

# Gauss 関数の最尤推定

2024 年 1 月 25 日

## 1 確率密度関数と最尤推定

ある確率変数  $x$  がパラメーター  $\theta$  を持つ関数  $f(x; \theta)$  に従う時、この関数を確率密度関数という。

$$x \sim P(x; \theta) \quad (1)$$

ここで、互いに独立な  $N$  個の要素を持つ変数  $x = (x_0, x_1, x_2, \dots, x_n)$  を用意し、各々の値に関する確率密度関数の値の積を  $\theta$  の関数とすると、これは  $\theta$  の尤度と呼ぶ。これは  $\theta$  の尤もらしさを示す。

$$L(\theta) = \prod_n^N P(x_n; \theta) \quad (2)$$

この尤度も最大にするようなパラメーター  $\theta$  の探索を最尤推定と呼ぶ。具体的には、あるパラメーター不明の確率分布に従うデータを得た時、最もデータをよく表すパラメーターを探索することである。

通常は簡単のため尤度を対数で表した対数尤度を用いる。

$$\log_e L(\theta) = \log_e \prod_n^N P(x_n; \theta) = \sum_n^N \log_e P(x_n; \theta) \quad (3)$$

## Gauss 関数の最尤推定

正規分布の確率関数は Gauss 関数であり、以下である。

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (4)$$

個数  $N$  のデータ  $x = (x_0, x_1, x_2, \dots, x_n)$  を用意すると尤度関数は以下となる。

$$L(x|\theta) = \sum_{n=0}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_n - \mu)^2}{2\sigma^2}\right] \quad (5)$$

Gauss 関数には 2 つのパラメーターが存在するため、2 回に分けて各々のパラメーターを決定する必要がある。

## 1.1 パラメーター $\mu$ についての最尤推定

(5) 式を  $\mu$  で偏微分するのだが、それに先んじて簡単のため両辺に対数を取って式を変形する.

$$\log_e L(x|\theta) = \log_e \sum_{n=0}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_n - \mu)^2}{2\sigma^2}\right] \quad (6)$$

$$= \sum_{n=0}^N \log_e \frac{1}{\sqrt{2\pi\sigma^2}} + \log_e e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} \quad (7)$$

$$= \sum_{n=0}^N -\frac{1}{2} \log_e(2\pi\sigma^2) - \frac{(x_n - \mu)^2}{2\sigma^2} \quad (8)$$

$$= -\frac{N}{2} \log_e(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^N (x_n - \mu)^2 \quad (9)$$

(9) 式を  $\sigma$  で偏微分する.

$$\frac{\partial}{\partial \sigma} L(x|\theta) = -\frac{1}{\sigma} \sum_{n=0}^N (x_n - \mu) \quad (10)$$

最後に (10) 式を 0 とし,  $\mu$  について解く.

$$-\frac{1}{\sigma} \sum_{n=0}^N (x_n - \mu) = 0 \quad (11)$$

$$-N\mu + \sum_{n=0}^N x_n = 0 \quad (12)$$

$$\mu = \frac{1}{N} \sum_{n=0}^N x_n \equiv \bar{x} \quad (13)$$

(13) をみると Gauss 関数のパラメーター  $\mu$  はデータ  $x$  の平均に等しいことがわかる. したがって  $\mu$  はデータの平均である.

## 1.2 パラメーター $\sigma$ についての最尤推定

$\mu$  の場合と同様にして、(9) 式を  $\sigma$  で偏微分する。対数に注意し、その引数を  $u$  とする。(合成関数の微分)

$$\frac{\partial}{\partial \sigma} \log_e L(x|\theta) = -\frac{N}{2} \frac{d}{du} \log_e u \frac{d}{d\sigma} 2\pi\sigma^2 - \frac{\partial}{\partial \sigma} \frac{1}{2\sigma^2} \sum_{n=0}^N (x_n - \mu)^2 \quad (14)$$

$$= -\frac{N}{2} \cdot \frac{1}{u} \cdot 4\pi\sigma + \frac{1}{\sigma^3} \sum_{n=0}^N (x_n - \mu)^2 \quad (15)$$

$$= -\frac{N}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 4\pi\sigma + \frac{1}{\sigma^3} \sum_{n=0}^N (x_n - \mu)^2 \quad (16)$$

$$= -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{n=0}^N (x_n - \mu)^2 \quad (17)$$

$$= \frac{1}{\sigma} \left\{ -N + \frac{1}{\sigma^2} \sum_{n=0}^N (x_n - \mu)^2 \right\} \quad (18)$$

最後に (18) 式を 0 とし、 $\sigma$  について解く。

$$-\frac{1}{\sigma} \left\{ -N + \frac{1}{\sigma^2} \sum_{n=0}^N (x_n - \mu)^2 \right\} = 0 \quad (19)$$

$$-N + \frac{1}{\sigma^2} \sum_{n=0}^N (x_n - \mu)^2 = 0 \quad (20)$$

$$N\sigma^2 = \sum_{n=0}^N (x_n - \mu)^2 \quad (21)$$

$$\sigma^2 = \frac{1}{N} \sum_{n=0}^N (x_n - \mu)^2 \quad (22)$$

(22) 式を見るとこれは分散の定義そのものである。従ってガウス関数のパラメーター  $\sigma$  はデータの分散である。

以上 2 回の最尤推定により、Gauss 関数のパラメーターを解析的に決定することができた。よって、適当なデータを用意し、その平均と分散とを (4) 式に代入することでヒストグラムを Gauss 関数で近似することができる。

## 2 勾配降下法

前節では Gauss 関数の最尤推定を行って解析的にパラメーターを導いたが、関数によっては解析的にパラメーターを導くことが困難、あるいは不可能といった場合が存在する。こうした場合においては数値解析的にパラメーターを求めることが可能である。この分野では目的関数を最小化することを目的とする。この最小化する目的関数を損失関数と呼ぶ。今回の場合においては、損失関

数是对数尤度関数と関係を持つ.

$$E(\theta) \equiv -\log_e L(\theta) \quad (23)$$

こうした微分可能な関数の数値解析において最も簡単な方法は勾配降下法と呼ばれるものである. これは次式のように, 関数の勾配を利用して解を求める.

$$\theta_{t+1} = \theta_t - \gamma \nabla E(\theta_t) \quad (24)$$

これを Gauss 関数に適用する. Gauss 関数にはパラメーター  $\mu$  および  $\sigma$  が存在した. これをまとめてよって  $\theta$  とする. 勾配降下法のためには  $\theta$  の初期値  $\theta_0$  を与えねばならない. これで準備は済んだので勾配降下法を行っていく. まずは  $\theta_1$  を求める.

$$\theta_1 = \theta_0 - \gamma \nabla E(\theta_0) \quad (25)$$

$$= \theta_0 - \gamma \begin{pmatrix} \frac{\partial E}{\partial \mu} \\ \frac{\partial E}{\partial \sigma} \end{pmatrix} \quad (26)$$

次に  $\theta_2$  を求めるのだが, 計算は全く同じで添字が変化するのみである.

$$\theta_2 = \theta_1 - \gamma \nabla E(\theta_1) \quad (27)$$

$$= \theta_1 - \gamma \begin{pmatrix} \frac{\partial E}{\partial \mu} \\ \frac{\partial E}{\partial \sigma} \end{pmatrix} \quad (28)$$

このように適当な値に収束するまで微分を繰り返していく.