PRML 2.4 指数型分布族

5501 酒井一徳
10/11/17

目次

2.4　指数型分布族

指数型分布族-定義

$\mathbf{x}$上の指数型分布族

$$p(\mathbf{x}|\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}$$

- **$\mathbf{x}$はスカラーでもベクトルでも、離散でも連続でもよい**
- $\boldsymbol{\eta}$　　：自然パラメータ
- $\mathbf{u}(\mathbf{x})$　：任意の関数
- $g(\boldsymbol{\eta})$　：正規化係数

$$g(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathrm{d}\mathbf{x} = 1$$

ベルヌーイ分布(1/2)

$$
\begin{aligned}
p(x|\mu) = \mathrm{Bern}(x|\mu) &= \mu^x (1-\mu)^{1-x} \\
&= \exp\{x \ln \mu + (1-x)\ln(1-\mu)\} \\
&= (1-\mu)\exp\left\{\ln\left(\frac{\mu}{1-\mu}\right)x\right\}
\end{aligned}
$$

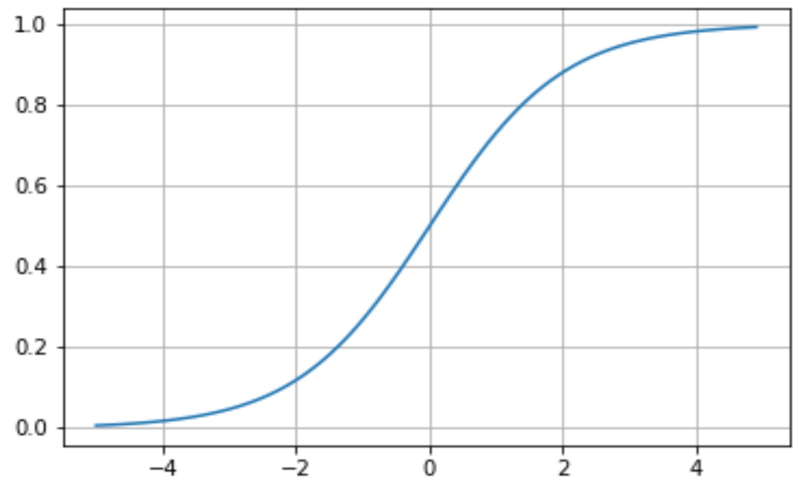
定義と
見比べて

$$\eta = \ln\left(\frac{\mu}{1-\mu}\right)$$

$$\mu = \sigma(\eta)$$

ロジスティックシグモイド関数

$$\sigma(\eta) = \frac{1}{1+\exp(-\eta)}$$



ベルヌーイ分布(2/2)

ベルヌーイ分布

$$p(x|\eta) = \sigma(-\eta)\exp(\eta x)$$

ただし、

$$u(x) = x$$
$$h(x) = 1$$
$$g(\eta) = \sigma(-\eta)$$

- ベルヌーイ分布は指数型分布族である

多項分布(カテゴリカル分布)(1/4)

$$p(\mathbf{x}|\boldsymbol{\mu}) = \mathrm{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^{M} \mu_k^{x_k}$$

$$= \exp\left\{\sum_{k=1}^{M} x_k \ln \mu_k\right\}$$

where $\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$

定義と
見比べて

$$\eta_k = \ln \mu_k$$

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)^{\mathrm{T}}$$

カテゴリカル分布

$$p(\mathbf{x}|\boldsymbol{\eta}) = \exp\left(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{x}\right)$$

ただし、

$$\mathbf{u}(\mathbf{x}) = \mathbf{x}$$
$$h(\mathbf{x}) = 1$$
$$g(\boldsymbol{\eta}) = 1$$

カテゴリカル分布も
指数型分布族である

多項分布(カテゴリカル分布)(2/4)

しかし、 $\mu_M = 1 - \sum_{k=1}^{M-1} \mu_k$ から、μ_k と η_k は独立ではない

残存する制約

$$0 \leqslant \mu_k \leqslant 1, \quad \sum_{k=1}^{M-1} \mu_k \leqslant 1$$

M-1個のパラメータで
表現することもできる

$$\begin{aligned}\exp\left\{\sum_{k=1}^{M} x_k \ln \mu_k\right\} &= \exp\left\{\sum_{k=1}^{M-1} x_k \ln \mu_k + \left(1 - \sum_{k=1}^{M-1} x_k\right) \ln\left(1 - \sum_{k=1}^{M-1} \mu_k\right)\right\} \\ &= \exp\left\{\sum_{k=1}^{M-1} x_k \ln\left(\frac{\mu_k}{1 - \sum_{j=1}^{M-1} \mu_j}\right) + \ln\left(1 - \sum_{k=1}^{M-1} \mu_k\right)\right\}\end{aligned}$$

多項分布(カテゴリカル分布)(3/4)

$$\ln\left(\frac{\mu_k}{1-\sum_j \mu_j}\right) = \eta_k$$

$$\Leftrightarrow \frac{\mu_k}{1-\sum_{j=1}^{M-1}\mu_j} = \exp\left(\eta_k\right)$$

全てのkについて
足し合わせる

$$(\text{右辺}) = \frac{1}{1-\sum_{j=1}^{M-1}\mu_j}$$

$$(\text{左辺}) = \sum_{k=1}^{M}\eta_k = 1 + \sum_{k=1}^{M-1}\eta_k$$

元の式に
代入しなおすと、

$$\mu_k = \frac{\exp\left(\eta_k\right)}{1+\sum_j \exp\left(\eta_j\right)}$$

ソフトマックス関数

$$f_k(\mathbf{x}) = \frac{\exp(x_k)}{\sum_{i=0}^{M}\exp(x_i)}$$

多項分布(カテゴリカル分布)(4/4)

カテゴリカル分布はソフトマックス関数を用いた表現においても

$$p(\mathbf{x}|\boldsymbol{\eta}) = \left(1 + \sum_{k=1}^{M-1} \exp\left(\eta_k\right)\right)^{-1} \exp\left(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{x}\right)$$

$$\text{where} \quad \boldsymbol{\eta} = \left(\eta_1, \ldots, \eta_{M-1}, 0\right)^{\mathrm{T}}$$

$$\mathbf{u}(\mathbf{x}) = \mathbf{x}$$
$$h(\mathbf{x}) = 1$$
$$g(\boldsymbol{\eta}) = \left(1 + \sum_{k=1}^{M-1} \exp\left(\eta_k\right)\right)^{-1}$$

によって対応付けられることから
指数型分布族の標準形になっている

1変数ガウス分布(1/2)

$$
\begin{aligned}
p(x|\mu,\sigma^2) &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x-\frac{1}{2\sigma^2}\mu^2\right\} \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}\mu^2\right\}\exp\left\{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x\right\} \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}\mu^2\right\}\exp\left\{\begin{pmatrix}\mu/\sigma^2 \\ -1/2\sigma^2\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}x \\ x^2\end{pmatrix}\right\}
\end{aligned}
$$

１変数ガウス分布(2/2)

つまり、１変数ガウス分布も以下の対応付けによって
指数型分布族であるといえる

$$\boldsymbol{\eta} = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix}$$

$$\mathbf{u}(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$$

$$h(x) = (2\pi)^{-1/2}$$

$$g(\boldsymbol{\eta}) = (-2\eta_2)^{1/2} exp\left(\frac{\eta_1^2}{4\eta_2}\right)$$

目次

2.4　指数型分布族

最尤推定量と十分統計量(1/4)

パラメータベクトル$\boldsymbol{\eta}$の推定をする

$$\mathrm{g}(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}d\mathbf{x}=1 \qquad (2.195)$$

$\boldsymbol{\eta}$について勾配をとる

$$\nabla\mathrm{g}(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}d\mathbf{x}+\mathrm{g}(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathbf{u}(\mathbf{x})d\mathbf{x}=0$$

両辺を(2.195)で割る

$$-\frac{\nabla\mathrm{g}(\boldsymbol{\eta})}{\mathrm{g}(\boldsymbol{\eta})}=\mathrm{g}(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathbf{u}(\mathbf{x})d\mathbf{x}=\mathbf{E}\left[\mathbf{u}(\mathbf{x})\right]$$

最尤推定量と十分統計量 (2/4)　　演習(2.58)

つまり、 $-\nabla \ln g(\boldsymbol{\eta}) = \mathbf{E}\left[\mathbf{u}(\mathbf{x})\right]$ が得られる (2.226)

また、共分散については

$$-\nabla \ln g(\boldsymbol{\eta}) = \mathbf{E}\left[\mathbf{u}(\mathbf{x})\right] = g(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathbf{u}(\mathbf{x})d\mathbf{x}$$

$$\begin{aligned}-\nabla\nabla \ln g(\boldsymbol{\eta}) &= \nabla g(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathbf{u}(\mathbf{x})d\mathbf{x} + g(\boldsymbol{\eta})\int h(\mathbf{x})\exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\}\mathbf{u}(\mathbf{x})\mathbf{u}(\mathbf{x})^{\mathrm{T}}d\mathbf{x} \\ &= -\mathbf{E}\left[\mathbf{u}(\mathbf{x})\right]\mathbf{E}\left[\mathbf{u}(\mathbf{x})^{\mathrm{T}}\right] + \mathbf{E}\left[\mathbf{u}(\mathbf{x})\mathbf{u}(\mathbf{x})^{\mathrm{T}}\right] \\ &= \mathrm{cov}\left[\mathbf{u}(\mathbf{x})\right]\end{aligned}$$

より、二次微分から得られる

指数型分布族の分布を正規化できたなら、微分で簡単に分布のモーメントがわかる

最尤推定量と十分統計量(3/4)

同分布に従う独立なデータの集合　$\mathbf{X} = \left\{ \mathbf{x}_1, \ldots, \mathbf{x}_N \right\}$

尤度関数　$$p(\mathbf{X} \mid \boldsymbol{\eta}) = \left(\prod_{n=1}^{N} h(\mathbf{x}_n) \right) g(\boldsymbol{\eta})^N \exp\left\{ \boldsymbol{\eta}^{\mathrm{T}} \sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n) \right\} \quad (2.227)$$

対数尤度関数　$$\ln p(\mathbf{X} \mid \boldsymbol{\eta}) = \sum_{n=1}^{N} \ln h(\mathbf{x}_n) + N \ln g(\boldsymbol{\eta}) + \boldsymbol{\eta}^{\mathrm{T}} \sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n)$$

$\ln p(\mathbf{X} \mid \boldsymbol{\eta})$　の $\boldsymbol{\eta}$ について勾配を 0 とすると、

$$-\nabla \ln g(\boldsymbol{\eta}_{ML}) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n)$$

原則として、
この式をとけば最尤推定量$\boldsymbol{\eta}_{ML}$が
得られる。

最尤推定量と十分統計量(4/4)

$$-\nabla \ln g(\boldsymbol{\eta}_{ML}) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{u}(\mathbf{x}_n)$$

最尤推定解は $\sum_n \mathbf{u}(\mathbf{x}_n)$ に依存する

$N \to \infty$の極限では…

上式の右辺は $\mathbf{E}[\mathbf{u}(\mathbf{x})]$ になるため、(2.226)から最尤推定量が
真の値に等しくなることがわかる

ベイズ推論においてもこの十分性が成立する(８章にて)

目次

2.4　指数型分布族

- 2.4.1 最尤推定と十分統計量
- 2.4.2 共役事前分布
- 2.4.3 無情報事前分布

共役事前分布(1/2)

一般に、ある確率分布 $p(\mathbf{x}|\boldsymbol{\eta})$ について、事後分布がその事前分布と同じ関数形になるような尤度関数と事前分布 $p(\boldsymbol{\eta})$ を求めることは可能である。

指数型分布族　$p(\mathbf{x}\,|\,\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta})\exp\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\}$　(2.194)

の共役事前分布とは

$$p(\boldsymbol{\eta}|\boldsymbol{\chi},\nu) = f(\boldsymbol{\chi},\nu)g(\boldsymbol{\eta})^{\nu}\exp\{\nu\boldsymbol{\eta}^{T}\boldsymbol{\chi}\}$$

$f(\boldsymbol{\chi},\nu)$　：正規化係数

$g(\boldsymbol{\eta})$　：(2.194)のものと同じ

共役事前分布(2/2)

尤度関数(2.227)と事前分布(2.229)の積を確認する

尤度関数 $p(\mathbf{X}|\boldsymbol{\eta}) = \left(\prod_{n=1}^{N} h(\mathbf{x}_n)\right) g(\boldsymbol{\eta})^N \exp\left\{\boldsymbol{\eta}^{\mathrm{T}} \sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n)\right\}$

事前分布 $p(\boldsymbol{\eta}|\chi,\nu) = f(\chi,\nu) g(\boldsymbol{\eta})^{\nu} \exp\left\{\nu\boldsymbol{\eta}^{\mathrm{T}}\chi\right\}$

$$p(\boldsymbol{\eta}|\mathbf{X},\chi,\nu) \propto g(\boldsymbol{\eta})^{\nu+N} \exp\left\{\boldsymbol{\eta}^T \left(\sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n) + \nu\boldsymbol{\chi}\right)\right\}$$

事後分布をみると, 事前分布のパラメータ ν は
有効な事前の仮想観測値 χ の数だとみなせる

目次

2.4　指数型分布族

無情報事前分布

事後分布 ∝ **尤度関数** × **事前分布**

何らかの知識やその信念の度合いを事前分布で表現可能

↓ なにも知見がないときは？

事前分布の影響を抑えたい

↓

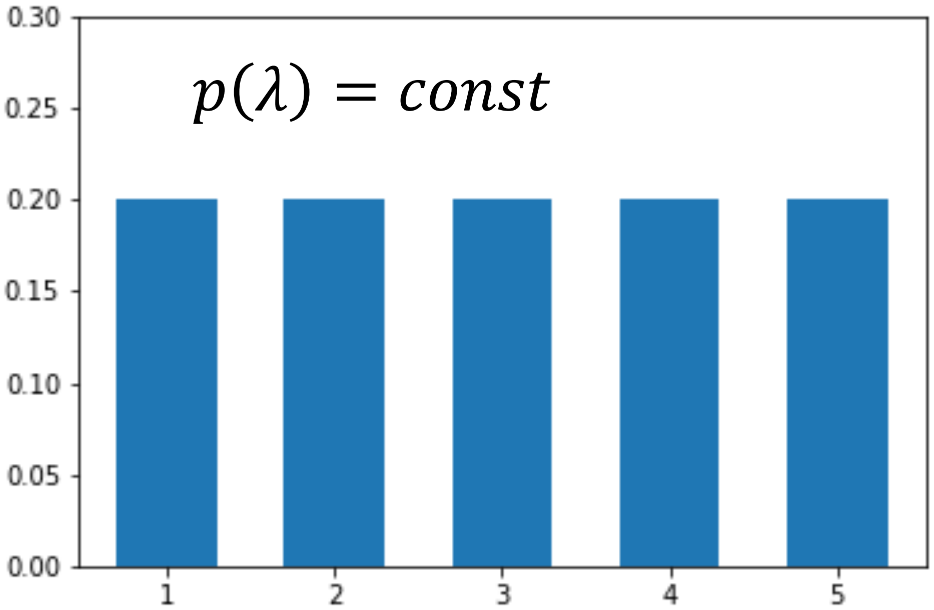
無情報事前分布

無情報事前分布

パラメータ λ で定められる分布 $p(x|\lambda)$ を考える

影響の少ない事前分布を選びたい

とりあえず一様分布に…　**離散変数であれば可能**



しかし

連続では二つの問題が…

無情報事前分布

連続パラメータλで $p(\lambda) = const$ を用いる際の**二つの問題**

1. λの定義域が有界でないとき、**λ上での積分が発散する** (正規化できない)

2. **非線形の変数変換をすると定数にならない**

無情報事前分布

変則事前分布　：正規化できない分布

得られる事後分布が**適切**であれば(正規化できるのであれば)
使われることが多い

無情報事前分布

変数変換にかかわる問題

関数:
$h(\lambda) = const$

確率密度:
$p_\lambda(\lambda) = const$

変数変換
$\lambda = \eta^2$

$h(\eta^2) = const$

問題なし

$$p_\lambda(\lambda) = p_\lambda(\lambda)\left|\frac{d\lambda}{d\eta}\right| = p_\lambda(\eta^2)2\eta \propto \eta$$

定数じゃない

最尤推定ではこの問題は生じない

定数の事前分布を用いる際にはパラメータは
適切に表現できるよう注意して用いる必要がある

無情報事前分布

平行移動不変性

モデル(尤度関数)：　$p(x|\mu) = f(x-\mu)$

μ　：位置パラメータ

定数分だけ移動した　$\hat{x} = x + c$　においても、

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

ただし　$\hat{\mu} \equiv \mu + c$

新しい変数においても確率密度は形状が同じ

無情報事前分布

平行移動不変性

事後分布 $\propto$ **尤度関数$\boldsymbol{p(x|\mu)}$** $\times$ **事前分布$\boldsymbol{p(\mu)}$**

事前分布にも平行移動不変性を

任意の定数cに対してμが区間 $A \leq \mu \leq B$ に入る確率と

区間 $A - c \leq \mu \leq B - c$ に入る確率が等しくなければならない

無情報事前分布

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

任意のAとBについて上記が成立することから次が得られる

$$p(\mu - c) = p(\mu)$$

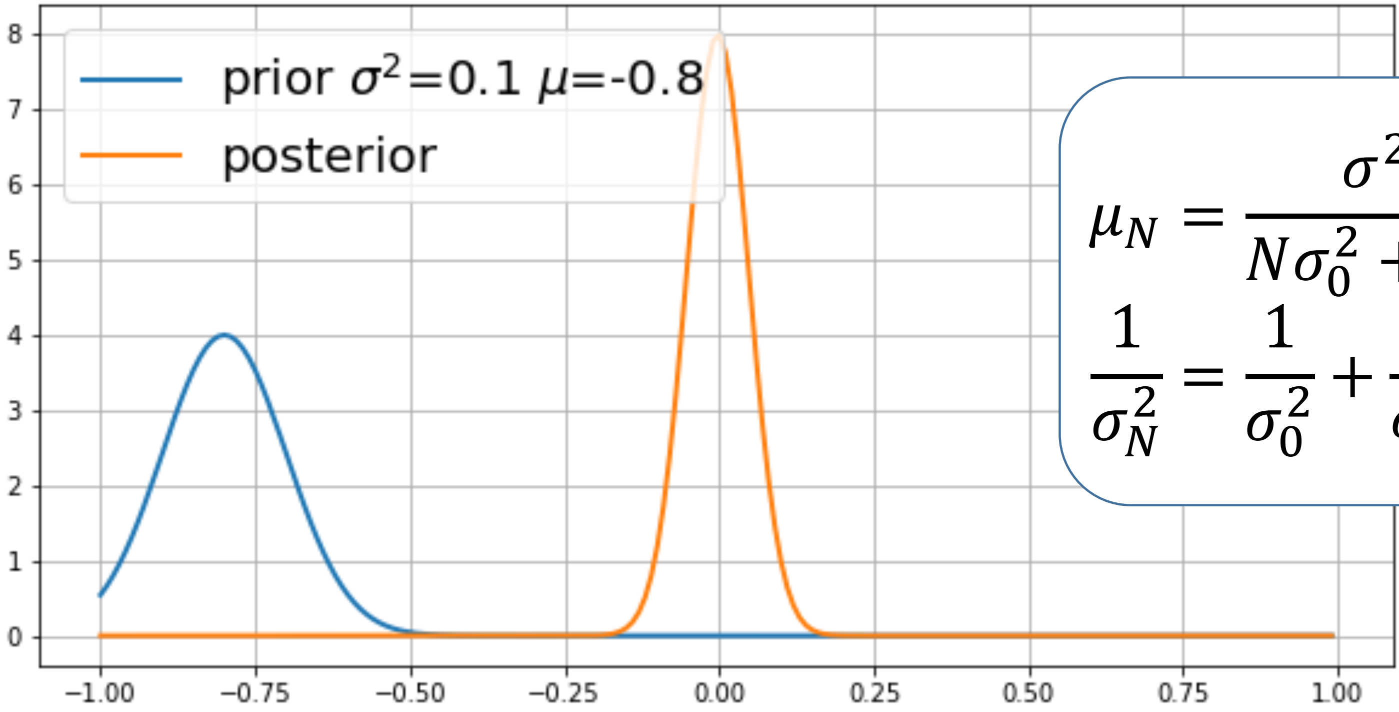
$$\Leftrightarrow p(\mu) = const$$

位置パラメータの例
ガウス分布の平均パラメータμなど

無情報事前分布

まずい共役事前分布を選んだとき

分散を既知としたガウス分布の
平均μに対するベイズ推論
事前分布もガウス分布
データは分散0.1、平均0.8の
ガウス分布から生成される



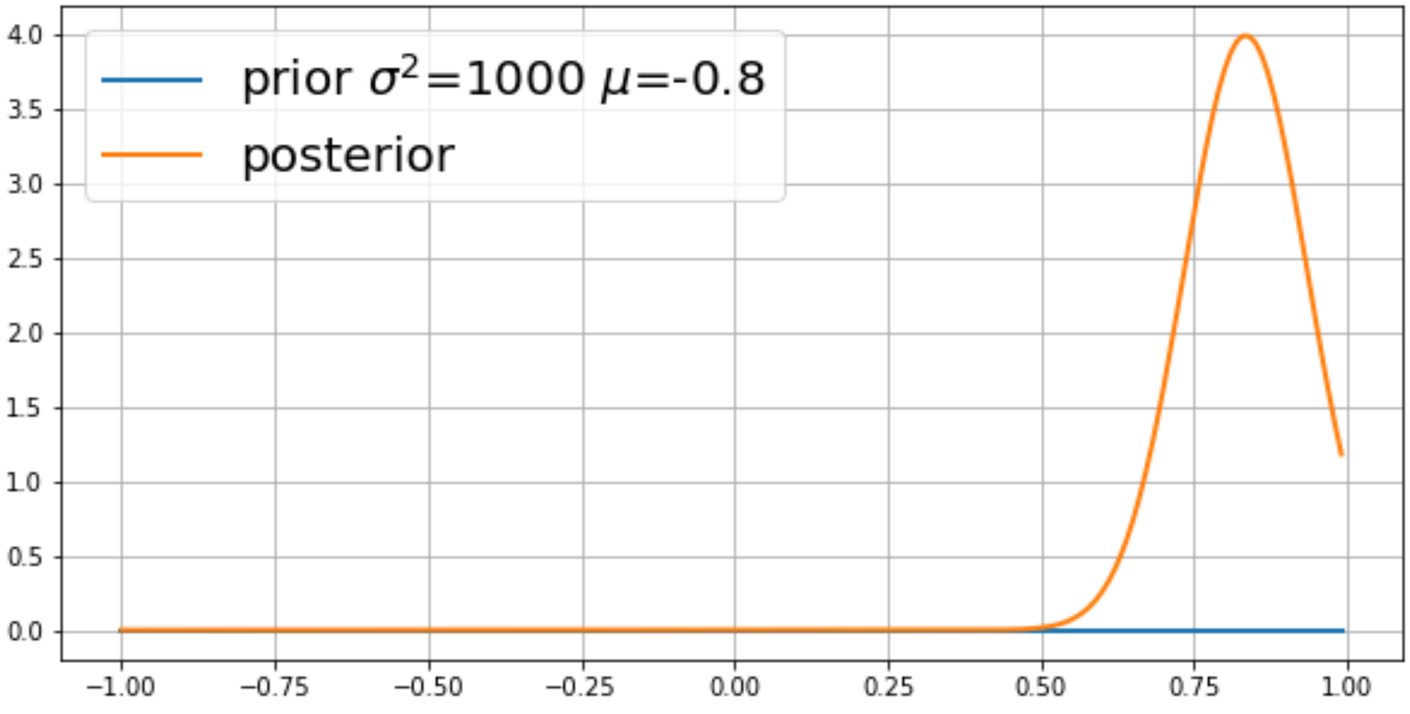
$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

無情報事前分布

無情報事前分布を選んだとき

同様のデータ生成モデルを使用しているが
事後分布に事前分布の影響がない



無情報事前分布

尺度不変性

モデル(尤度関数)：　$p(x|\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$

σ　：**尺度パラメータ**

定数倍だけ拡大縮小した　$\hat{x} = cx$　においても、

$$p(\hat{x}|\hat{\sigma}) = \frac{1}{\hat{\sigma}} f\left(\frac{\hat{x}}{\hat{\sigma}}\right)$$

ただし　$\hat{\sigma} \equiv c\sigma$

新しい変数においても
確率密度は形状が同じ

[m]から[km]への
尺度の変換などともとれる

無情報事前分布　　　再掲

尺度不変性

事後分布 $\propto$ 尤度関数$p(x|\sigma)$ $\times$ 事前分布$p(\sigma)$

事前分布にも尺度不変性を

任意の定数cに対してμが区間 $A \leq \sigma \leq B$ 　に入る確率と

区間　$A/c \leq \sigma \leq B/c$ 　に入る確率が等しくなければならない

無情報事前分布

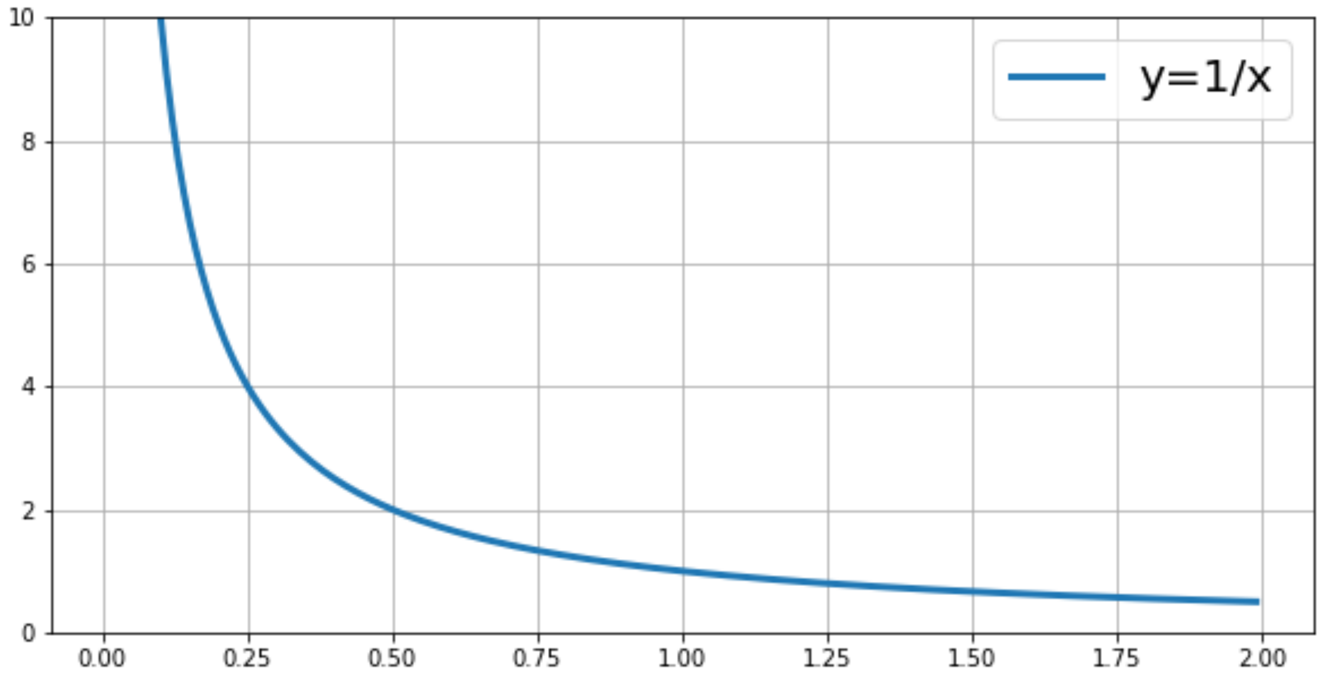
$$\int_A^B p(\sigma)d\sigma = \int_{A/c}^{B/c} p(\sigma)d\sigma = \int_A^B p\left(\frac{1}{c}\sigma\right)\frac{1}{c}d\sigma$$

任意のAとBについて上記が成立することから次が得られる

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

$$\Leftrightarrow p(\sigma) \propto \frac{1}{\sigma}$$

積分が発散するため
変則事前分布



無情報事前分布

パラメータの対数を考えたほうが便利なことも

新しい変数 η **：** $\eta = \ln \sigma$

(1.27)より $$p_\sigma(\sigma) = p_\eta(\eta)\left|\frac{d\eta}{d\sigma}\right| = p_\eta(\ln \sigma)\frac{1}{\sigma}$$

さらに、 $$p(\sigma) \propto \frac{1}{\sigma}$$

よって $$p(\ln \sigma) = const$$

$10 \leq \sigma \leq 100$ **に入る確率と** $100 \leq \sigma \leq 1000$ **に入る確率は等しくなる**

無情報事前分布

尺度パラメータの例：
位置パラメータ考慮済みのガウス分布の標準偏差 σ

$$\mathcal{N}(x|\mu,\sigma^2) \propto \sigma^{-1} exp\{-(\tilde{x}/\sigma)^2\}$$ ただし、 $\tilde{x} = x - \mu$

精度をつかうなら(1.27)から

$$\lambda = 1/\sigma^2, \qquad p_\sigma(\sigma) = p_\lambda(\lambda)\left|\frac{d\lambda}{d\sigma}\right| = -p_\lambda(\lambda)\frac{3}{\sigma^3} \propto \frac{1}{\sigma}$$

$$-p_\lambda(\lambda)\frac{3}{\sigma^2} = const$$

$$\Leftrightarrow p(\lambda) \propto \sigma^2 = \frac{1}{\lambda}$$

無情報事前分布

λ の共役事前分布は(2.146)のガンマ分布　$Gam(\lambda|a_0, b_0)$

$a_0 = b_0 = 0$ とすると無情報事前分布に

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

$$b_N = b_0 + \frac{N}{2}\sigma_{ML}^2$$

データ由来の項のみに依存することがわかる

(2.150)および(2.151)参照