正規分布を使った
ベイズ的モデリング

正田 備也

masada@rikkyo.ac.jp

Contents

単変量正規分布

- 単変量正規分布は、$\mathbb{R} = (-\infty, \infty)$ 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 $x \sim \mathcal{N}(\mu, \sigma^2)$ の確率密度関数：

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

多変量正規分布

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

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

多変量正規分布（共分散行列が対角行列の場合）

- 実際には共分散行列を対角行列と仮定することも多い
  - Σ の扱いがしばしば数値計算的に難しいため
- $\boldsymbol{\Sigma}$ が $\sigma_1^2, \ldots, \sigma_d^2$ を対角成分とする対角行列のとき、密度関数は単変量正規分布の密度関数の積となる

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left\{ -\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{2} \right\}$$
$$= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right) \quad (3)$$

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 $\mathcal{D} \equiv \{x_1, \ldots, x_N\}$ とする
  - 各 x_i は、$-\infty < x_i < \infty$ を満たす実数値とする
- 各観測データ x_i を同じ正規分布 $\mathcal{N}(\mu, \sigma^2)$ に独立にしたがうものとしてモデル化することにする（つまり i.i.d. を仮定）
- このとき、データセット $\mathcal{D}$ の尤度は以下のように μ と σ の関数として書くことができる：

$$p(\mathcal{D}; \mu, \sigma) = \prod_{i=1}^{N} p(x_i; \mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x_i - \mu)^2}{2\sigma^2} \right) \tag{4}$$

単変量正規分布の最尤推定

- 式 (4) より、観測データ $\mathcal{D} = \{x_1, \dots, x_N\}$ の対数尤度は

$$\ln p(\mathcal{D}; \mu, \sigma) = -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2} \tag{5}$$

- この対数尤度を最大化する μ と σ を求めると

$$\hat{\mu} = \frac{\sum_i x_i}{N} = \bar{x} \ , \quad \hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{N} \tag{6}$$

多変量正規分布の最尤推定 (1/2)

- 観測データ $\mathcal{D} \equiv \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}, \boldsymbol{x}_i \in \mathbb{R}^d$ の対数尤度は

$$p(\mathcal{D}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right] \tag{7}$$

- この対数尤度を最大化する $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ を求めると

$$\hat{\boldsymbol{\mu}} = \frac{\sum_i \boldsymbol{x}_i}{N} = \bar{\boldsymbol{x}} \ , \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_i (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\intercal \tag{8}$$

多変量正規分布の最尤推定 (2/2)

- 共分散行列が対角行列だと仮定する
- 観測データ $\mathcal{D} \equiv \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}, \boldsymbol{x}_i \in \mathbb{R}^d$ の対数尤度は

$$\ln p(\mathcal{D}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{N}{2}\sum_{j=1}^{d} \ln(2\pi\sigma_j^2) - \sum_{i=1}^{N}\sum_{j=1}^{d} \frac{(x_{i,j} - \mu_j)^2}{2\sigma_j^2} \tag{9}$$

- この対数尤度を最大化する $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ を求めると

$$\hat{\mu}_j = \frac{\sum_i x_{i,j}}{N} = \bar{x}_j \ , \quad \hat{\sigma}_j^2 = \frac{\sum_i (x_{i,j} - \bar{x}_j)^2}{N} \tag{10}$$

Contents

ベイズ的なモデリングとは

- 統計モデルは観測データの不確かさ uncertainty を表現する
- だが、ベイズ的な統計モデリングでは、観測データをもとにして統計モデルの**パラメータを決めること自体にも不確かさ uncertainty がある**と考える
- そこで、パラメータも確率変数とみなし、パラメータも確率分布にしたがっているものとしてモデリングする
- そこで導入されるのが事前分布である
- 事前分布はパラメータがしたがう確率分布として導入される

単変量正規分布を使うベイズ的モデリング（1）

- 観測データ $\mathcal{D} \equiv \{x_1, \ldots, x_N\}$ の尤度は $p(\mathcal{D}|\mu, \sigma)$
  - 事前分布を使わないときは $p(\mathcal{D}; \mu, \sigma)$ と書いていた
  - ベイズ的モデリングでは、$p(\mathcal{D}|\mu, \sigma)$ と、条件付き確率として書く
  - 観測変数 x_i だけでなく、μ と σ も確率変数となるからである
- まず、μ についてだけ、それがしたがう事前分布を導入する
  - つまり、σ は自由パラメータのままとする
- μ の事前分布として、正規分布 $\mathcal{N}(\mu_0, \sigma_0^2)$ を選ぶ
- このとき、正規分布が共役事前分布となる
  - このことを次の 2 枚のスライドで示す

共役事前分布としての正規分布（1）

$$
\begin{aligned}
p(\mathcal{D}|\mu;\sigma)p(\mu;\mu_0,\sigma_0) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] \\
&= \frac{1}{(\sqrt{2\pi})^N \sigma^N} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] \\
&= \frac{1}{(\sqrt{2\pi})^N \sigma^N} \exp\left[-\frac{N\mu^2 - 2\sum_i x_i\mu + \sum_i x_i^2}{2\sigma^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{\mu^2 - 2\mu_0\mu + \mu_0^2}{2\sigma_0^2}\right] \\
&= \frac{1}{(\sqrt{2\pi})^{N+1} \sigma^N \sigma_0} \exp\left[-\left(\frac{N}{2\sigma^2} + \frac{1}{2\sigma_0^2}\right)\mu^2 + 2\left(\frac{\sum_i x_i}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}\right)\mu - \frac{\sum_i x_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\sigma_0^2}\right] \qquad (11)
\end{aligned}
$$

よって

$$
p(\mu|\mathcal{D};\sigma,\mu_0,\sigma_0) \propto \exp\left[-\left(\frac{N}{2\sigma^2} + \frac{1}{2\sigma_0^2}\right)\mu^2 + 2\left(\frac{\sum_i x_i}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}\right)\mu\right] \qquad (12)
$$

指数関数の中身に注目すると・・・

$$
\begin{aligned}
&\left(\frac{N}{2\sigma^2}+\frac{1}{2\sigma_0^2}\right)\mu^2-2\left(\frac{\sum_i x_i}{2\sigma^2}+\frac{\mu_0}{2\sigma_0^2}\right)\mu=\frac{N\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left(\mu^2-2\frac{N\sigma_0^2\bar{x}+\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}\mu\right)\\
&=\frac{N\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left(\mu-\frac{N\sigma_0^2\bar{x}+\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}\right)^2+const.
\end{aligned}
\tag{13}
$$

以上より、

$$
p(\mu|\mathcal{D};\sigma,\mu_0,\sigma_0)\propto\exp\left[-\frac{N\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left(\mu-\frac{N\sigma_0^2\bar{x}+\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}\right)^2\right]
\tag{14}
$$

共役事前分布としての正規分布（2）

- 事後分布 $p(\mu|\mathcal{D};\sigma,\mu_0,\sigma_0)$ は、下に示した平均と分散を持つ正規分布 $\mathcal{N}(\mu_{\mathcal{D}},\sigma^2_{\mathcal{D}})$ であることが分かった

$$\mu_{\mathcal{D}} = \frac{N\sigma_0^2\bar{x} + \sigma^2\mu_0}{N\sigma_0^2 + \sigma^2} \;,\; \sigma^2_{\mathcal{D}} = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2 + \sigma^2} \tag{15}$$

- 平均 $\frac{N\sigma_0^2\bar{x}+\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}$ は、$\bar{x}$ と μ_0 を $\frac{N}{\sigma^2}$ 対 $\frac{1}{\sigma_0^2}$ の割合で混ぜたもの
- 分散 $\frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}$ の逆数は、$\frac{N}{\sigma^2}$ と $\frac{1}{\sigma_0^2}$ の和
  - 分散の逆数を精度 (precision) という

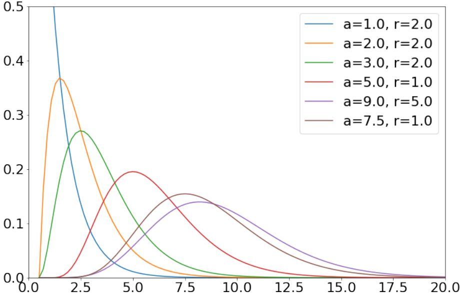
単変量正規分布を使うベイズ的モデリング（2）

- 観測データ $\mathcal{D} \equiv \{x_1, \ldots, x_N\}$ の尤度は $p(\mathcal{D}|\mu, \sigma)$
- 今度は、μ と σ^2 の両方について事前分布を導入する
- ただし、分散 σ^2 については、その逆数である精度 precision $\tau \equiv \sigma^{-2}$ がしたがう事前分布を導入する
- このとき、正規ガンマ分布 normal-gamma distribution が共役事前分布となる
- 正規ガンマ分布の確率密度関数は

$$p(\mu, \tau; \mu_0, \lambda_0, \alpha, \beta) = \frac{\beta^\alpha \sqrt{\lambda_0}}{\Gamma(\alpha)\sqrt{2\pi}} \tau^{\alpha - \frac{1}{2}} e^{-\beta\tau} e^{-\frac{\lambda_0 \tau (\mu - \mu_0)^2}{2}} \tag{16}$$

ガンマ分布

- ガンマ分布は非負実数 $[0, \infty)$ 上に定義される確率分布
- パラメータ
  - shape パラメータ α
  - rate パラメータ β
- ガンマ分布の確率密度関数は



$$p(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{17}$$

正規ガンマ分布の密度関数の見方

$$
\begin{aligned}
p(\mu,\tau;\mu_0,\lambda_0,\alpha,\beta) &= \frac{\beta^\alpha\sqrt{\lambda_0}}{\Gamma(\alpha)\sqrt{2\pi}}\tau^{\alpha-\frac{1}{2}}e^{-\beta\tau}e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau}\times\frac{\sqrt{\lambda_0}}{\sqrt{2\pi}}\tau^{\frac{1}{2}}e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau}\times\frac{\sqrt{\lambda_0\tau}}{\sqrt{2\pi}}\exp\left(-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}\right) \qquad (18)
\end{aligned}
$$

- ガンマ分布の密度関数と、正規分布の密度関数との、積
- $\lambda_0\tau$ が、正規分布の精度（分散の逆数）に対応

共役事前分布としての正規ガンマ分布

以下、事後分布 $p(\mu,\tau|\mathcal{D};\mu_0,\lambda_0,\alpha,\beta)$ も正規ガンマ分布であることを示す。

ベイズ則より $p(\mu,\tau|\mathcal{D};\mu_0,\lambda_0,\alpha,\beta) \propto p(\mathcal{D}|\mu,\tau)p(\mu,\tau;\mu_0,\lambda_0,\alpha,\beta)$ である。右辺は、

$$
\begin{aligned}
&p(\mathcal{D}|\mu,\tau)p(\mu,\tau;\mu_0,\lambda_0,\alpha,\beta)\\
&= \prod_{i=1}^{N}\sqrt{\frac{\tau}{2\pi}}e^{-\frac{\tau(x_i-\mu)^2}{2}} \times \frac{\beta^\alpha\sqrt{\lambda_0}}{\Gamma(\alpha)\sqrt{2\pi}}\tau^{\alpha-\frac{1}{2}}e^{-\beta\tau}e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}}\\
&\propto \tau^{\alpha+\frac{N}{2}-\frac{1}{2}}e^{-\beta\tau}e^{-\frac{\tau}{2}\left(\sum_{i=1}^{N}(x_i-\mu)^2+\lambda_0(\mu-\mu_0)^2\right)}\\
&= \tau^{\alpha+\frac{N}{2}-\frac{1}{2}}e^{-\beta\tau}e^{-\frac{\tau}{2}\{(\lambda_0+N)\mu^2-2(\lambda_0\mu_0+\sum_i x_i)\mu+\lambda_0\mu_0^2+\sum_i x_i^2\}}\\
&\propto \tau^{\alpha+\frac{N}{2}-\frac{1}{2}}\exp\left[-\tau\left(\beta+\frac{(\lambda_0\mu_0^2+\sum_i x_i^2)(\lambda_0+N)-(\lambda_0\mu_0+N\bar{x})^2}{2(\lambda_0+N)}\right)\right]\\
&\qquad \times \exp\left[-\frac{\tau}{2}(\lambda_0+N)\left(\mu-\frac{\lambda_0\mu_0+N\bar{x}}{\lambda_0+N}\right)^2\right] \qquad (19)
\end{aligned}
$$

ここで標本分散を s とおくと、$s = \frac{\sum_i x_i^2}{N} - \bar{x}^2$ となるから、

$$
\begin{aligned}
&(\lambda_0\mu_0^2 + \sum_i x_i^2)(\lambda_0 + N) - (\lambda_0\mu_0 + N\bar{x})^2 \\
&= \lambda_0 \sum_i x_i^2 + N\lambda_0\mu_0^2 + N\sum_i x_i^2 - 2\lambda_0\mu_0 N\bar{x} - N^2\bar{x}^2 \\
&= \lambda_0 N(s + \bar{x}^2) + N\lambda_0\mu_0^2 - 2\lambda_0\mu_0 N\bar{x} + N^2 s \\
&= \lambda_0 N(\bar{x} - \mu_0)^2 + Ns(\lambda_0 + N)
\end{aligned}
\tag{20}
$$

よって

$$
\begin{aligned}
&p(\mu, \tau|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta) \propto p(\mathcal{D}|\mu, \tau)p(\mu, \tau; \mu_0, \lambda_0, \alpha, \beta) \\
&\propto \tau^{\alpha + \frac{N}{2} - \frac{1}{2}} \exp\left[-\tau\left(\beta + \frac{Ns}{2} + \frac{\lambda_0 N(\bar{x} - \mu_0)^2}{2(\lambda_0 + N)}\right)\right] \exp\left[-\frac{\tau}{2}(\lambda_0 + N)\left(\mu - \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N}\right)^2\right]
\end{aligned}
\tag{21}
$$

この式は、事後分布も正規ガンマ分布であることを示している。

多変量正規分布を使ったベイズ的モデリング

- 多変量正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ の場合も、平均パラメータ $\boldsymbol{\mu}$ については正規分布 $\mathcal{N}(\boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda})^{-1})$ を事前分布として使う
- 精度行列 $\boldsymbol{\Lambda}$（ただし $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$）については、次のような密度関数を持つウィシャート分布を事前分布として使う

$$\mathcal{W}(\boldsymbol{W}, \nu) = B|\boldsymbol{\Lambda}|^{(\nu-D-1)/2} \exp\left(-\frac{1}{2}\mathrm{Tr}(\boldsymbol{W}^{-1}\boldsymbol{\Lambda})\right) \tag{22}$$

- B は規格化定数で、以下のような $\boldsymbol{W}$ と ν の関数である

$$B(\boldsymbol{W}, \nu) = |\boldsymbol{W}|^{-\nu/2}\left(2^{\nu d/2}\pi^{d(d-1)/4}\prod_{i=1}^{d}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1} \tag{23}$$

共役事前分布としての正規ウィシャート分布

- 正規ウィシャート分布が多変量正規分布の共役事前分布になっていることの証明は割愛する
- Christopher M. Bishop, *Pattern Recognition and Machine Learning* の Exercise 2.45 参照

Contents

Rosenthal and Jacobson (1968) の実験

- IQ スコアの変化の分析
  - この例は、STA 360/602: Bayesian Methods and Modern Statistics @ Duke University の Module 4 から取った
  - いわゆる「ピグマリオン効果」を明らかにした実験らしい
- 実験の設定
  - 先生の期待は学生の学修に影響するかを調べたい。まず、年度初めに IQ テストを実施。そして、各クラスから 2 割の学生を無作為に選び、先生に「この学生は伸びる学生 spurter だ」と告げる。年度終わりにまた IQ テストを実施。IQ スコアの変化を調べる。

データ例

```
#spurters
x <- c(18, 40, 15, 17, 20, 44, 38)
#controls
y <- c(-4, 0, -19, 24, 19, 10, 5, 10,
        29, 13, -9, -8, 20, -1, 12, 21,
        -7, 14, 13, 20, 11, 16, 15, 27,
        23, 36, -33, 34, 13, 11, -19, 21,
        6, 25, 30, 22, -28, 15, 26, -1, -2,
        43, 23, 22, 25, 16, 10, 29)
iqData <- data.frame(Treatment =
        c(rep("Spurters", length(x)),
        rep("Controls", length(y))),
        Gain = c(x, y))
```

分析の方法

- 知りたいのは、spurters の平均スコア μ_S と、controls の平均スコア μ_C とについて、$\mu_S > \mu_C$ となる確率
- サンプル数が少ないため、spurters と controls それぞれの分散をちゃんと推定できなさそう
- そこで、spurters と controls それぞれの平均と分散に、別々の正規ガンマ分布を事前分布として使う
- 観測データをもとに事後分布を計算、その事後分布から 10 万のサンプル対を draw し、spurters の事後分布からのサンプルのほうが大きかった割合を求める
  - 分布からのサンプリングについては「統計モデリング 2」で

Contents

指数型分布族 exponential family

- 以下のような形の確率密度関数を持つ確率分布をまとめて、指数型分布族と呼ぶ

$$p(\boldsymbol{x}|\boldsymbol{\eta}) = h(\boldsymbol{x})g(\boldsymbol{\eta})\exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) \tag{24}$$

- $\boldsymbol{\eta}$ は分布のパラメータだが、指数型分布族については特に、自然パラメータ (natural parameter) と呼ばれる
- $g(\boldsymbol{\eta})$ は、規格化のために導入されている係数とみなせる
  - $g(\boldsymbol{\eta})$ は $\boldsymbol{x}$ を含まないことに注意
- 確率変数 $\boldsymbol{x}$ がとる値は、スカラーでもベクトルでもよいし、離散値でも連続値でもよい。

例. ベルヌーイ分布

- 確率質量関数が式 (24) の形を持つことを確かめる。

$$
\begin{aligned}
p(x|\phi) &= \phi^x (1-\phi)^{1-x} \\
&= \exp(x \ln \phi + (1-x)\ln(1-\phi)) \\
&= (1-\phi)\exp\left(\ln\left(\frac{\phi}{1-\phi}\right) \times x\right)
\end{aligned} \tag{25}
$$

注. 例えばコイン投げの場合、$x = 1$ は表、$x = 0$ は裏が出ること表す。

- $\eta = \ln\left(\frac{\phi}{1-\phi}\right)$ とすればよい。
- すると $g(\eta) = 1 - \phi = \frac{1}{1+e^{\eta}} = \sigma(-\eta)$ となる。
- そして $h(x) = 1$ とすれば式 (24) の形になる。

指数型分布族の対数尤度

- 観測データ $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ が独立に同じ指数型分布族の分布にしたがうとき、$\mathcal{D}$ の対数尤度は

$$
\begin{aligned}
\ln p(\mathcal{D}|\boldsymbol{\eta}) &= \ln \prod_{i=1}^{N} p(\boldsymbol{x}_i|\boldsymbol{\eta}) \\
&= \sum_{i=1}^{N} \ln h(\boldsymbol{x}_i) + N \ln g(\boldsymbol{\eta}) + \boldsymbol{\eta}^\intercal \sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i) \qquad (26)
\end{aligned}
$$

- このとき、$\frac{\partial}{\partial \boldsymbol{\eta}} \ln p(\mathcal{D}|\boldsymbol{\eta}) = 0$ とおいて $\boldsymbol{\eta}$ を求める計算（最尤推定でおこなう計算）は、どういう計算になるだろうか？

$$\frac{\partial}{\partial \boldsymbol{\eta}} \ln p(\mathcal{D}|\boldsymbol{\eta}) = \frac{N}{g(\boldsymbol{\eta})} \frac{\partial g(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} + \sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i) \tag{27}$$

ここで、$\frac{1}{g(\boldsymbol{\eta})} = \int h(\boldsymbol{x}) \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x}$（$\because$ 密度関数なので積分すると 1）より

$$\begin{aligned} -\frac{1}{(g(\boldsymbol{\eta}))^2} \frac{\partial g(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= \frac{\partial}{\partial \boldsymbol{\eta}} \int h(\boldsymbol{x}) \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x} = \int h(\boldsymbol{x}) \frac{\partial}{\partial \boldsymbol{\eta}} \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x} \\ &= \int h(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x}) \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x} \end{aligned} \tag{28}$$

よって

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\eta}} \ln p(\mathcal{D}|\boldsymbol{\eta}) &= -N g(\boldsymbol{\eta}) \int h(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x}) \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x} + \sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i) \\ &= -N \int \boldsymbol{u}(\boldsymbol{x}) h(\boldsymbol{x}) g(\boldsymbol{\eta}) \exp(\boldsymbol{\eta}^\intercal \boldsymbol{u}(\boldsymbol{x})) d\boldsymbol{x} + \sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i) \\ &= -N \left(\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{\eta})}[\boldsymbol{u}(\boldsymbol{x})] - \frac{\sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i)}{N} \right) \end{aligned} \tag{29}$$

指数型分布族の最尤推定

- 以上より、$\frac{\partial}{\partial \boldsymbol{\eta}} \ln p(\mathcal{D}|\boldsymbol{\eta}) = 0$ とおくと、下の結果を得る。

$$\mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{\eta})}[\boldsymbol{u}(\boldsymbol{x})] = \frac{\sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i)}{N} \tag{30}$$

- この結果は、指数型分布族に含まれる確率分布であれば、どの分布についても当てはまる。
  - 例. ベルヌーイ分布の場合、式 (30) は何を意味するか？

共役事前分布

- 式 (24) のような密度関数を持つどの確率分布に対しても、以下の形の密度関数を持つ共役事前分布が存在する

$$p(\boldsymbol{\eta}; \boldsymbol{\chi}, \nu) = f(\boldsymbol{\chi}, \nu) g(\boldsymbol{\eta})^{\nu} \exp(\nu \boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{\chi}) \tag{31}$$

- つまり、式 (31) の形の密度関数を持つ確率分布を事前分布とすると、事後分布が事前分布と同じ形の密度関数を持つ
- $f(\boldsymbol{\chi}, \nu)$ は、規格化のために導入されている係数とみなせる
  - $f(\boldsymbol{\chi}, \nu)$ は $\boldsymbol{\eta}$ を含まないことに注意

例. ベルヌーイ分布の共役事前分布

- $\eta = \ln\left(\frac{\phi}{1-\phi}\right)$ および $g(\eta) = 1-\phi$ だったので

$$
\begin{aligned}
p(\eta;\chi,\nu) &\propto (1-\phi)^{\nu}\exp\left(\nu\ln\left(\frac{\phi}{1-\phi}\right)\chi\right) \\
&= (1-\phi)^{\nu}\phi^{\nu\chi}(1-\phi)^{-\nu\chi} \\
&= \phi^{\nu\chi}(1-\phi)^{\nu(1-\chi)} \qquad (32)
\end{aligned}
$$

- この式の形は、ベータ分布の密度関数の式の形と、同じ！
  - $\nu = \alpha+\beta-2$ および $\chi = \frac{\alpha-1}{\alpha+\beta-2}$ と置き換えればよい

共役事前分布を用いたときの事後分布

- 式 (24) の密度関数を持つ確率分布に対して・・・
- 式 (31) の事前分布を使うと・・・
- 観測データ $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ が所与のときの事後分布は、以下のようになる

$$p(\boldsymbol{\eta}|\mathcal{D}, \boldsymbol{\chi}, \nu) \propto g(\boldsymbol{\eta})^{\nu+N} \exp\left(\boldsymbol{\eta}^\intercal\left(\sum_{i=1}^{N} \boldsymbol{u}(\boldsymbol{x}_i) + \nu\boldsymbol{\chi}\right)\right) \quad (33)$$

問. このことを示せ（cf. PRML, Sec. 2.4.2）

課題

- 二項分布も、試行の総数を表すパラメータ n が固定されているならば、指数型分布族に属する
- そこで、n が固定されている二項分布の共役事前分布を、式 (31) をもとに求めてみよう
- ヒント：確率質量関数 $p(k;\boldsymbol{\phi},n)=\frac{n!}{k!(n-k)!}\phi_1^k(1-\phi_1)^{n-k}$ を、まず式 (24) の形に変形する。