2.4.3 無情報事前分布

パターン認識と機械学習

茂木 哲矢

情報工学科

2008 **年** 07 **月** 18 **日**

目次

無情報事前分布

事前分布が分からない

ある変数に確率 0 を割り当てる．
　$\longrightarrow$ 事後分布でもその確率は 0．

無情報事前分布の特徴

- 事後分布に与える影響が少ない事前分布のことを<u>無情報事前分布</u>という．
- 分布の形状に知見がない場合に用いる．

$p(\lambda) = \mathrm{const}$ とした場合

- λ が K 個の状態をとる離散変数の場合 $p(\lambda) = 1/K$ で OK．
- 連続変数の場合は問題あり．
  1. λ が有界でない場合
     $$\int p(\lambda) d\lambda = \infty$$
     これを <u>変則事前分布</u> という．
  2. 非線形な変数変換をした場合 $\lambda = \eta^2$
     $p_\lambda(\lambda) = \mathrm{const}$ とすると，$p_\eta(\eta)$ は式 (1.27) より，
     $$p_\eta(\eta) = p_\lambda(\lambda)\left|\frac{d\lambda}{d\eta}\right| = p_\lambda(\eta^2) 2\eta \propto \eta \tag{2.231}$$

無情報事前分布の例

1. 平行移動不変性を持つ分布．
2. 尺度不変性を持つ分布．

平行移動不変性 I

$$p(x|\mu) = f(x-\mu) \tag{2.232}$$

- μ は 位置パラメータ.
- 平行移動不変性 を持つ.
- $\hat{x} = x + c$ としても，同じ形になる.

$$p(x|\mu) = f(x-\mu) = f((x+c)-(\mu+c)) = f(\hat{x}-\hat{\mu}) = p(\hat{x}|\hat{\mu})$$

ただし，$\hat{\mu} = \mu + c$

平行移動不変性 II

平行移動不変性を持つ確率分布を選ぶ．区間 $A \le \mu \le B$ に入る確率が区間 $A - c \le \mu \le B - c$ に入る確率と等しくなる．

$$\int_A^B p(\mu) d\mu = \int_{A-c}^{B-c} p(\mu) d\mu = \int_A^B p(\mu - c) d\mu \tag{2.234}$$

よって，

$$p(\mu - c) = p(\mu) \tag{2.235}$$

- $p(\mu) = \mathrm{const}$

位置パラメータの例

ガウス分布の平均 μ

- **式** (2.138) **より，共役事前分布は** $\mathcal{N}(\mu|\mu_0, \sigma_0^2)$
- **この事前分布で** $\sigma_0^2 \to \infty$ **の極限を取れば，無情報事前分布．**
- **式** (2.141)**，**(2.142) **より，**

$$\mu_N = \frac{\sigma^2}{N\infty + \sigma^2}\mu_0 + \frac{N}{N + \sigma^2/\infty}\mu_{ML} = \mu_{ML}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\infty} + \frac{N}{\sigma^2} = \frac{N}{\sigma^2}$$

よって，事後分布 $\mathcal{N}(\mu|\mu_N, \sigma_N^2)$ **は事前分布に影響されない．**

尺度不変性 I

$$p(x|\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) \quad (\sigma > 0) \tag{2.236}$$

- σ は尺度パラメータ.
- 尺度不変性を持つ.
- $\hat{x} = cx$ としても，同じ形式になる.

$$p_{\hat{x}}(\hat{x}) = p_x(x)\left|\frac{dx}{d\hat{x}}\right| = p_x\left(\frac{\hat{x}}{c}\right)\frac{1}{c} = \frac{1}{c\sigma} f\left(\frac{\hat{x}}{c\sigma}\right) = p_x(\hat{x}|\hat{\sigma})$$

ただし，$\hat{\sigma} = c\sigma$.

尺度不変性 II

尺度不変性をもつような事前分布を選ぶ．事前分布が区間 $A \le \sigma \le B$ と拡大縮小後の区間 $A/c \le \sigma \le B/c$ に入る確率と等しい．

$$\int_A^B p(\sigma)d\sigma = \int_{A/c}^{B/c} p(\sigma)d\sigma = \int_A^B p\left(\frac{1}{c}\sigma\right)\frac{1}{c}d\sigma \tag{2.238}$$

よって，

$$p(\sigma) = p\left(\frac{1}{c}\sigma\right)\frac{1}{c} \tag{2.239}$$

- $p(\sigma) \propto 1/\sigma$
- $\int_0^\infty p(\sigma)d\sigma = \infty$
- **式** (1.27) **より** $p(\ln\sigma) = \mathrm{const}$

尺度パラメータの例

位置パラメータ μ を考慮したガウス分布での標準偏差 σ

$$\mathcal{N}(x|\mu,\sigma^2) \propto \sigma^{-1} \exp\left\{-\left(\frac{\hat{x}}{\sigma}\right)^2\right\}$$

ただし，$\hat{x} = x - \mu$．

- $p(\lambda) \propto 1/\lambda$．**ただし，**$1/\sigma^2 = \lambda$．
- **式** (2.146) **より，**λ **の共役事前分布は** $\mathrm{Gam}(\lambda|a_0, b_0)$．
- **この事前分布で** $a_0 = b_0 = 0$ **とすると，無情報事前分布．**
- **式** (2.150)**，**(2.151) **より，**

$$a_N = 0 + \frac{N}{2} = \frac{N}{2}$$

$$b_N = 0 + \frac{N}{2}\sigma_{ML}^2 = \frac{N}{2}\sigma_{ML}^2$$

よって，事後分布 $\mathrm{Gam}(\lambda|a_N, b_N)$ **は事前分布に依存しない．**