ベイズ推論

3.3

3.3. 1次元ガウス分布の学習.

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

… パラメータは (μ,σ^2)

⇝ ① 平均値 μ のみを学習する.
② 分散 σ^2 のみを学習する.
③ 両方 (μ,σ^2) を学習する.

ここで, 精度 (precision) パラメータ $\lambda=\sigma^{-2}$ を導入する.

3.3.1. 平均が未知の場合.

仮定

データ : 正規分布, N個
… 精度パラメータ $\lambda\in\mathbb{R}_+$: 固定.
⇝ μ のみ事前分布を設定.

観測値 $x\in\mathbb{R}$ の分布

$$p(x|\mu)=\mathcal{N}(x|\mu,\lambda^{-1})\left(=\sqrt{\frac{\lambda}{2\pi}}\exp\left(-\frac{\lambda(x-\mu)^2}{2}\right)\right) \quad (3.47)$$

μの事前分布

$$p(\mu)=\mathcal{N}(\mu|m,\lambda_\mu^{-1})\left(=\sqrt{\frac{\lambda_\mu}{2\pi}}\exp\left(-\frac{\lambda_\mu(\mu-m)^2}{2}\right)\right) \quad (3.48)$$

where
$m\in\mathbb{R}$, $\lambda_\mu\in\mathbb{R}_+$ are hyper parameters.

Observed Data
$\mathbb{X}=\{x_1,\dots,x_N\}$

$$p(\mu|\mathbf{X}) \propto p(\mathbf{X}|\mu)\,p(\mu) \quad (\because \text{Bayes})$$

$$= \left(\prod_{i=1}^{N} p(x_i|\mu)\right) p(\mu)$$

$$= \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \lambda^{-1})\,\mathcal{N}(\mu|m, \lambda_\mu^{-1}) \quad (3.49)$$

対数をとる、

$$\ln p(\mu|\mathbf{X}) = \sum_{i=1}^{N} \ln \mathcal{N}(x_i|\mu, \lambda^{-1}) + \ln \mathcal{N}(\mu|m, \lambda_\mu^{-1}) + \text{const.}$$

$$= \sum_{i=1}^{N} \ln \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda(x_i-\mu)^2}{2}\right) + \ln \sqrt{\frac{\lambda_\mu}{2\pi}} \exp\left(-\frac{\lambda_\mu(\mu-m)^2}{2}\right) + \text{const.}$$

$$= \sum_{i=1}^{N}\left[\underline{\frac{1}{2}(\ln\lambda - \ln 2\pi)} - \frac{\lambda(x_i-\mu)^2}{2}\right] + \underline{\frac{1}{2}(\ln\lambda_\mu - \ln 2\pi)} - \frac{\lambda_\mu(\mu-m)^2}{2} + \underline{\text{const.}}$$

$$= -\frac{\lambda}{2}\sum_{i=1}^{N}(x_i^2 - 2x_i\mu + \mu^2) - \frac{\lambda_\mu}{2}(\mu^2 - 2\mu m + \underline{m^2}) + \underline{\text{const.}}$$

$$= -\frac{1}{2}\left[\underline{(N\lambda + \lambda_m)}\mu^2 - 2\left(\underline{\sum_{i=1}^{N} x_i\lambda + m\lambda_\mu}\right)\mu\right] + \text{const.}$$

(3.50)

$\rightsquigarrow$ μに関する上に凸の2次関数.

事後分布が2次元のガウス分布だとする。

$$p(\mu|X) = \mathcal{N}(\mu|\hat{m}, \hat{\lambda}_\mu^{-1}) \quad \left(= \sqrt{\frac{\hat{\lambda}_\mu}{2\pi}} \exp\left(- \frac{\hat{\lambda}_\mu(\mu-\hat{m})^2}{2} \right) \right) \tag{3.51}$$

対数をとると、

$$\ln p(\mu|X) = \underline{\frac{1}{2}(\ln\hat{\lambda}_\mu - \ln 2\pi)} - \frac{\hat{\lambda}_\mu}{2}(\mu - \hat{m})^2$$

$$= -\frac{1}{2}\left(\underline{\hat{\lambda}_\mu}\mu^2 - 2\underline{\hat{m}\hat{\lambda}_\mu}\mu\right) + \text{const}. \tag{3.52}$$

(3.50), (3.52) より、

$$\hat{\lambda}_\mu = N\lambda + \lambda_\mu \tag{3.53}$$

$$\hat{m} = \frac{\lambda \sum_{i=1}^N x_i + \lambda_\mu m}{\hat{\lambda}_\mu} \tag{3.54}$$

が得られる。

② 平均 μ の 事後分布の精度 $\hat{\lambda}_\mu$

… 事前分布の精度に $N\lambda$ だけ加えたもの

⇝ データ数 N が大きくなると事後分布の精度が大きくなる。

③ 平均 μ の事後分布の平均 $\hat{m}$

… 事前分布による知識 m と観測データの和 $\sum_{i=1}^N x_i$ の重み付き和。

⇒ データを観測するほど、事前分布の平均 m の影響はうすれ、データによって計算される平均値 $\frac{1}{N}\sum_{i=1}^N x_i$ が支配的。

λ_μ が小さいときもデータが支配的。

③ 未観測データ x_* に対する予測分布。

$$p(x_*) = \int \overbrace{p(x_* \mid \mu)\, p(\mu)}^{p(x_*,\, \mu)}\, d\mu$$

$$= \int \mathcal{N}(x_* \mid \mu, \lambda^{-1})\, \mathcal{N}(\mu \mid m, \lambda_\mu^{-1})\, d\mu \qquad (3.55)$$

→ 対数を使って計算する。

・予測分布 $p(x_*)$ と事前分布 $p(\mu)$ の間の関係。

$$p(\mu \mid x_*) = \frac{p(x_* \mid \mu)\, p(\mu)}{p(x_*)} \quad (\because \text{Bayes}) \qquad (3.56)$$

対数をとり、$p(x_*)$ に関して解く。

$$\ln p(\mu \mid x_*) = \ln p(x_* \mid \mu) + \ln p(\mu) - \ln p(x_*)$$

$$\therefore \ln p(x_*) = \ln p(x_* \mid \mu) - \ln p(\mu \mid x_*) + \text{const.} \qquad (3.57)$$

(3.57) において、$p(\mu \mid x_*)$ は、(3.51) の $N \leftarrow 1$ としたものに等しい。

$$\therefore p(\mu \mid x_*) = \mathcal{N}(\mu \mid m(x_*), \bar{\lambda}) \qquad (3.58)$$

where　　　平均値は x_* の関数になっている。

$$\bar{\lambda} = \lambda + \lambda_\mu$$

$$m(x_*) = \frac{\lambda x_* + \lambda_\mu m}{\lambda + \lambda_\mu} \qquad (3.59)$$

さらに、$p(x_* \mid \mu)$ に関しては、(3.47) 式。

5.7.

$$\ln p(x_*) = \ln p(x_*|\mu) - \ln(\mu|x_*) + \text{const.}$$

$$= \ln \mathcal{N}(x_*|\mu, \lambda^{-1}) - \ln \mathcal{N}(\mu|m(x_*), \lambda+\lambda_\mu) + \text{const.}$$

$$= \ln\left[\sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda(x_*-\mu)^2}{2}\right)\right]$$

$$- \ln\left[\sqrt{\frac{\lambda+\lambda_\mu}{2\pi}} \exp\left(-\frac{(\lambda+\lambda_\mu)(\mu-m(x_*))^2}{2}\right)\right] + \text{const.}$$

$$= \underline{\frac{1}{2}(\ln\lambda - \ln 2\pi)} - \frac{\lambda}{2}(x_*-\mu)^2$$

$$\underline{-\frac{1}{2}(\ln(\lambda+\lambda_\mu)} - \ln 2\pi) - \frac{1}{2}(\lambda+\lambda_\mu)(\mu-m(x_*))^2$$

$$+ \text{const.}$$

$$= -\frac{1}{2}\left[\lambda(x_*-\mu)^2 - (\lambda+\lambda_\mu)(\mu-m(x_*))^2\right] + \text{const.}$$

$$= -\frac{1}{2}\left[\lambda(x_*^2 - 2x_*\mu + \mu^2) - (\lambda+\lambda_\mu)\left(\mu - \frac{\lambda x_* + \lambda_\mu m}{\lambda+\lambda_\mu}\right)^2\right]$$

$$+ \text{const.}$$

$$\lambda^2 x_*^2 + 2\lambda\lambda_\mu m x_* + \lambda_\mu^2 m^2$$

$$= -\frac{1}{2}\left[\lambda x_*^2 - \cancel{2\lambda\mu x_*} + \lambda\mu^2 - (\lambda+\lambda_\mu)\mu^2 + 2\mu(\cancel{\lambda x_*} + \lambda_\mu m) - \frac{(\lambda x_* + \lambda_\mu m)^2}{\lambda+\lambda_\mu}\right]$$

$$+ \text{const.}$$

$$= -\frac{1}{2}\left[\left(\lambda - \frac{\lambda^2}{\lambda+\lambda_\mu}\right)x_*^2 - \frac{2m\lambda\lambda_\mu}{\lambda+\lambda_\mu}x_*\right] + \text{const.}$$

$$= -\frac{1}{2}\left[\frac{\lambda\lambda_\mu}{\lambda+\lambda_\mu}x_*^2 - \frac{2m\lambda\lambda_\mu}{\lambda+\lambda_\mu}x_*\right] + \text{const}, \quad (3.60)$$

よって、予測分布 $p(x_*)$ は、

$$p(x_*) = \mathcal{N}(x_* | \mu_*, \lambda_*^{-1}) \quad (3.61)$$

where

$$\begin{cases} \lambda_* = \dfrac{\lambda\lambda_\mu}{\lambda+\lambda_\mu} \\ \mu_* = m \end{cases} \quad (3.62)$$

$$\ln \mathcal{N}(x_* | \mu_*, \lambda_*^{-1}) = -\frac{1}{2}[\lambda_* x_*^2 - 2\mu_*\lambda_* x_*] + \text{const.}$$

より、

$$\lambda_* = \frac{\lambda\lambda_\mu}{\lambda+\lambda_\mu}, \quad \mu_*\lambda_* = \frac{m\lambda\lambda_\mu}{\lambda+\lambda_\mu}$$

ここで、

$$\frac{1}{\lambda_*} = \frac{\lambda+\lambda_\mu}{\lambda\lambda_\mu} = \frac{1}{\lambda_\mu} + \frac{1}{\lambda} \quad (3.63)$$

⇒ 予測分布の平均パラメータ：事前分布の平均
　　〃　　不確かさ　：観測分布と事前分布の不確かさの和

☆ データ観測後の予測分布 $p(x_* | X)$
　… $m \leftarrow \hat{m}$, $\lambda_\mu \leftarrow \hat{\lambda}_\mu$

このとき、学習後の予測分布の分散は、$\lambda^{-1} + \lambda_\mu^{-1}$.

◎ パラメータ μ に関する精度 λ_μ は
データが増えるほど大きくなる
but
データの生成自体に関する精度 λ は変わらない。

⇝ λ の分布も学習する場合、ここで事前分布 $p(\lambda)$ を
モデルに追加して推論。

3.3.2. 精度が未知

仮定

- μ : 既知
- λ : 未知 ⇒ データから学習
  $\mathcal{X} = \{x_1, \ldots, x_N\}$

観測モデル

$$p(x \mid \underline{\underline{\lambda}}) = \mathcal{N}(x \mid \mu, \lambda^{-1}) \left(= \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda(x-\mu)^2}{2} \right) \right) \tag{3.64}$$

変数として明示

λ の事前分布 : ガンマ分布

$$p(\lambda) = \mathrm{Gam}(\lambda \mid a, b) \left(= \underbrace{C_G(a, b)}_{\frac{b^a}{\Gamma(a)}} \lambda^{a-1} e^{-b\lambda} \right) \tag{3.65}$$

⇝ ガンマ分布はガウス分布の精度パラメータの共役事前分布。

以下、事後分布を求めることでこれを示す。

$$\begin{aligned} p(\lambda \mid \mathcal{X}) &\propto p(\mathcal{X} \mid \lambda)\, p(\lambda) \\ &= \prod_{i=1}^{N} p(x_i \mid \lambda)\, p(\lambda) \\ &= \prod_{i=1}^{N} \mathcal{N}(x_i \mid \mu, \lambda^{-1})\, \mathrm{Gam}(\lambda \mid a, b) \end{aligned} \tag{3.66}$$

対数をとる。

$$\ln p(\lambda|\mathbf{X}) = \sum_{i=1}^{N} \ln\left(\sqrt{\frac{\lambda}{2\pi}}\exp\left(-\frac{\lambda(x_i-\mu)^2}{2}\right)\right) + \ln\left(C_G(a,b)\,\lambda^{a-1}e^{-b\lambda}\right) + \text{const.}$$

$$= \frac{N}{2}(\ln\lambda - \underline{\ln 2\pi}) - \sum_{i=1}^{N}\frac{\lambda(x_i-\mu)^2}{2} + \underline{\ln C_G(a,b)} + (a-1)\ln\lambda - b\lambda + \underline{\text{const.}}$$

$$= \left(\frac{N}{2} + a - 1\right)\ln\lambda - \left(\frac{1}{2}(x_i-\mu)^2 + b\right)\lambda + \text{const.} \quad (3.67)$$

ここで、$\mathrm{Gam}(\lambda|\hat{a},\hat{b})$ の対数は、

$$\ln \mathrm{Gam}(\lambda|\hat{a},\hat{b}) = (\hat{a}-1)\ln\lambda - \hat{b}\lambda + \text{const.}$$

であることを考えると、事後分布 $p(\lambda|\mathbf{X})$ は、

$$p(\lambda|\mathbf{X}) = \mathrm{Gam}(\lambda|\hat{a},\hat{b}) \quad (3.68)$$

where

$$\begin{cases} \hat{a} = \frac{N}{2} + a \\ \hat{b} = \frac{1}{2}(x_i-\mu)^2 + b \end{cases} \quad (3.69)$$

次に、予測分布を考える。事前分布を利用した予測分布は、

$$p(x_*) = \int p(x_*|\lambda)\,p(\lambda)\,d\lambda. \quad (3.70)$$

対数をとることで簡易的に計算できないか検討する。

ベイズの定理より、

$$p(\lambda|x_*) = \frac{p(x_*|\lambda)\,p(\lambda)}{p(x_*)} \quad (3.71)$$

対数をとると、

$$\ln p(\lambda|x_*) = \ln p(x_*|\lambda) + \ln p(\lambda) - \ln p(x_*)$$

$p(x_*)$ について考える。

$$\ln p(x_*) = \ln p(x_*|\lambda) - \ln p(\lambda|x_*) + \text{const.} \quad (3.72)$$

ここで，$p(\lambda|x_*)$ を 1点 x_* を観測した後の事後分布と考えると，

$$p(\lambda|x_*) = \mathrm{Gam}\left(\lambda \middle| \frac{1}{2}+a,\ b(x_*)\right) \qquad (3.73)$$

where $\qquad (\because (3.69))$

$$b(x_*) = \frac{1}{2}(x_* - \mu)^2 + b \qquad (3.74)$$

よって，

$$\begin{aligned}
\ln p(x_*) &= \ln p(x_*|\lambda) - \ln p(\lambda|x_*) + \text{const.} \\
&= \ln \mathcal{N}(x_*|\mu, \lambda^{-1}) - \ln \mathrm{Gam}\left(\lambda \middle| \frac{1}{2}+a,\ b(x_*)\right) + \text{const.} \\
&= \underline{\frac{1}{2}(\ln\lambda - \ln 2\pi)} - \frac{\lambda}{2}(x_* - \mu)^2 \\
&\quad - \left[\ln C_G\left(\frac{1}{2}+a,\ b(x_*)\right) + \underline{\left(\frac{1}{2}+a-1\right)\ln\lambda} - b(x_*)\lambda\right] \underline{+\text{const.}} \\
&= -\cancel{\frac{\lambda}{2}(x_* - \mu)^2} - \left[\cancel{\frac{1}{2}(x_* - \mu)^2} + \underline{b}\right]\underline{\lambda} \\
&\quad - \left(\ln\left(b + \frac{1}{2}(x_* - \mu)^2\right)^{\frac{1}{2}+a} \underline{- \ln\Gamma\left(\frac{1}{2}+a\right)}\right) \\
&\qquad \underline{+\text{const.}} \\
&= -\frac{2a+1}{2} \ln b\left(1 + \frac{1}{2b}(x_* - \mu)^2\right) + \text{const.} \\
&= -\frac{2a+1}{2} \ln\left(1 + \frac{1}{2b}(x_* - \mu)^2\right) + \text{const.} \qquad (3.75)
\end{aligned}$$

〜) スチューデントのt分布の対数，

$$\mathrm{St}(x|\mu_s, \lambda_s, \nu_s) = \frac{\Gamma\left(\frac{\nu_s+1}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)} \left(\frac{\lambda_s}{\pi\nu_s}\right)^{\frac{1}{2}} \left[1 + \frac{\lambda_s}{\nu_s}(x - \mu_s)^2\right]^{-\frac{\nu_s+1}{2}}$$

$$(3.76)$$

(3.76) の対数をとる。

$$
\begin{aligned}
&\ln \mathrm{St}(x \mid \mu_s, \lambda_s, \nu_s) \\
&= \ln \Gamma\left(\frac{\nu_s+1}{2}\right) - \ln \Gamma\left(\frac{\nu_s}{2}\right) + \frac{1}{2}\ln\left(\frac{\lambda_s}{\pi \nu_s}\right) \\
&\qquad - \frac{\nu_s+1}{2}\ln\left[1 + \frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right] \\
&= -\frac{\nu_s+1}{2}\ln\left[1 + \frac{\lambda_s}{\nu_s}(x-\mu_s)^2\right] + \text{const.} \qquad (3.77)
\end{aligned}
$$

(3.75), (3.77) を比較する。

$$
p(x_*) = \mathrm{St}(x_* \mid \mu_s, \lambda_s, \nu_s) \qquad (3.78)
$$

where

$$
\begin{cases} \nu_s = 2a \\ \lambda_s / \nu_s = 1/2b \\ \mu_s = \mu \end{cases} \quad \therefore \quad \begin{cases} \mu_s = \mu \\ \lambda_s = a/b \\ \nu_s = 2a \end{cases} \qquad (3.79)
$$

3.3.3. 平均・精度が未知の場合。

観測モデル

$$
p(x \mid \mu, \lambda) = \mathcal{N}(x \mid \mu, \lambda^{-1}) \left(= \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda(x-\mu)^2}{2}\right) \right) \qquad (3.80)
$$

→ 事前分布は ガウス・ガンマ分布

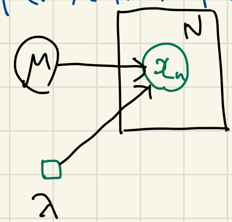
$$
\begin{aligned}
p(\mu, \lambda) &= \mathrm{NG}(\mu, \lambda \mid m, \beta, a, b) \\
&= \mathcal{N}(\mu \mid m, (\beta\lambda)^{-1})\, \mathrm{Gam}(\lambda \mid a, b) \qquad (3.81)
\end{aligned}
$$

モデルは、

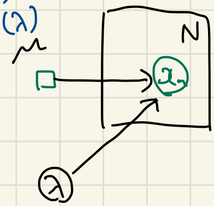
$$
p(X, \mu, \lambda) = p(X \mid \mu, \lambda)\, p(\mu, \lambda)
$$

1次元ガウス分布の学習モデル

$p(x,\lambda,\mu) = p(x|\mu,\lambda)\,p(\mu)p(\lambda)$

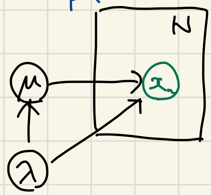


。平均未知



。精度未知

$p(x,\mu,\lambda)$
$= p(x|\mu,\lambda)\,p(\mu|\lambda)\,p(\lambda)$
$\| \; p(x|\mu,\lambda)\,p(\mu,\lambda)$



。平均・精度未知.
平均パラメータも
精度パラメータに依存する.

事後分布を求める.

① μのみに着目する.
μの事前分布が
$p(\mu) = \mathcal{N}(\mu | m, (\beta\lambda)^{-1})$
で与えられるとして, 実際に確かめる.

$$p(\mu|\underline{\underline{\lambda}}, X) \propto p(X|\mu,\underline{\underline{\lambda}})\,p(\mu)$$

fixと思う　　　　fixと思う

$$= \underbrace{\prod_{i=1}^{N} \mathcal{N}(x_i|\mu,\lambda^{-1})\,\mathcal{N}(\mu|m,(\beta\lambda)^{-1})}_{}$$

(3.49)において $\lambda_\mu \leftarrow (\beta\lambda)$ とした形.

よって, 事後分布は, (3.51), (3.53), (3.54)より,

◎ $$p(\mu|\lambda, X) = \mathcal{N}(\mu|\hat{m}, \hat{\lambda}^{-1}) = \mathcal{N}(\mu|\hat{m}, (\hat{\beta}\lambda)^{-1}) \quad (3.82)$$

where

$$\hat{\lambda} = N\lambda + \beta\lambda = (N+\beta)\lambda, \quad \hat{\beta} = N + \beta$$

$$\hat{m} = \frac{\lambda\sum_{i=1}^{N} x_i + m\beta\lambda}{(N+\beta)\lambda} = \frac{1}{\hat{\beta}}\left(\sum_{i=1}^{N} x_i + \beta m\right) \quad (3.83)$$

である.

② λ の事後分布

同時分布を条件付分布の積として書き下す。

$$p(\mathbf{X}, \lambda, \mu) = p(\mu \mid \lambda, \mathbf{X})\, p(\lambda \mid \mathbf{X})\, p(\mathbf{X}) \qquad (3.84)$$

これより、λ の事後分布は、

$$p(\lambda \mid \mathbf{X}) \propto \frac{p(\mathbf{X}, \lambda, \mu)}{p(\mu \mid \lambda, \mathbf{X})\, p(\mathbf{X})}$$

$$\propto \frac{p(\mathbf{X}, \lambda, \mu)}{p(\mu \mid \lambda, \mathbf{X})} \qquad (3.85)$$

ここで、同時分布は

$$p(\mathbf{X}, \mu, \lambda) = p(\mathbf{X} \mid \mu, \lambda)\, p(\mu, \lambda)$$

である。

よって、(3.85) の対数をとると、

$$\ln p(\lambda \mid \mathbf{X})$$

$$= \ln p(\mathbf{X} \mid \mu, \lambda) + \ln p(\mu, \lambda) - \ln p(\mu \mid \lambda, \mathbf{X}) + \text{const.}$$

$$= \underbrace{\sum_{i=1}^{N} \ln \mathcal{N}(x_i \mid \mu, \lambda^{-1})}_{(3.80)} + \underbrace{\ln \mathcal{N}(\mu \mid m, (\beta\lambda)^{-1}) + \ln \mathrm{Gam}(\lambda \mid a, b)}_{(3.81)}$$

$$- \underbrace{\ln \mathcal{N}(\mu \mid \hat{m}, (\hat{\beta}\lambda)^{-1})}_{(3.82)} + \text{const},$$

$$= \sum_{i=1}^{N} \left[\frac{1}{2}(\ln\lambda - \ln 2\pi) - \frac{1}{2}\lambda(x_i - \mu)^2 \right]$$

$$+ \frac{1}{2}\left(\overset{\ln\beta + \cancel{\ln\lambda}}{\ln(\beta\lambda)} - \ln 2\pi\right) - \frac{1}{2}\beta\lambda(\mu - m)^2 + (a-1)\ln\lambda - b\lambda + \ln C_G(a,b)$$

$$- \left[\frac{1}{2}\left(\overset{\ln\hat{\beta} + \cancel{\ln\lambda}}{\ln(\hat{\beta}\lambda)} - \ln 2\pi\right) - \frac{1}{2}\hat{\beta}\lambda(\mu - \hat{m})^2 \right] + \text{const.}$$

$$= \left(\frac{N}{2} + a - 1\right)\ln\lambda$$

$$+ \left[-\frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^2 - \frac{1}{2}\beta(\mu - m)^2 - b + \frac{1}{2}\hat{\beta}(\mu - \hat{m})^2 \right]\lambda$$

$$+ \boxed{\text{const.}}$$

ここで、

$$-\frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$= -\frac{1}{2}\sum_{i=1}^{N}(x_i^2 - 2\mu x_i + \mu^2) = -\frac{1}{2}\sum_{i=1}^{N} x_i^2 + \cancel{\mu\sum_{i=1}^{N} x_i} \cancel{- \frac{1}{2}N\mu^2}$$

$$-\frac{1}{2}\beta(\mu - m)^2 - b$$

$$= -\frac{1}{2}\beta(\mu^2 - 2\mu m + m^2) - b$$

$$= \cancel{-\frac{1}{2}\beta\mu^2} + \cancel{\beta\mu m} - \frac{1}{2}\beta m^2 - b$$

$$\frac{1}{2}\hat{\beta}(\mu - \hat{m})^2$$

$$= \frac{1}{2}\hat{\beta}(\mu^2 - 2\mu\hat{m} + \hat{m}^2)$$

$$= \frac{1}{2}\hat{\beta}\mu^2 - \hat{\beta}\mu \cdot \frac{1}{\hat{\beta}}\left(\sum_{i=1}^{N} x_i + \beta m\right) + \frac{1}{2}\hat{\beta}\hat{m}^2$$

$$= \cancel{\frac{1}{2}\hat{\beta}\mu^2} \cancel{- \mu\sum_{i=1}^{N} x_i} \cancel{- \beta\mu m} + \frac{1}{2}\hat{\beta}\hat{m}^2$$

$$(\because \hat{\beta} = N + \beta)$$

なので、λの1次は、

$$\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2 - \frac{1}{2}\beta m^2 + \frac{1}{2}\hat{\beta}\hat{m}^2 - b\right)\lambda$$

$$= -\left[\frac{1}{2}\left(\sum_{i=1}^{N} x_i^2 + \beta m^2 - \hat{\beta}\hat{m}^2\right) + b\right]\lambda$$

以上より、

$$\ln p(\lambda \mid \mathbf{X}) = \left(\frac{N}{2} + a - 1\right)\ln\lambda - \left[\frac{1}{2}\left(\sum_{i=1}^{N} x_i^2 + \beta m^2 - \hat{\beta}\hat{m}^2\right) + b\right]\lambda + \text{const.} \quad (3.86)$$

これは、ガンマ分布の対数と同じ形。

$$\left(\ln \mathrm{Gam}(\lambda \mid \bar{a}, \bar{b}) = (\bar{a} - 1)\ln\lambda - \bar{b}\lambda + \text{const.}\right)$$

よって、事後分布は、

$$p(\lambda \mid \mathbf{X}) = \mathrm{Gam}(\lambda \mid \hat{a}, \hat{b}) \quad (3.87)$$

where

$$\hat{a} = \frac{N}{2} + a$$

$$\hat{b} = \frac{1}{2}\left(\sum_{i=1}^{N} x_i^2 + \beta m^2 - \hat{\beta}\hat{m}^2\right) + b \quad (3.88)$$

$\underset{(3.82)}{p(\mu \mid \mathbf{X}, \lambda)}$, $\underset{(3.86)}{p(\lambda \mid \mathbf{X})}$ は、事後分布が (3.81) の事前分布と同じ形式。

⇝ ハイパーパラメータ m, β, a, b が観測データによって更新される形。

③ 予測分布の計算.

$$p(x_*) = \iint p(x_* \mid \mu, \lambda)\, p(\mu, \lambda)\, d\mu\, d\lambda \qquad (3.89)$$

~> ベイズの定理を使って $p(x_*)$ を求められないか?

$$p(\mu, \lambda \mid x_*) = \frac{p(x_* \mid \mu, \lambda)\, p(\mu, \lambda)}{p(x_*)}$$

より、対数をとると、

$$\ln p(\mu, \lambda \mid x_*) = \ln p(x_* \mid \mu, \lambda) + \ln p(\mu, \lambda) - \ln p(x_*)$$

$$\therefore \ln p(x_*) = \ln p(x_* \mid \mu, \lambda) - \ln p(\mu, \lambda \mid x_*) + \mathrm{const.} \qquad (3.90)$$

ここで、$p(\mu, \lambda \mid x_*)$ を、1点 x_* を観測した後の事後分布と考える。

$$p(\mu, \lambda \mid x_*) = \underbrace{p(\mu \mid x_*, \lambda)}_{\substack{(3.82)\text{式で} \\ X \leftarrow x_* \text{としたもの}}} \underbrace{p(\lambda \mid x_*)}_{\substack{(3.87)\text{式で} \\ X \leftarrow x_* \text{としたもの}}}$$

$$= \mathcal{N}\left(\mu \mid m(x_*), \{(1+\beta)\lambda\}^{-1}\right) \mathrm{Gam}\left(\lambda \,\middle|\, \frac{1}{2} + a, b(x_*)\right) \qquad (3.91)$$

ただし、

$$\hat{m} = \frac{x_* + \beta m}{1 + \beta} =: m(x_*)$$

$$\hat{b} = \frac{1}{2}\left(x_*^2 + \beta m^2 - (1+\beta) \cdot \frac{(x_* + \beta m)^2}{(1+\beta)^2}\right) + b$$

$$= \frac{1}{2}\left[x_*^2 + \beta m^2 - \frac{x_*^2 + 2\beta m x_* + \beta^2 m^2}{1+\beta}\right] + b$$

$$= \frac{1}{2}\left[\frac{(1+\beta-1)x_*^2 - 2\beta m x_* + (\beta(1+\beta) - \beta^2)m^2}{1+\beta}\right] + b$$

$$= \frac{\beta}{2(1+\beta)}(x_* - m)^2 + b =: b(x_*) \qquad (3.92)$$

(3.90) に (3.80), (3.91) を代入.

$$\ln p(x_*)$$

$$= \ln p(x_* \mid \mu, \lambda) - \ln (\mu, \lambda \mid x_*) + \mathrm{const.}$$

$$= \ln \mathcal{N}(x_* \mid \mu, \lambda)$$

$$- \ln \mathcal{N}(\mu \mid m(x_*), \{(1+\beta)\lambda\}^{-1})$$

$$- \ln \mathrm{Gam}\left(\lambda \,\middle|\, \frac{1}{2} + a,\ b(x_*)\right) + \mathrm{const.}$$

$$= \frac{1}{2}(\ln\lambda - \ln 2\pi) - \frac{1}{2}\lambda(x_* - \mu)^2$$

$$-\left[\frac{1}{2}(\ln(1+\beta)\lambda - \ln 2\pi) - \frac{1}{2}(1+\beta)\lambda(\mu - m(x_*))^2\right]$$

$$-\left(\left(\frac{1}{2}+a-1\right)\ln\lambda - b(x_*)\lambda + \ln C_G\left(\tfrac{1}{2}+a, b(x_*)\right) + \mathrm{const.}\right.$$

$$= -\frac{1}{2}\lambda(x_*^2 - 2\mu x_* + \mu^2) + \frac{1}{2}(1+\beta)\lambda\left(\mu - \frac{x_* + \beta m}{1+\beta}\right)^2$$

$$+ \frac{\beta}{2(1+\beta)}(x_* - m)^2\lambda + \left(\frac{1}{2}+a\right)\ln b(x_*) - \ln\Gamma\left(\frac{1}{2}+a\right) + \mathrm{const.}$$

ここで、

$$\frac{1}{2}(1+\beta)\lambda\left(\mu - \frac{x_* + \beta m}{1+\beta}\right)^2$$

$$= \frac{1}{2}(1+\beta)\lambda\left(\mu^2 - \frac{2(x_* + \beta m)}{1+\beta}\mu + \frac{(x_* + \beta m)^2}{(1+\beta)^2}\right)$$

$$= \frac{1}{2}(1+\beta)\lambda\mu^2 - \lambda\mu(x_* + \beta m) + \frac{\lambda(x_* + \beta m)^2}{2(1+\beta)}$$

$$= \frac{1}{2}(1+\beta)\lambda\mu^2 - \lambda\mu x_* - \lambda\beta\mu m + \frac{\lambda}{2(1+\beta)}(x_*^2 + 2\beta m x_* + \beta^2 m^2)$$

$$= \frac{1}{2}(1+\beta)\lambda\mu^2 - \lambda\mu x_* - \lambda\beta m + \frac{\lambda}{2(1+\beta)}x_*^2 + \frac{\beta m\lambda}{1+\beta}x_* + \frac{\beta^2\lambda m^2}{2(1+\beta)}$$

$$\frac{\beta}{2(1+\beta)}(x_*-m)^2$$
$$= \frac{\beta\lambda}{2(1+\beta)}(x_*^2 - 2x_*m + m^2)$$

$$\left(\frac{1}{2}+a\right)\ln b(x_*)$$
$$= \left(\frac{1}{2}+a\right)\ln\left[\frac{\beta}{2(1+\beta)}(x_*-m)^2 + b\right]$$

より、x_*に関する項で整理.

$$\frac{\lambda}{2(1+\beta)} + \frac{\beta\lambda}{2(1+\beta)} = \frac{1}{2}$$

$$\ln p(x_*)$$
$$= \cancel{-\frac{1}{2}\lambda x_*^2} + \cancel{\mu\lambda x_*}$$
$$\cancel{-\lambda\mu x_*} + \cancel{\frac{\lambda}{2(1+\beta)}x_*^2} + \cancel{\frac{\beta m\lambda}{1+\beta}x_*}$$
$$+ \cancel{\frac{\beta\lambda}{2(1+\beta)}x_*^2} - \cancel{\frac{\beta m\lambda}{1+\beta}x_*} + \left(\frac{1}{2}+a\right)\ln\left[1+\frac{\beta}{2(1+\beta)b}(x_*-m)^2\right]$$
$$+ \text{const.}$$
$$= \left(\frac{1}{2}+a\right)\ln\left[\frac{\beta}{2(1+\beta)b}(x_*-m)^2\right] + \text{const.} \qquad (3.93)$$

これは、t−分布の対数と同じ形.

$$p(x_*) = \mathrm{St}(x_* \mid \mu_s, \lambda_s, \nu_s) \qquad (3.94)$$

where
$$\mu_s = m$$
$$\lambda_s = \frac{\beta a}{(1+\beta)b} \qquad (3.95)$$
$$\nu_s = 2a$$

$\Rightarrow$ データを観測後の $p(x_* \mid X)$ は、事前分布の代わりに事後分布のパラメータを入れる.