{N}$

Contents

二変量正規分布 bivariate normal distribution

- パラメータ
  - 平均ベクトル $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$
  - 共分散行列 $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ （ただし Σ は正定値行列）
- 確率密度関数（ただし $|\boldsymbol{\Sigma}| \equiv \mathrm{det}\boldsymbol{\Sigma}$）

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^2 |\boldsymbol{\Sigma}|}} \exp\left\{ -\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{2} \right\} \tag{13}$$

多変量正規分布

- パラメータ
  - 平均ベクトル $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix}$
  - 共分散行列 $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1d} \\ \vdots & \ddots & \vdots \\ \sigma_{1d} & \cdots & \sigma_{dd} \end{bmatrix}$ （ただし Σ は正定値行列）
- 確率密度関数（ただし $|\boldsymbol{\Sigma}| \equiv \mathrm{det}\boldsymbol{\Sigma}$）

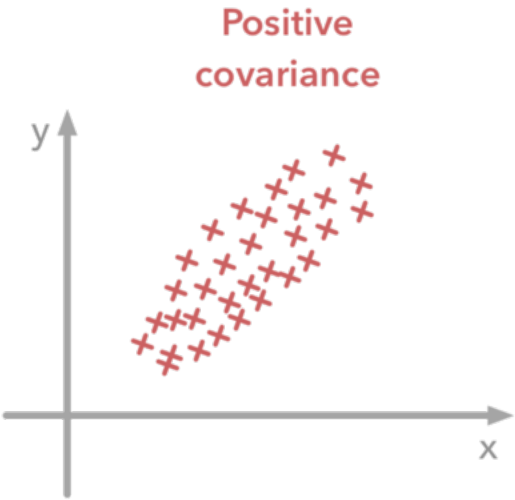
$$p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left\{ -\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{2} \right\} \tag{14}$$

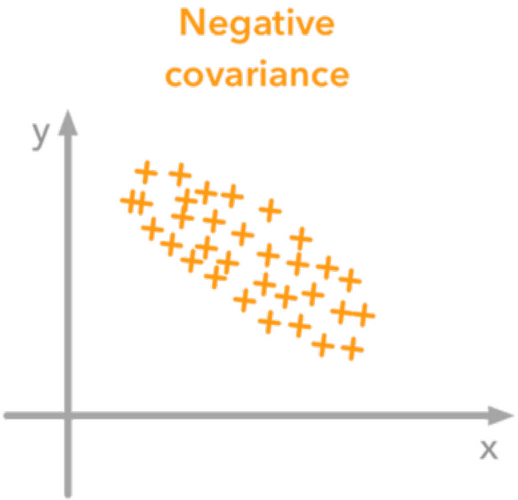
共分散行列 covariance matrix

- 二変量以上の場合、2 つ以上成分をもつベクトルをモデル化
- ベクトルの第 1 成分、第 2 成分、等々について、各々単独での散らばり方は、当然考慮する
  - これが、共分散行列の対角成分、つまり分散に対応する
- これに加えて、第 1 成分と第 2 成分、第 1 成分と第 3 成分など、異なる成分間の関連も考慮する
  - これが、共分散行列の非対角成分、つまり共分散に対応する
- 共分散行列は分散共分散行列とも呼ばれる

共分散 covariance の直感的な意味

- 共分散は共分散行列の非対角成分に現れている値
- ゼロだと、対応する二つの成分は独立に分布
- 正の値だと、一方の成分が平均より大きいとき、他方の成分も平均より大きくなることが多い
- 負の値だと、一方の成分が平均より大きいとき、他方の成分は平均より小さくなることが多い





https://www.kdnuggets.com/2018/10/preprocessing-deep-learning-covariance-matrix-image-whitening.html

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$$

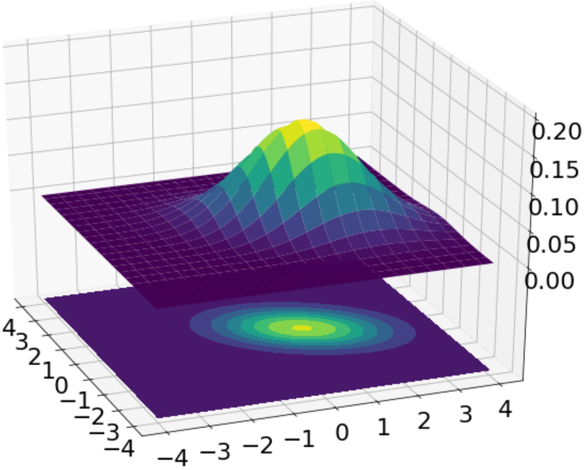


Figure: 二変量正規分布の密度関数の例

問題4-1

- $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ の行列式 $\det\boldsymbol{\Sigma}$ を求めよ。
- $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ の逆行列 $\boldsymbol{\Sigma}^{-1}$ を求めよ。

二変量正規分布の別のパラメータ化(1/2)

- 同じ確率分布でも、パラメータ化 parameterization の方法が複数あることがある
- 二変量正規分布では、共分散行列を以下のようにパラメータ化することがある

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{15}$$

二変量正規分布の別のパラメータ化(2/2)

- このパラメータ化のもとでは、$\boldsymbol{\Sigma}$ の行列式と逆行列は

$$\mathrm{det}\boldsymbol{\Sigma} = \sigma_1^2\sigma_2^2(1-\rho^2) \tag{16}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\mathrm{det}\boldsymbol{\Sigma}}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \tag{17}$$

Contents

多変量正規分布の最尤推定

- 観測データ $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ の各 $\boldsymbol{x}_i$ が独立に同じ正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ にしたがうと仮定すると、$\mathcal{D}$ の尤度は

$$
\begin{aligned}
p(\mathcal{D}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{N} p(\boldsymbol{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right) \quad (18)
\end{aligned}
$$

- 最尤推定によって各パラメータを推定（次スライドから）

  cf. The Matrix Cookbook

行列やベクトルの偏微分の基本を再確認

$$\frac{\partial x_{kl}}{\partial x_{ij}} = \delta_{ik}\delta_{lj} \tag{19}$$

ただし、$i = k$ ならば $\delta_{ik} = 1$、そうでなければ $\delta_{ik} = 0$

$$\left[\frac{\partial \boldsymbol{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \tag{20}$$

$$\left[\frac{\partial x}{\partial \boldsymbol{y}}\right]_i = \frac{\partial x}{\partial y_i} \tag{21}$$

$$\left[\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j} \tag{22}$$

$$
\begin{aligned}
\ln p(\mathcal{D}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{i=1}^{N} \ln p(\boldsymbol{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= -\frac{Nd}{2} \ln(2\pi) - \frac{N}{2} \ln(|\boldsymbol{\Sigma}|) - \frac{1}{2} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \qquad (23)
\end{aligned}
$$

ここで、$\bar{\boldsymbol{x}} = \frac{\sum_i \boldsymbol{x}_i}{N}$ として、

$$
\sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) = \sum_i \boldsymbol{x}_i^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_i - 2N \bar{\boldsymbol{x}}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + N \boldsymbol{\nu}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad (24)
$$

より、

$$
\frac{\partial \ln p(\mathcal{D}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -N \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{x}} + N \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad (25)
$$

$$
\therefore \boldsymbol{\mu} = \bar{\boldsymbol{x}} \qquad (26)
$$

$\frac{\partial \ln p(\mathcal{D};\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}}$ の計算は、The Matrix Cookbook を見ながらおこなう。

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} \ln(|\boldsymbol{\Sigma}|) = \boldsymbol{\Sigma}^{-1} \tag{27}$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} (\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) = -\boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} \tag{28}$$

以上より、

$$\begin{aligned} \frac{\partial \ln p(\mathcal{D};\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} &= -\frac{N}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Sigma}^{-1} \\ &= -\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left\{N - \left(\sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Sigma}^{-1}\right\} \end{aligned} \tag{29}$$

$$\therefore \boldsymbol{\Sigma} = \frac{1}{N}\sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \tag{30}$$

Contents

https://www.amazon.co.jp/dp/4061529080

異常検知への応用：ホテリングの T^2 法の概要

- 異常値を含まない（あるいはほとんど含まない）データ集合を使って多変量正規分布のパラメータを最尤推定する
- 異常値かどうかを調べたいデータについて、最尤推定の結果を使って標本平均とのマハラノビス距離を求める
- マハラノビス距離が、あらかじめ計算しておいた閾値を超えたとき、警報を出す
- 解説付きの Python 実装例
  - https://gochikika.ntt.com/Application/anomaly_density_basic.html

観測データの異常度

- 最尤推定で求めた平均ベクトルを $\hat{\boldsymbol{\mu}}$、共分散行列を $\hat{\boldsymbol{\Sigma}}$ とする
- この推定値を使うと正規分布の密度関数は

$$p(\boldsymbol{x};\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) = \frac{1}{\sqrt{(2\pi)^d|\hat{\boldsymbol{\Sigma}}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{\mu}})^\intercal\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}-\hat{\boldsymbol{\mu}})\right\} \tag{31}$$

- そして、観測データ $\boldsymbol{x}$ の異常度を $-\ln p(\boldsymbol{x};\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})$ と定義する
  - 直感的には、推定された平均ベクトル $\hat{\boldsymbol{\mu}}$ から遠いほど異常
  - ただし、ユークリッド距離を使っているのではない

マハラノビス距離 Mahalanobis distance

- $\boldsymbol{x}$ と $\hat{\boldsymbol{\mu}}$ との間のマハラノビス距離 $a(\boldsymbol{x})$ は、以下のように定義される

$$a(\boldsymbol{x}) = \sqrt{(\boldsymbol{x} - \hat{\boldsymbol{\mu}})^\intercal \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{\mu}})} \tag{32}$$

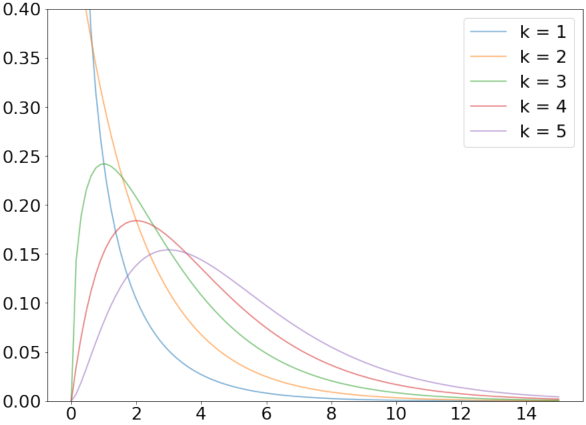
- マハラノビス距離の大小関係は、異常度 $-\ln p(\boldsymbol{x}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ の大小関係に、一致する
- そこで、$a(\boldsymbol{x})$ が大きいほど、$\boldsymbol{x}$ はより一層異常だとみなすことにする
  - $\hat{\Sigma}$ を使っているので、観測データのベクトルの成分間の関連も考慮した距離になっている

ホテリングのT^2法

- 同じ d 次元正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ からの N 個の独立なサンプルだと仮定された観測データに基づき、最尤推定により平均ベクトルを $\hat{\boldsymbol{\mu}}$、共分散行列を $\hat{\boldsymbol{\Sigma}}$ と、それぞれ推定したとする
- このとき、$T^2 \equiv \frac{N-d}{(N+1)d} a(\boldsymbol{x})^2$ により定義される統計量 T^2 は、自由度 $(d, N-d)$ の F 分布にしたがう
- $N \gg d$ の場合、$a(\boldsymbol{x})^2 \equiv (\boldsymbol{x} - \hat{\boldsymbol{\mu}})^\intercal \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{\mu}})$ は近似的に自由度 d のカイ 2 乗分布にしたがう

カイ２乗分布

- 独立に標準正規分布にしたがう k 個の確率変数の二乗和がしたがう分布を、自由度 k のカイ２乗分布とよぶ

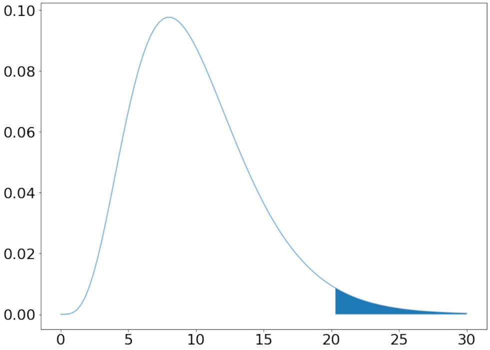


閾値の決め方

- あらかじめ誤報率 α を決めておく
- 下の等式を満たす a_0 を閾値として設定する

$$1 - \alpha = \int_0^{a_0} \chi^2(x; d)dx \tag{33}$$

  - ただし、$\chi^2(x; d)$ は自由度 d のカイ 2 乗分布の確率密度関数とする



- $a(\boldsymbol{x})^2$ と a_0 を比較して異常を検知

注意：共分散行列の推定

- マハラノビス距離の計算には、データ集合から推定された共分散行列を使う。
- この推定に最尤推定を使うのがベストとは限らない。
- 特に、データ集合に初めから外れ値が多く含まれる場合は、最尤推定以外の方法を使うほうが良い場合が多い。
- 例えば、scikit-learn の Web サイトの下記ページを参照。
    - https://scikit-learn.org/stable/auto_examples/covariance/plot_mahalanobis_distances.html

課題: 二変量正規分布に従う乱数を発生させる

- 二変量の正規分布を適当に選ぼう。
  - 平均パラメータや共分散行列を適当に決める、ということ。
- 上で決めた正規分布に従う乱数を発生させる Python のコードを書こう。
  - ただし、まず標準正規分布に従う乱数を発生させて、それを演算によって加工することで、乱数を作成すること。
- 発生させた乱数（たくさんの 2 次元ベクトル）の散布図を描き、だいたい思った通りの散らばり方をしているか、確認しよう。