3.4.3 平均・精度が未知の場合.

- 観測モデル $p(x|\mu,\Lambda) = \mathcal{N}(x|\mu,\Lambda^{-1})$.
- 共役事前分布は, 次のGauss-Wishart分布.:

$$p(\mu,\Lambda) = \mathrm{NW}(\mu,\Lambda|m,\beta,\nu,W)$$
$$:= \mathcal{N}(\mu|m,(\beta\Lambda)^{-1})\mathcal{W}(\Lambda|\nu,W).$$

- 事後分布の計算.

データ $\mathcal{X} = \{x_1,\cdots,x_N\}$ を観測したとする. Bayes' Thm. より,

$$p(\mu,\Lambda|\mathcal{X}) = \frac{p(\mathcal{X}|\mu,\Lambda)p(\mu,\Lambda)}{p(\mathcal{X})}.$$

また, $p(\mu,\Lambda|\mathcal{X}) = p(\mu|\Lambda,\mathcal{X})p(\Lambda|\mathcal{X})$ なので,

$$p(\mu|\Lambda,\mathcal{X})p(\Lambda|\mathcal{X}) = \frac{p(\mathcal{X}|\mu,\Lambda)p(\mu,\Lambda)}{p(\mathcal{X})}$$

$\therefore \log p(\Lambda|\mathcal{X})$

$$= \underbrace{\log p(\mathcal{X}|\mu,\Lambda)}_{\text{わかる}} + \underbrace{\log p(\mu,\Lambda)}_{\text{わかる}} - \underbrace{\log p(\mu|\Lambda,\mathcal{X})}_{\text{わからない}} + \text{const}.$$

→ $p(\mu|\Lambda,\mathcal{X})$ が求まれば, 上式から $p(\Lambda|\mathcal{X})$ も求まり,

ひいては $p(\mu,\Lambda|\mathcal{X}) = p(\mu|\Lambda,\mathcal{X})p(\Lambda|\mathcal{X})$ が計算できる.

多次元Gauss分布，平均未知の場合.

- $p(\boldsymbol{\mu}|\boldsymbol{\Lambda}, \mathcal{X})$ について. $p(\boldsymbol{\mu}|\boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\mu}|\mathbf{m}, (\beta\boldsymbol{\Lambda})^{-1})$ なので,

$$p(\boldsymbol{\mu}|\boldsymbol{\Lambda}, \mathcal{X}) = \mathcal{N}(\boldsymbol{\mu}|\hat{\mathbf{m}}, (\hat{\beta}\boldsymbol{\Lambda})^{-1}),$$

$$\hat{\beta} := N + \beta, \quad \hat{\mathbf{m}} := \frac{1}{\hat{\beta}}\left(\sum_{n=1}^{N} \mathbf{x}_n + \beta\mathbf{m}\right).$$

$\det(\hat{\beta}\boldsymbol{\Lambda}) = \hat{\beta}^D \det\boldsymbol{\Lambda}$.

$$\therefore \log p(\boldsymbol{\mu}|\boldsymbol{\Lambda}, \mathcal{X})$$

$$= -\frac{1}{2}\left((\boldsymbol{\mu}-\hat{\mathbf{m}})^T(\hat{\beta}\boldsymbol{\Lambda})(\boldsymbol{\mu}-\hat{\mathbf{m}}) - \log(\det(\hat{\beta}\boldsymbol{\Lambda}))\right) + \text{const.}$$

$$= -\frac{1}{2}\Big(\hat{\beta}(\boldsymbol{\mu}-\hat{\mathbf{m}})^T\boldsymbol{\Lambda}(\boldsymbol{\mu}-\hat{\mathbf{m}}) \underbrace{- D\log\hat{\beta}}_{\text{const.}} - \log(\det\boldsymbol{\Lambda})\Big) + \text{const.}$$

$$= -\frac{1}{2}\left(\hat{\beta}(\boldsymbol{\mu}-\hat{\mathbf{m}})^T\boldsymbol{\Lambda}(\boldsymbol{\mu}-\hat{\mathbf{m}}) - \log(\det\boldsymbol{\Lambda})\right) + \text{const.}$$

- $p(\boldsymbol{\Lambda}|\mathcal{X})$ について,

$\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$

$$\log p(\mathcal{X}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \sum_{n=1}^{N} \log p(\mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Lambda})$$

$$= \sum_{n=1}^{N}\left(-\frac{1}{2}\left((\mathbf{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Lambda}(\mathbf{x}_n-\boldsymbol{\mu}) - \log(\det\boldsymbol{\Lambda})\right)\right) + \text{const.}$$

$$= -\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Lambda}(\mathbf{x}_n-\boldsymbol{\mu}) + \frac{N}{2}\log(\det\boldsymbol{\Lambda}) + \text{const.}$$

$$\log p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \log\mathcal{N}(\boldsymbol{\mu}|\mathbf{m}, (\beta\boldsymbol{\Lambda})^{-1}) + \log\mathcal{W}(\boldsymbol{\Lambda}|\nu, \mathbf{W})$$

$$= -\frac{1}{2}\left((\boldsymbol{\mu}-\mathbf{m})^T(\beta\boldsymbol{\Lambda})(\boldsymbol{\mu}-\mathbf{m}) - \log(\det(\beta\boldsymbol{\Lambda}))\right)$$

$$+ \frac{\nu-D-1}{2}\log(\det\boldsymbol{\Lambda}) - \frac{1}{2}\mathrm{tr}(\mathbf{W}^{-1}\boldsymbol{\Lambda}) + \text{const.}$$

$$= -\frac{1}{2}\Big(\beta(\boldsymbol{\mu}-\mathbf{m})^T\boldsymbol{\Lambda}(\boldsymbol{\mu}-\mathbf{m}) \underbrace{- D\log\beta}_{\text{const.}} - \log(\det\boldsymbol{\Lambda})\Big)$$

$$+ \frac{\nu-D-1}{2}\log(\det\boldsymbol{\Lambda}) - \frac{1}{2}\mathrm{tr}(\mathbf{W}^{-1}\boldsymbol{\Lambda}) + \text{const.}$$

$$= -\frac{1}{2}\beta(\boldsymbol{\mu}-\boldsymbol{m})^T\Lambda(\boldsymbol{\mu}-\boldsymbol{m}) + \frac{\nu-D}{2}\log(\det\Lambda)$$
$$-\frac{1}{2}\mathrm{tr}(W^{-1}\Lambda) + \mathrm{const}.$$

$\therefore\ \log p(\Lambda|\mathcal{X})$

$$= -\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{x}_n-\boldsymbol{\mu})^T\Lambda(\boldsymbol{x}_n-\boldsymbol{\mu}) + \frac{N}{2}\log(\det\Lambda) \quad \leftarrow \log p(\mathcal{X}|\boldsymbol{\mu},\Lambda)$$
$$\left.\begin{aligned}&-\frac{1}{2}\beta(\boldsymbol{\mu}-\boldsymbol{m})^T\Lambda(\boldsymbol{\mu}-\boldsymbol{m}) + \frac{\nu-D}{2}\log(\det\Lambda)\\ &-\frac{1}{2}\mathrm{tr}(W^{-1}\Lambda)\end{aligned}\right\} \log p(\boldsymbol{\mu},\Lambda)$$
$$+\frac{1}{2}\left(\hat{\beta}(\boldsymbol{\mu}-\hat{\boldsymbol{m}})^T\Lambda(\boldsymbol{\mu}-\hat{\boldsymbol{m}}) - \log(\det\Lambda)\right) + \mathrm{const}. \quad \leftarrow \log p(\boldsymbol{\mu}|\Lambda,\mathcal{X})$$

$$= \frac{N+\nu-D-1}{2}\log(\det\Lambda)$$
$$-\frac{1}{2}\left(\sum_{n=1}^{N}\mathrm{tr}\left((\boldsymbol{x}_n-\boldsymbol{\mu})^T\Lambda(\boldsymbol{x}_n-\boldsymbol{\mu})\right) + \beta\,\mathrm{tr}\left((\boldsymbol{\mu}-\boldsymbol{m})^T\Lambda(\boldsymbol{\mu}-\boldsymbol{m})\right)\right.$$
$$\left. - \hat{\beta}\,\mathrm{tr}\left((\boldsymbol{\mu}-\hat{\boldsymbol{m}})^T\Lambda(\boldsymbol{\mu}-\hat{\boldsymbol{m}})\right) + \mathrm{tr}(W^{-1}\Lambda)\right) + \mathrm{const}.$$

$\mathrm{tr}(A+B) = \mathrm{tr}(A)+\mathrm{tr}(B)$,
$\mathrm{tr}(\boldsymbol{x}^TA\boldsymbol{y}) = \mathrm{tr}(\boldsymbol{y}\boldsymbol{x}^TA)$

$$= \frac{N+\nu-D-1}{2}\log(\det\Lambda)$$
$$-\frac{1}{2}\mathrm{tr}\left(\left(\sum_{n=1}^{N}(\boldsymbol{x}_n-\boldsymbol{\mu})(\boldsymbol{x}_n-\boldsymbol{\mu})^T + \beta(\boldsymbol{\mu}-\boldsymbol{m})(\boldsymbol{\mu}-\boldsymbol{m})^T\right.\right.$$
$$\left.\left. - \hat{\beta}(\boldsymbol{\mu}-\hat{\boldsymbol{m}})(\boldsymbol{\mu}-\hat{\boldsymbol{m}})^T + W^{-1}\right)\Lambda\right) + \mathrm{const}.$$

Wishart分布のpdfのlogをとったもの.

$$\therefore\ p(\Lambda|\mathcal{X}) = \mathcal{W}(\Lambda|\hat{\nu},\hat{W}), \quad \hat{\nu} := N+\nu,$$

$$\hat{W}^{-1}$$
$$:= \sum_{n=1}^{N}(\boldsymbol{x}_n-\boldsymbol{\mu})(\boldsymbol{x}_n-\boldsymbol{\mu})^T + \beta(\boldsymbol{\mu}-\boldsymbol{m})(\boldsymbol{\mu}-\boldsymbol{m})^T - \hat{\beta}(\boldsymbol{\mu}-\hat{\boldsymbol{m}})(\boldsymbol{\mu}-\hat{\boldsymbol{m}})^T + W^{-1}.$$

$$
\begin{aligned}
&= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^{\top} - \cancel{\left(\sum_{n=1}^{N} \mathbf{x}_n\right) \boldsymbol{\mu}^{\top}} - \cancel{\boldsymbol{\mu} \left(\sum_{n=1}^{N} \mathbf{x}_n\right)^{\top}} + \cancel{N \boldsymbol{\mu}\boldsymbol{\mu}^{\top}} \\
&\quad + \cancel{\beta \boldsymbol{\mu}\boldsymbol{\mu}^{\top}} - \cancel{\beta \boldsymbol{\mu} \mathbf{m}^{\top}} - \cancel{\beta \mathbf{m} \boldsymbol{\mu}^{\top}} + \beta \mathbf{m}\mathbf{m}^{\top} \\
&\quad - \cancel{(N+\beta) \boldsymbol{\mu}\boldsymbol{\mu}^{\top}} + \cancel{\hat{\beta} \boldsymbol{\mu} \hat{\mathbf{m}}^{\top}} + \cancel{\hat{\beta} \hat{\mathbf{m}} \boldsymbol{\mu}^{\top}} - \hat{\beta} \hat{\mathbf{m}} \hat{\mathbf{m}}^{\top} \quad + \mathbf{W}^{-1} \\
&= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^{\top} + \beta \mathbf{m}\mathbf{m}^{\top} - \hat{\beta} \hat{\mathbf{m}} \hat{\mathbf{m}}^{\top} + \mathbf{W}^{-1}.
\end{aligned}
$$

・ $p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \mathbf{X})$

$$
\begin{aligned}
&= p(\boldsymbol{\mu} \mid \boldsymbol{\Lambda}, \mathbf{X}) \, p(\boldsymbol{\Lambda} \mid \mathbf{X}) \\
&= \mathcal{N}(\boldsymbol{\mu} \mid \hat{\mathbf{m}}, (\hat{\beta}\boldsymbol{\Lambda})^{-1}) \, \mathcal{W}(\boldsymbol{\Lambda} \mid \hat{\nu}, \hat{\mathbf{W}}) \\
&= \mathrm{NW}(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \hat{\mathbf{m}}, \hat{\beta}, \hat{\nu}, \hat{\mathbf{W}}),
\end{aligned}
$$

$$
\hat{\beta} := N + \beta, \quad \hat{\mathbf{m}} := \frac{1}{\hat{\beta}} \left(\sum_{n=1}^{N} \mathbf{x}_n + \beta \mathbf{m} \right),
$$

$$
\hat{\nu} := N + \nu,
$$

$$
\hat{\mathbf{W}}^{-1} := \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^{\top} + \beta \mathbf{m}\mathbf{m}^{\top} - \hat{\beta} \hat{\mathbf{m}} \hat{\mathbf{m}}^{\top} + \mathbf{W}^{-1}.
$$

→ 事後分布も Gauss-Wishart 分布になっていることが確認できた.

◦ 予測分布の計算.

$$
p(\mathbf{x}_*) = \iint p(\mathbf{x}_* \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}) \, p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \, d\boldsymbol{\mu} \, d\boldsymbol{\Lambda}.
$$

→ これを直接計算せずに求める.

Bayes' Thm. より.

x_* についての項のみ残す.

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

$$\log p(x_* | \mu, \Lambda)$$

$$= -\frac{1}{2}\left((x_* - \mu)^T \Lambda (x_* - \mu) - \underbrace{\log(\det \Lambda)}_{\text{const.}}\right) + \text{const.}$$

$$= -\frac{1}{2}(x_* - \mu)^T \Lambda (x_* - \mu) + \text{const.}$$

また.

$$p(\mu, \Lambda | x_*) = \mathrm{NW}(\mu, \Lambda | m(x_*), 1+\beta, 1+\nu, W(x_*)),$$

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

$$W(x_*)^{-1}$$

$$:= x_* x_*^T + \beta m m^T - (1+\beta)\hat{m}\hat{m}^T + W^{-1}.$$

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

$$= \frac{1}{1+\beta}\left((\cancel{1} + \beta) x_* x_*^T + (\beta + \cancel{\beta^2}) m m^T - \cancel{x_* x_*^T} - \beta x_* m^T - \beta m x_*^T - \cancel{\beta^2 m m^T}\right)$$
$$+ W^{-1}$$

$$= \frac{\beta}{1+\beta}\left(x_* x_*^T - x_* m^T - m x_*^T + m m^T\right) + W^{-1}$$

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

であるから,

$$\log p(\mu, \Lambda \mid x_*)$$

$$= \log \mathcal{N}(\mu \mid m(x_*), ((1+\beta)\Lambda)^{-1}) + \log \mathcal{W}(\Lambda \mid 1+\nu, W(x_*)^{-1})$$

$$= -\frac{1}{2}\left((\mu - m(x_*))^T((1+\beta)\Lambda)(\mu - m(x_*)) - \underbrace{\log(\det((1+\beta)\Lambda))}_{\text{const.}}\right)$$

$$+ \underbrace{\frac{1+\nu-D-1}{2}\log(\det\Lambda)}_{\text{const.}} - \frac{1}{2}\mathrm{tr}\left(W(x_*)^{-1}\Lambda\right)$$

正規化定数も x_* に依存することに注意.

$$+ \log C_{\mathcal{W}}(1+\nu, W(x_*)) + \text{const.}$$

$$= -\frac{1}{2}(1+\beta)(\mu - m(x_*))^T\Lambda(\mu - m(x_*)) - \frac{1}{2}\mathrm{tr}\left(W(x_*)^{-1}\Lambda\right)$$

$$- \frac{1+\nu}{2}\log(\det W(x_*)) + \text{const.}$$

$$\therefore \quad \log p(x_*)$$

$$= -\frac{1}{2}(x_* - \mu)^T\Lambda(x_* - \mu)$$

$$+ \frac{1}{2}(1+\beta)(\mu - m(x_*))^T\Lambda(\mu - m(x_*)) + \frac{1}{2}\mathrm{tr}\left(W(x_*)^{-1}\Lambda\right)$$

$$+ \frac{1+\nu}{2}\log(\det W(x_*)) + \text{const.}$$

ここで,

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

$$+ \frac{1}{2}\mathrm{tr}\left(W(x_*)^{-1}\Lambda\right)$$

$$= -\frac{1}{2}\mathrm{tr}\left(\left((x_* - \mu)(x_* - \mu)^T - (1+\beta)(\mu - m(x_*))(\mu - m(x_*))^T - W(x_*)^{-1}\right)\Lambda\right)$$

$$
\begin{aligned}
= -\frac{1}{2}\mathrm{tr}\Bigg(\bigg(& x_* x_*^T - x_* \mu^T - \mu x_*^T + \underbrace{\mu\mu^T - (1+\beta)\mu\mu^T}_{\text{const.}} \\
&+ (1+\beta)\mu m(x_*)^T + (1+\beta) m(x_*)\mu^T - (1+\beta) m(x_*) m(x_*)^T \\
&- \frac{\beta}{1+\beta}\Big(x_* x_*^T - x_* m^T - m x_*^T + \underbrace{m m^T}_{\text{const.}}\Big) + \underbrace{W^{-1}}_{\text{const.}}\bigg)\Lambda\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
= -\frac{1}{2}\mathrm{tr}\Bigg(\bigg(& \cancel{x_* x_*^T} - \cancel{x_* \mu^T} - \cancel{\mu x_*^T} + \mu\big(\cancel{x_*} + \underbrace{\beta m}_{\text{const.}}\big)^T + \big(\cancel{x_*} + \underbrace{\beta m}_{\text{const.}}\big)\mu^T \\
&- \frac{1}{1+\beta}\Big(\cancel{x_* x_*^T} + \cancel{\beta x_* m^T} + \cancel{\beta m x_*^T} + \underbrace{\beta^2 m m^T}_{\text{const.}}\Big) \\
&- \frac{\beta}{1+\beta}\Big(\cancel{x_* x_*^T} - \cancel{x_* m^T} - \cancel{m x_*^T}\Big)\bigg)\Lambda\Bigg) + \text{const.}
\end{aligned}
$$

$$= \text{const.}$$

ゆえに,

$$\log p(x_*)$$

$$= \frac{1+\nu}{2}\log\left(\det W(x_*)\right) + \text{const.}$$

$$= -\frac{1+\nu}{2}\log\left(\det W(x_*)\right)^{-1} + \text{const.}$$

$(\det A)^{-1} = \det(A^{-1})$

$$= -\frac{1+\nu}{2}\log\left(\det W(x_*)^{-1}\right) + \text{const.}$$

$$= -\frac{1+\nu}{2}\log\left(\det\left(\frac{\beta}{1+\beta}(x_* - m)(x_* - m)^T + W^{-1}\right)\right) + \text{const.}$$

W^{-1}のくくり出し

$$= -\frac{1+\nu}{2}\log\left(\det\left(\left(\frac{\beta}{1+\beta}(x_* - m)(x_* - m)^T W + I_D\right)W^{-1}\right)\right) + \text{const}$$

$\det(AB) = (\det A)(\det B)$

$$= -\frac{1+\nu}{2}\log\left(\det\left(\frac{\beta}{1+\beta}(x_* - m)(x_* - m)^T W + I_D\right)\det W^{-1}\right) + \text{const.}$$

$$= -\frac{1+\nu}{2}\log\left(\det\left(I_D + \frac{\beta}{1+\beta}(x_* - m)(x_* - m)^T W\right)\right)$$

$$\underbrace{-\frac{1+\nu}{2}\log\left(\det W^{-1}\right)}_{\text{const.}} + \text{const.}$$

$$= -\frac{1+\nu}{2}\log\left(\det\left(I_D + \frac{\beta}{1+\beta}(x_* - m)(x_* - m)^T W\right)\right) + \text{const.}$$

detの公式

$$= -\frac{1+\nu}{2}\log\left(\det\left(I_1 + \frac{\beta}{1+\beta}(x_* - m)^T W^T (x_* - m)\right)\right) + \text{const.}$$

1次正方行列の det は要素そのもの. $W^T = W$.

$$= -\frac{1+\nu}{2}\log\left(1 + \frac{\beta}{1+\beta}(x_* - m)^T W (x_* - m)\right) + \text{const.}$$

Studentのt分布のpdfのlogをとったもの

$$\therefore\ p(x_*) = \mathrm{St}(x_* \mid \mu_s, \Lambda_s, \nu_s),$$

$$\mu_s := m,\quad \Lambda_s := \frac{(1-D-\nu)\beta}{1+\beta}W,\quad \nu_s := 1 - D + \nu.$$