

「パターン認識と機械学習」

2.3.6 ガウス分布に対するベイズ推論

小野寺喜行

# ガウス分布の平均 $\mu$ の推定

分散 $\sigma^2$ は既知

与えられた $N$ 個の観測値集合 $\mathbf{x}=\{x_1, \dots, x_N\}$

→ 平均 $\mu$ を推定することを考える

# ガウス分布の平均 $\mu$ の推定

$\mu$ が与えられたときに観測データが生じる確率である尤度関数

$$p(\mathbf{x} | \mu) = \prod_{n=1}^N p(x_n | \mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2\right\}$$

事前分布 $p(\mu)$ (ガウス分布):  $\uparrow$ の共役事前分布

$$p(\mu) = N(\mu | \mu_0, \sigma_0^2)$$

事後分布

$$p(\mu | \mathbf{x}) \propto p(\mathbf{x} | \mu) p(\mu) \rightarrow p(\mu | \mathbf{x}) = N(\mu | \mu_N, \sigma_N^2)$$

# ガウス分布の平均 $\mu$ の推定

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0 + \frac{N\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_{ML} \quad (\text{平均})$$

→  $N=0$  のとき事前分布の平均

$N \rightarrow \infty$  のとき最尤推定解

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \quad (\text{精度})$$

→  $N$ が増えるにつれ増加

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^N x_n \quad (\text{最尤推定解})$$

観測値が無限にある極限でのベイズ的な定式化によって得られる

# 平均 $\mu$ の逐次的な更新式

最終データ点 $x_N$ の影響を分け事後分布を書くと

$$p(\mu | \mathbf{x}) \propto \left[ p(\mu) \prod_{n=1}^{N-1} p(x_n | \mu) \right] p(x_N | \mu)$$

事前分布とみなす

ベイズの定理で結合

N個のデータ点を観測した後の事後分布と見なせる

# ガウス分布の精度 $\lambda$ の推定

平均 $\mu$ は既知

与えられた $N$ 個の観測値集合 $\mathbf{x}=\{x_1, \dots, x_N\}$

→ 分散平均 $\sigma^2$ を推定することを考える

↓ 精度 $\lambda \equiv 1/\sigma^2$ で操作する方が便利

精度で計算する

# ガウス分布の精度 $\lambda$ の推定

$\lambda$ についての尤度関数

$$p(\mathbf{x} | \lambda) = \prod_{n=1}^N N(x_n | \mu, \lambda^{-1}) \propto \lambda^{N/2} \exp\left\{-\frac{\lambda}{2} \sum_{n=1}^N (x_n - \mu)^2\right\}$$

事前分布(ガンマ分布):  $\uparrow$ の共役事前分布

$$\text{Gam}(\lambda | a, b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda)$$

$$\text{平均: } \mathbf{E}[\lambda] = \frac{a}{b} \quad \text{分散: } \text{var}[\lambda] = \frac{a}{b^2}$$

# ガウス分布の精度 $\lambda$ の推定

事前分布 $\text{Gam}(\lambda \mid a_0, b_0)$ に尤度関数をかけると

$$p(\lambda \mid \mathbf{x}) \propto \lambda^{a_0-1} \lambda^{N/2} \exp \left\{ -b_0 \lambda - \frac{\lambda}{2} \sum_{n=1}^N (x_n - \mu)^2 \right\}$$



(事後分布)

$\text{Gam}(\lambda \mid a_N, b_N)$  (ガンマ分布)

# ガウス分布の精度 $\lambda$ の推定

$N$ 個のデータ点を観測すると

$$a_N = a_0 + \frac{N}{2}$$

→ 係数 $a$ を $N/2$ だけ増やす

$$b_N = b_0 + \frac{1}{2} \sum_{n=1}^N (x_n - \mu)^2 = b_0 + \frac{N}{2} \sigma_{ML}^2$$

→  $N\sigma_{ML}^2/2$ だけ $b$ に影響を与える

# 尤度関数の $\mu$ と $\lambda$ への依存関係

平均 $\mu$ は精度 $\lambda$ は未知

共役事前分布を求めるために考える

尤度関数

$$p(\mathbf{x} | \mu, \lambda) = \prod_{n=1}^N \left( \frac{\lambda}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda}{2} (x_n - \mu)^2 \right\}$$
$$\propto \left[ \lambda^{1/2} \exp \left( -\frac{\lambda \mu^2}{2} \right) \right]^N \exp \left\{ \lambda \mu \sum_{n=1}^N x_n - \frac{\lambda}{2} \sum_{n=1}^N x_n^2 \right\}$$

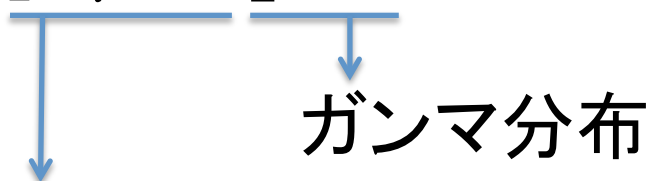
# 尤度関数の $\mu$ と $\lambda$ への依存関係

尤度関数と同じ $\mu, \lambda$ への関数依存性を備えた事前分布を求めるので、次の形式でなくてはならない

$$\begin{aligned} p(\mu, \lambda) &\propto \left[ \lambda^{1/2} \exp\left(-\frac{\lambda\mu^2}{2}\right) \right]^N \exp\{c\lambda\mu - d\lambda\} \\ &= \exp\left\{-\frac{\beta\lambda}{2}(\mu - c/\beta)^2\right\} \lambda^{\beta/2} \exp\left\{-\left(d - \frac{c^2}{2\beta}\right)\lambda\right\} \end{aligned}$$

# 尤度関数の $\mu$ と $\lambda$ のへの依存関係

$p(\mu, \lambda) = p(\mu | \lambda) p(\lambda)$  に対応する部分を見いだすと



精度が $\lambda$ の線形関数であるガウス分布

$$p(\mu, \lambda) = N\left(\mu \mid \mu_0, (\beta\lambda)^{-1}\right) \text{Gam}(\lambda \mid a, b)$$


$$\mu_0 = c / \beta \quad a = 1 + \beta / 2 \quad b = d - c^2 / 2\beta$$

正規-ガンマ分布 (ガウス-ガンマ分布)

# $D$ 次元変数の多変量ガウス分布


$D$ 次元変数の多変量ガウス分布  $N(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$

精度が既知のとき: 平均 $\boldsymbol{\mu}$ の共役事前分布

  
ガウス分布

# $D$ 次元変数の多変量ガウス分布

平均が既知のとき: 精度行列 $\Lambda$ の共役事前分布


$$W(\Lambda | \mathbf{W}, \nu) = B |\Lambda|^{(\nu-D-1)/2} \exp\left(-\frac{1}{2} \text{Tr}(\mathbf{W}^{-1} \Lambda)\right)$$

(ウィシャート分布)

$\nu$ : 分布の自由度パラメータ

$\mathbf{W}$ :  $D \times D$ の尺度行列     $\text{Tr}(\cdot)$ : トレース

$B$ : 正規化定数

$$B(\mathbf{W}, \nu) = |\mathbf{W}|^{-\nu/2} \left( 2^{\nu D/2} \pi^{D(D-1)/4} \prod_{i=1}^D \Gamma\left(\frac{\nu+1-i}{2}\right) \right)^{-1}$$

# $D$ 次元変数の多変量ガウス分布

平均と精度の両方が未知の場合  
の共役事前分布

$$p(\mu, \Lambda | \mu_0, \beta, \mathbf{W}, \nu) = N\left(\mu | \mu_0, (\beta\Lambda)^{-1}\right) W(\Lambda | \mathbf{W}, \nu)$$



正規-ウィシャート分布 (ガウス-ウィシャート分布)