§3.4 多次元Gauss分布の学習と予測

D次元の多次元Gauss分布のパラメータの学習をする.

$\Lambda := \Sigma^{-1}$ (精度行列) とする.

3.4.1 平均が未知の場合.

$\boldsymbol{\mu} \in \mathbb{R}^D$: unknown. $\Lambda \in M_D(\mathbb{R})$, $\Lambda > 0$: given.

$$p(\mathbf{x} \mid \boldsymbol{\mu}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Lambda^{-1}).$$

$\boldsymbol{\mu}$の共役事前分布はGauss分布:

$$p(\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu} \mid \mathbf{m}, \Lambda_{\mu}^{-1}).$$ $\mathbf{m}, \Lambda_{\mu}^{-1}$: ハイパーパラメータ

- データ $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ を観測. 事後分布は,

$$\begin{aligned}
p(\boldsymbol{\mu} \mid \mathcal{X}) &\propto p(\mathcal{X} \mid \boldsymbol{\mu}) p(\boldsymbol{\mu}) \\
&= \left(\prod_{n=1}^{N} p(\mathbf{x}_n \mid \boldsymbol{\mu}) \right) p(\boldsymbol{\mu}) \\
&= \left(\prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}, \Lambda^{-1}) \right) \mathcal{N}(\boldsymbol{\mu} \mid \mathbf{m}, \Lambda_{\mu}^{-1})
\end{aligned}$$

対数をとる

$$\begin{aligned}
&\log p(\boldsymbol{\mu} \mid \mathcal{X}) \\
&= \sum_{n=1}^{N} \log \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}, \Lambda^{-1}) + \log \mathcal{N}(\boldsymbol{\mu} \mid \mathbf{m}, \Lambda_{\mu}^{-1}) + \text{const.} \\
&= \sum_{n=1}^{N} \left(-\frac{1}{2} \left((\mathbf{x}_n - \boldsymbol{\mu})^T \Lambda (\mathbf{x}_n - \boldsymbol{\mu}) - \underbrace{\log(\det \Lambda)}_{\text{const.}} \right) \right) \\
&\quad - \frac{1}{2} \left((\boldsymbol{\mu} - \mathbf{m})^T \Lambda_{\mu} (\boldsymbol{\mu} - \mathbf{m}) - \underbrace{\log(\det \Lambda_{\mu})}_{\text{const.}} \right) + \text{const.}
\end{aligned}$$

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

$$= \log \mathcal{N}(x_* \mid \mu, \Lambda^{-1}) - \log \mathcal{N}(\mu \mid \mathbf{m}(x_*), (\Lambda + \Lambda_\mu)^{-1}) + \text{const.}$$

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

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

$$+ \text{const.}$$

$$\left((\Lambda+\Lambda_\mu)^{-1}\right)^T = \left((\Lambda+\Lambda_\mu)^T\right)^{-1} = (\Lambda+\Lambda_\mu)^{-1}$$

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

$$+ \frac{1}{2}\left(-2\mu^T(\cancel{\Lambda x_*} + \underbrace{\Lambda_\mu \mathbf{m}) + \mathbf{m}(x_*)^T(\Lambda x_* + \Lambda_\mu \mathbf{m})}_{\text{const.}}\right) + \text{const.}$$

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

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

↑ 多次元Gauss分布のpdfの対数をとったものと形が同じ.

ここで,

$p(x_*) := \mathcal{N}(x_* \mid \mu_*, \Lambda_*^{-1})$ とすると,

$$\log p(x_*) = -\frac{1}{2}\left(x_*^T \Lambda_* x_* - 2x_*^T \Lambda_* \mu_*\right) + \text{const.}$$

$$\therefore \quad \Lambda_* = \Lambda - \Lambda(\Lambda + \Lambda_\mu)^{-1}\Lambda = (\Lambda^{-1} + \Lambda_\mu^{-1})^{-1}.$$

↑ Woodburyの公式 $\left(\begin{array}{l} A = \Lambda^{-1},\ B = \Lambda_\mu^{-1} \\ U = V = I_D \end{array}\right)$

$$\mu_* = \Lambda_*^{-1}\Lambda(\Lambda + \Lambda_\mu)^{-1}\Lambda_\mu \mathbf{m}$$

↓ Woodbury

$$= \Lambda_*^{-1}\Lambda(\Lambda^{-1} - \Lambda^{-1}(\Lambda_\mu^{-1} + \Lambda^{-1})^{-1}\Lambda^{-1})\Lambda_\mu \mathbf{m}$$

$$= (\Lambda_*^{-1} - \Lambda_*^{-1}\Lambda_* \Lambda^{-1})\Lambda_\mu \mathbf{m} = \Lambda_\mu^{-1}\Lambda_\mu \mathbf{m} = \mathbf{m}.$$

$$= -\frac{1}{2}\sum_{n=1}^{N}\left(\underbrace{x_n^T \Lambda x_n}_{\text{const.}} - x_n^T\Lambda\mu - \mu^T\Lambda x_n + \mu^T\Lambda\mu\right)$$

$$-\frac{1}{2}\left(\mu^T\Lambda_\mu\mu - \mu^T\Lambda_\mu m - m^T\Lambda_\mu\mu + \underbrace{m^T\Lambda_\mu m}_{\text{const.}}\right) + \text{const.}$$

Λ, Λ_μ は対称.
$x^T\Lambda y = (x^T\Lambda y)^T = y^T\Lambda^T x.$

$$= -\frac{1}{2}\sum_{n=1}^{N}\left(\mu^T\Lambda\mu - 2\mu^T\Lambda x_n\right) - \frac{1}{2}\left(\mu^T\Lambda_\mu\mu - 2\mu^T\Lambda_\mu m\right) + \text{const.}$$

$$= -\frac{1}{2}\left(\mu^T(N\Lambda + \Lambda_\mu)\mu - 2\mu^T\left(\Lambda\left(\sum_{n=1}^{N}x_n\right) + \Lambda_\mu m\right)\right) + \text{const.} \quad -①$$

↑ 多次元Gauss分布のpdfの対数をとったものと形が同じ.

ここで,

$p(\mu|X) = \mathcal{N}(\mu|\hat{m}, \hat{\Lambda}_\mu^{-1})$ とすると,

$$\log p(\mu|X) = -\frac{1}{2}\left(\mu^T\hat{\Lambda}_\mu\mu - 2\mu^T\hat{\Lambda}_\mu\hat{m}\right) + \text{const.} \quad -②$$

①②を比較すれば,

1次元のときの結果と対応がとれる.

$$\hat{\Lambda}_\mu := N\Lambda + \Lambda_\mu, \quad \hat{m} := \hat{\Lambda}_\mu^{-1}\left(\Lambda\left(\sum_{n=1}^{N}x_n\right) + \Lambda_\mu m\right).$$

- 予測分布 $p(x_*)$.

Bayes' Thm. から,

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

- $p(\mu|x_*)$ は上の結果から

$$p(\mu|x_*) = \mathcal{N}(\mu|m(x_*), (\Lambda + \Lambda_\mu)^{-1}),$$

$$m(x_*) := (\Lambda + \Lambda_\mu)^{-1}(\Lambda x_* + \Lambda_\mu m).$$

- $p(x_*|\mu)$ はモデルの式そのもの: $p(x_*|\mu) = \mathcal{N}(x_*|\mu, \Lambda^{-1}).$

cf. Sherman-Morrison-Woodbury の公式の証明.

Th. (Sherman-Morrison-Woodbury)

$A \in M_n(\mathbb{R})$: regular, $D \in M_m(\mathbb{R})$: regular.

$B \in M_{n,m}(\mathbb{R})$, $C \in M_{m,n}(\mathbb{R})$.

$D^{-1} + CA^{-1}B$: regular ← $D+DCA^{-1}BD$ の正則性はこの条件により保証される.

$$\Rightarrow (A+BDC)^{-1}$$

$$= A^{-1} - A^{-1}B(D^{-1}+CA^{-1}B)^{-1}CA^{-1}$$

D^{-1}が現れない式.

$$= A^{-1} - A^{-1}BD(D+DCA^{-1}BD)^{-1}DCA^{-1}. \qquad \square$$

pf. 行列の基本変形を考えると,

列の交換

行の交換

$$\begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix}\begin{pmatrix} A & -B \\ C & D^{-1} \end{pmatrix}\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix} = \begin{pmatrix} D^{-1} & C \\ -B & A \end{pmatrix}$$

が成立. 両辺逆行列をとると,

$$\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix}^{-1}\begin{pmatrix} A & -B \\ C & D^{-1} \end{pmatrix}^{-1}\begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix}^{-1} = \begin{pmatrix} D^{-1} & C \\ -B & A \end{pmatrix}^{-1}$$

i.e.

$$\begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix}\begin{pmatrix} A & -B \\ C & D^{-1} \end{pmatrix}^{-1}\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix} = \begin{pmatrix} D^{-1} & C \\ -B & A \end{pmatrix}^{-1}.$$

ところで，逆行列は行列の基本変形をくり返すことで求められる．

$$\left(\begin{array}{cc|cc} A & -B & I_n & O \\ C & D^{-1} & O & I_m \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_n & -A^{-1}B & A^{-1} & O \\ C & D^{-1} & O & I_m \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_n & -A^{-1}B & A^{-1} & O \\ O & D^{-1}+CA^{-1}B & -CA^{-1} & I_m \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_n & -A^{-1}B & A^{-1} & O \\ O & I_m & -(D^{-1}+CA^{-1}B)^{-1}CA^{-1} & * \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_n & O & A^{-1}+A^{-1}B(D^{-1}+CA^{-1}B)^{-1}CA^{-1} & * \\ O & I_m & * & * \end{array}\right)$$

$$\therefore \begin{pmatrix} A & -B \\ C & D^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1}+A^{-1}B(D^{-1}+CA^{-1}B)^{-1}CA^{-1} & * \\ * & * \end{pmatrix}.$$

$$\left(\begin{array}{cc|cc} D^{-1} & C & I_m & O \\ -B & A & O & I_n \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} I_m & DC & D & O \\ -B & A & O & I_n \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_m & DC & D & O \\ O & A+BDC & * & I_n \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_m & DC & D & O \\ O & I_n & * & (A+BDC)^{-1} \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|cc} I_m & O & * & * \\ O & I_n & * & (A+BDC)^{-1} \end{array}\right)$$

$$\therefore \begin{pmatrix} D^{-1} & C \\ -B & A \end{pmatrix}^{-1} = \begin{pmatrix} * & * \\ * & (A+BDC)^{-1} \end{pmatrix}.$$

以上を

$$\begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix}\begin{pmatrix} A & -B \\ C & D^{-1} \end{pmatrix}^{-1}\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix} = \begin{pmatrix} D^{-1} & C \\ -B & A \end{pmatrix}^{-1}$$

に代入すると,

$$(A+BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1}+CA^{-1}B)^{-1}CA^{-1}.$$

また,

$$\begin{aligned}(D^{-1}+CA^{-1}B)^{-1} &= DD^{-1}(D^{-1}+CA^{-1}B)^{-1}D^{-1}D \\ &= D(D(D^{-1}+CA^{-1}B)D)^{-1}D \\ &= D(D+DCA^{-1}BD)^{-1}D.\end{aligned}$$

なので, 後半も成立する. ▨

<u>Cor.</u> (Sherman-Morrison)

$A \in M_n(\mathbb{R})$: regular. $\mathbb{c}^T A^{-1} \mathbb{b} \neq -1$

$$\Rightarrow (A+\mathbb{b}\mathbb{c}^T)^{-1} = A^{-1} - A^{-1}\mathbb{b}\mathbb{c}^T A^{-1}(1+\mathbb{c}^T A^{-1}\mathbb{b})^{-1}$$ □

pf. SMW公式で $B = \mathbb{b}$, $C = \mathbb{c}^T$, $D = 1$ とすればよい. ▨

3.4.2 精度が未知の場合

$\mu \in \mathbb{R}^D$: given. $\Lambda \in M_D(\mathbb{R})$, $\Lambda > O$: unknown $\to$ Λを推定.

観測モデル: $p(x|\Lambda) = \mathcal{N}(x|\mu, \Lambda^{-1})$.

$\Lambda > O$ を生成する分布 $\to$ Wishart分布. これを事前分布とする.

$$p(\Lambda) = \mathcal{W}(\Lambda|\nu, W). \quad \nu > D-1, \ W \in M_D(\mathbb{R}), W > O.$$

$\mathcal{X} = \{x_1, \cdots, x_N\}$: 観測データ (given)

- 事後分布の計算.

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

Bayes' Thm.

$$= \log p(\mathcal{X}|\Lambda) + \log p(\Lambda) + \text{const.}$$

$$= \sum_{n=1}^{N} \log p(x_n|\Lambda) + \log p(\Lambda) + \text{const.}$$

$$= \sum_{n=1}^{N} \log \mathcal{N}(x_n|\mu, \Lambda^{-1}) + \log \mathcal{W}(\Lambda|\nu, W) + \text{const.}$$

$$= \sum_{n=1}^{N}\left(-\frac{1}{2}\left((x_n-\mu)^T\Lambda(x_n-\mu) - \log(\det\Lambda)\right)\right) + \frac{\nu-D-1}{2}\log(\det\Lambda) - \frac{1}{2}\mathrm{tr}(W^{-1}\Lambda) + \text{const.}$$

$$= \frac{N+\nu-D-1}{2}\log(\det\Lambda) - \frac{1}{2}\left(\sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu) + \mathrm{tr}(W^{-1}\Lambda)\right) + \text{const.}$$

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

$\mathrm{tr}(x^TAy)=\mathrm{tr}(yx^TA)$

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

$\mathrm{tr}(A+B)=\mathrm{tr}(A)+\mathrm{tr}(B)$

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

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

Wishart分布のpdfの対数をとった形

$$\therefore\ p(\Lambda|\mathcal{X})=\mathcal{W}(\Lambda|\hat{\nu},\hat{W}),$$
$$\hat{\nu}:=N+\nu,\quad \hat{W}^{-1}:=\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T+W^{-1}.$$

- 予測分布の計算.

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

- $p(\Lambda|x_*)$は上の式を用いると,

$$p(\Lambda|x_*)=\mathcal{W}(\Lambda|1+\nu,W(x_*)),$$
$$W(x_*)^{-1}:=(x_*-\mu)(x_*-\mu)^T+W^{-1}.$$

・ $p(x_*|\Lambda)$ はモデルの式より, $p(x_*|\Lambda) = \mathcal{N}(x_*|\mu, \Lambda^{-1})$.

∴ $\log p(x_*)$

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

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

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

← Wishart分布の正規化項に含まれる項. x_* に依存するので考慮する必要あり.

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

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

$\mathrm{tr}(x^TAy) = \mathrm{tr}(yx^TA)$

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

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

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

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

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

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

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

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

$$= -\frac{1+\nu}{2}\log\left(\det(I_D + (x_*-\mu)(x_*-\mu)^TW)\right)$$

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

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

ここで，次の公式を示しておく．

<u>Prop.</u> $A, B \in M_{m,n}(\mathbb{R})$. $\det(I_m + AB^T) = \det(I_n + A^T B)$. $\square$

pf. $$\det\begin{pmatrix} I_m & -A \\ B^T & I_n \end{pmatrix} = \det\begin{pmatrix} I_m & -A \\ B^T & I_n \end{pmatrix}^T = \det\begin{pmatrix} I_m & B \\ -A^T & I_n \end{pmatrix}.$$

行列の基本変形を行い，detを計算する．

列基本変形

$$\det\begin{pmatrix} I_m & -A \\ B^T & I_n \end{pmatrix} = \det\begin{pmatrix} I_m + AB^T & -A \\ O & I_n \end{pmatrix}$$

$$= \det(I_m + AB^T)\det I_n$$

$$= \det(I_m + AB^T).$$

行基本変形

$$\det\begin{pmatrix} I_m & B \\ -A^T & I_n \end{pmatrix} = \det\begin{pmatrix} I_m & B \\ O & I_n + A^T B \end{pmatrix}$$

$$= \det I_m \det(I_n + A^T B)$$

$$= \det(I_n + A^T B).$$

$$\therefore \det(I_m + AB^T) = \det(I_n + A^T B). \quad \blacksquare$$

これを用いると，

1×1の行列のdetは行列の要素そのもの

$$\log p(x_*)$$

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

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

$\downarrow$ W : sym.

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

多次元t分布のpdf.の対数をとったもの

Def. (Studentのt分布, 多次元版)

$$\mu_s \in \mathbb{R}^D,\ \Lambda_s \in M_D(\mathbb{R}),\ \Lambda_s \succ 0,\ \nu_s > 0.$$

$$\mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)$$

$$= \frac{\Gamma\left(\frac{\nu_s+D}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)} \frac{(\det \Lambda_s)^{\frac{1}{2}}}{(\pi\nu_s)^{\frac{D}{2}}} \left(1+\frac{1}{\nu_s}(x-\mu_s)^T \Lambda_s (x-\mu_s)\right)^{-\frac{\nu_s+D}{2}}. \qquad \square$$

$$\log \mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)$$

$$= \frac{\nu_s+D}{2}\log\left(1+\frac{1}{\nu_s}(x-\mu_s)^T \Lambda_s (x-\mu_s)\right) + \text{Const.}$$

以上より,

$\nu > D-1$ だった!

$$\mu_s := \mu,\quad \Lambda_s := (1-D+\nu)W,\quad \nu_s := 1-D+\nu\ (>0).$$

cf. 多次元Studentのt分布の正規化項, 平均, 分散の計算.

- 正規化項が

$$C(\mu_s, \Lambda_s, \nu_s) := \frac{\Gamma\left(\frac{\nu_s+D}{2}\right)}{\Gamma\left(\frac{\nu_s}{2}\right)} \frac{(\det \Lambda_s)^{\frac{1}{2}}}{(\pi\nu_s)^{\frac{D}{2}}}$$

になるのを確認する.

$$I = \int_{\mathbb{R}^D} \left(1 + \frac{1}{\nu_s}(x - \mu_s)^T \Lambda_s (x - \mu_s)\right)^{-\frac{\nu_s + D}{2}} dx = C(\mu_s, \Lambda_s, \nu_s)^{-1}$$

を示せばよい.

$\Lambda_s > 0$ より $\Lambda_s = LL^T$ と Cholesky分解する.

$y := \frac{1}{\sqrt{\nu_s}} L^T (x - \mu_s)$ とおく. このとき, $x = \sqrt{\nu_s}(L^T)^{-1} y + \mu_s$

$\frac{\partial x}{\partial y} = \sqrt{\nu_s} L^{-1}$. $\left(\because \frac{\partial}{\partial x} Ax = A^T\right)$

Cholesky分解の性質から.

また, $\det \Lambda_s = (\det L)^2$ で, $\det L > 0$ より

$\det L = (\det \Lambda_s)^{\frac{1}{2}}$.

この変換のJacobianは,

$$J(y) = \det\left(\frac{\partial x}{\partial y}\right) = \det\left(\sqrt{\nu_s} L^{-1}\right) = \nu_s^{\frac{D}{2}}\left(\det L^{-1}\right)$$

$$= \nu_s^{\frac{D}{2}}(\det L)^{-1} = \nu_s^{\frac{D}{2}}(\det \Lambda_s)^{-\frac{1}{2}} \ (>0).$$

よって,

$$I = \int_{\mathbb{R}^D} \left(1 + y^T y\right)^{-\frac{\nu_s + D}{2}} |J(y)| dy$$

$$= \nu_s^{\frac{D}{2}} (\det \Lambda_s)^{-\frac{1}{2}} \int_{\mathbb{R}^D} \left(1 + y^T y\right)^{-\frac{\nu_s + D}{2}} dy.$$

ここで, 次の主張を示す:

[Claim] $\displaystyle\int_{\mathbb{R}^D}\left(1+y^Ty\right)^{-\frac{\nu_s+D}{2}}dy=\frac{\pi^{\frac{D}{2}}\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s+D}{2}\right)}.$

pf. 次元 D についての数学的帰納法で示す.

- $D=1$ のときは、1次元 t 分布のときに計算している.
- $D-1$ 次元で成立するとする:

$$\int_{\mathbb{R}^{D-1}}\left(1+y^Ty\right)^{-\frac{\nu_s+D-1}{2}}dy=\frac{\pi^{\frac{D-1}{2}}\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s+D-1}{2}\right)}.$$

$y=(y_1,\cdots,y_D)^T$ とし、$y_{D-1}:=(y_1,\cdots,y_{D-1})^T\in\mathbb{R}^{D-1}$ と定める.

$$\int_{\mathbb{R}^D}\left(1+y^Ty\right)^{-\frac{\nu_s+D}{2}}dy$$

$$=\int_{\mathbb{R}^D}\left(1+y_{D-1}^Ty_{D-1}+y_D^2\right)^{-\frac{\nu_s+D}{2}}dy_{D-1}dy_D$$

$$=\int_{\mathbb{R}^{D-1}}\left(\int_{-\infty}^{\infty}\left(1+y_{D-1}^Ty_{D-1}+y_D^2\right)^{-\frac{\nu_s+D}{2}}dy_D\right)dy_{D-1}.$$

ここで、$1+y_{D-1}^Ty_{D-1}=:Y_{D-1}$ とおいて、

$$\int_{-\infty}^{\infty}\left(Y_{D-1}+y_D^2\right)^{-\frac{\nu_s+D}{2}}dy_D$$

偶関数

$$=2\int_0^{\infty}\left(Y_{D-1}+y_D^2\right)^{-\frac{\nu_s+D}{2}}dy_D.$$

$Y_{D-1}\left(Y_{D-1}+y_D^2\right)^{-1}=:t\in(0,1]$ とおくと、

$$y_D^2 = Y_{D-1}\frac{1-t}{t} \qquad y_D > 0 \text{ より}$$

$$y_D = \sqrt{Y_{D-1}\frac{1-t}{t}}\ .$$

$$dy_D = \frac{1}{2}\left(Y_{D-1}\frac{1-t}{t}\right)^{-\frac{1}{2}}\cdot\left(-Y_{D-1}\frac{1}{t^2}\right)dt.$$

$$= -\frac{1}{2}Y_{D-1}^{\frac{1}{2}}t^{-\frac{3}{2}}(1-t)^{-\frac{1}{2}}dt$$

$$\therefore \int_{-\infty}^{\infty}\left(Y_{D-1}+y_D^2\right)^{-\frac{\nu_s+D}{2}}dy_D$$

$$= \int_0^1\left(\frac{t}{Y_{D-1}}\right)^{\frac{\nu_s+D}{2}}Y_{D-1}^{\frac{1}{2}}t^{-\frac{3}{2}}(1-t)^{-\frac{1}{2}}dt$$

$$= Y_{D-1}^{-\frac{\nu_s+D-1}{2}}\int_0^1 t^{\frac{\nu_s+D-1}{2}-1}(1-t)^{\frac{1}{2}-1}dt$$

$$\downarrow \text{ ベータ関数の定義}$$

$$= Y_{D-1}^{-\frac{\nu_s+D-1}{2}}B\left(\frac{\nu_s+D-1}{2},\frac{1}{2}\right).$$

よって,

$$\int_{\mathbb{R}^D}\left(1+y^Ty\right)^{-\frac{\nu_s+D}{2}}dy$$

$$= \int_{\mathbb{R}^{D-1}}Y_{D-1}^{-\frac{\nu_s+D-1}{2}}B\left(\frac{\nu_s+D-1}{2},\frac{1}{2}\right)dy_{D-1}$$

$$= B\left(\frac{\nu_s+D-1}{2},\frac{1}{2}\right)\int_{\mathbb{R}^{D-1}}\left(1+y_{D-1}^Ty_{D-1}\right)^{-\frac{\nu_s+D-1}{2}}dy_{D-1}$$

$$\downarrow \text{ 帰納法の仮定}$$

$$= B\left(\frac{\nu_s+D-1}{2},\frac{1}{2}\right)\frac{\pi^{\frac{D-1}{2}}\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s+D-1}{2}\right)}$$

$$\downarrow \text{ ベータ関数の性質 } B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$= \frac{\cancel{\Gamma\left(\frac{\nu_s+D-1}{2}\right)}\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_s+D}{2}\right)}\frac{\pi^{\frac{D-1}{2}}\Gamma\left(\frac{\nu_s}{2}\right)}{\cancel{\Gamma\left(\frac{\nu_s+D-1}{2}\right)}}$$

$$\downarrow\ \Gamma\left(\frac{1}{2}\right)=\pi^{\frac{1}{2}}$$

$$= \frac{\pi^{\frac{1}{2}} \pi^{\frac{D-1}{2}} \Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s + D}{2}\right)}$$

$$= \frac{\pi^{\frac{D}{2}} \Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s + D}{2}\right)} . \qquad ▨$$

以上より，

$$I = \nu_s^{\frac{D}{2}} (\det \Lambda_s)^{-\frac{1}{2}} \frac{\pi^{\frac{D}{2}} \Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s + D}{2}\right)}$$

$$= C(\mu_s, \Lambda_s, \nu_s)^{-1}$$

となって，確かに $\int_{\mathbb{R}^D} \mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)\, dx = 1$ となることが分かる．

- 平均．

$$\mathbb{E}[X]$$

$$= \int_{\mathbb{R}^D} x\, \mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)\, dx$$

$$= \mu_s \underbrace{\int_{\mathbb{R}^D} \mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)\, dx}_{=1.} + \int_{\mathbb{R}^D} (x - \mu_s) \mathrm{St}(x \mid \mu_s, \Lambda_s, \nu_s)\, dx$$

$$= \mu_s$$

$$+ C(\mu_s, \Lambda_s, \nu_s) \int_{\mathbb{R}^D} (x - \mu_s) \left(1 + \frac{1}{\nu_s} (x - \mu_s)^T \Lambda_s (x - \mu_s)\right)^{-\frac{\nu_s + D}{2}} dx.$$

以下，$I_1 := \int_{\mathbb{R}^D} (x - \mu_s) \left(1 + \frac{1}{\nu_s} (x - \mu_s)^T \Lambda_s (x - \mu_s)\right)^{-\frac{\nu_s + D}{2}} dx$

を計算する．

$\Lambda_s = LL^T$ と Cholesky 分解する. $\det L = (\det \Lambda_s)^{\frac{1}{2}}$.

$y := \frac{1}{\sqrt{\nu_s}} L^T (x - \mu_s)$ とおく. このとき, $x = \sqrt{\nu_s} (L^T)^{-1} y + \mu_s$ で,

$\frac{\partial x}{\partial y} = \sqrt{\nu_s} L^{-1}$.

Jacobian は, $J(y) = \nu_s^{\frac{D}{2}} (\det \Lambda_s)^{-\frac{1}{2}}$.

$$
\begin{aligned}
I_1 &= \sqrt{\nu_s} (L^T)^{-1} \int_{\mathbb{R}^D} y (1 + y^T y)^{-\frac{\nu_s + D}{2}} \nu_s^{\frac{D}{2}} (\det \Lambda_s)^{-\frac{1}{2}} dy \\
&= \nu_s^{\frac{D+1}{2}} (\det \Lambda_s)^{-\frac{1}{2}} (L^T)^{-1} \int_{\mathbb{R}^D} y (1 + y^T y)^{-\frac{\nu_s + D}{2}} dy.
\end{aligned}
$$

$I_2 := \int_{\mathbb{R}^D} y (1 + y^T y)^{-\frac{\nu_s + D}{2}} dy$ を計算する.

第 i 成分に注目すると,

$$
\begin{aligned}
(I_2)_i &= \int_{\mathbb{R}^D} y_i (1 + y^T y)^{-\frac{\nu_s + D}{2}} dy \\
&= \int_{\mathbb{R}^D} y_i \left(1 + y_i^2 + \sum_{j \neq i} y_j^2\right)^{-\frac{\nu_s + D}{2}} dy.
\end{aligned}
$$

添字のつけかえをしているだけ

ここで, $\mathbb{z}_D := (z_1, \cdots, z_D) := (y_i, y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_D)^T$

$$\mathbb{z}_k := (z_1, \cdots, z_k) \in \mathbb{R}^k \quad (k = 1, \cdots, D-1)$$

とすれば,

$$(I_2)_i = \int_{\mathbb{R}^{D-1}} z_1 \left(\int_{-\infty}^{\infty} (1 + \mathbb{z}_D^T \mathbb{z}_D)^{-\frac{\nu_s + D}{2}} dz_D \right) d\mathbb{z}_{D-1}$$

ここで，正規化定数の計算で行ったとおり，

$$\int_{-\infty}^{\infty}\left(1+\mathbb{z}_D^T\mathbb{z}_D\right)^{-\frac{\nu_s+D}{2}}d\mathbb{z}_D = B\left(\frac{\nu_s+D-1}{2},\frac{1}{2}\right)\left(1+\mathbb{z}_{D-1}^T\mathbb{z}_{D-1}\right)^{-\frac{\nu_s+D-1}{2}}$$

であり，これを繰り返し用いることで，

$$\begin{aligned}(I_2)_i &= \left(\prod_{j=1}^{D-1}B\left(\frac{\nu_s+D-j}{2},\frac{1}{2}\right)\right)\int_{-\infty}^{\infty}z_1\left(1+z_1^2\right)^{-\frac{\nu_s+1}{2}}dz_1 \\ &= \left(\prod_{j=1}^{D-1}B\left(\frac{\nu_s+D-j}{2},\frac{1}{2}\right)\right)\int_{-\infty}^{\infty}y_i\left(1+y_i^2\right)^{-\frac{\nu_s+1}{2}}dy_i .\end{aligned}$$

さて，1次元t分布の平均の計算でみたとおり，

$$\int_{-\infty}^{\infty}y_i\left(1+y_i^2\right)^{-\frac{\nu_s+1}{2}}dy_i$$

$$= \begin{cases} 0 & (\nu_s > 1) \\ \text{indeterminate} & (0 < \nu_s \leq 1) \end{cases}$$

であったから，結局

$$I_1 = \begin{cases} 0 & (\nu_s > 1) \\ \text{indeterminate} & (0 < \nu_s \leq 1) . \end{cases}$$

$$\therefore\ \mathbb{E}[X] = \begin{cases} \mu_s & (\nu_s > 1) \\ \text{undefined} & (0 < \nu_s \leq 1) . \end{cases}$$

・分散の計算.

平均は $\nu_s > 1$ で存在するので, この範囲で考える.

$$\mathbb{V}[X]$$

$$= \int_{\mathbb{R}^D} (\boldsymbol{x}-\boldsymbol{\mu}_s)(\boldsymbol{x}-\boldsymbol{\mu}_s)^T \mathrm{St}(\boldsymbol{x}|\boldsymbol{\mu}_s, \Lambda_s, \nu_s)\, d\boldsymbol{x}$$

$$= C(\boldsymbol{\mu}_s, \Lambda_s, \nu_s)$$

$$\times \int_{\mathbb{R}^D} (\boldsymbol{x}-\boldsymbol{\mu}_s)(\boldsymbol{x}-\boldsymbol{\mu}_s)^T \left(1 + \frac{1}{\nu_s}(\boldsymbol{x}-\boldsymbol{\mu}_s)^T \Lambda_s (\boldsymbol{x}-\boldsymbol{\mu}_s)\right)^{-\frac{\nu_s+D}{2}} d\boldsymbol{x}.$$

積分

$$\mathcal{I}_1 := \int_{\mathbb{R}^D} (\boldsymbol{x}-\boldsymbol{\mu}_s)(\boldsymbol{x}-\boldsymbol{\mu}_s)^T \left(1 + \frac{1}{\nu_s}(\boldsymbol{x}-\boldsymbol{\mu}_s)^T \Lambda_s (\boldsymbol{x}-\boldsymbol{\mu}_s)\right)^{-\frac{\nu_s+D}{2}} d\boldsymbol{x}.$$

について考える.

$\Lambda_s = LL^T$ と Cholesky 分解する. $\det L = (\det \Lambda_s)^{\frac{1}{2}}$.

$\boldsymbol{y} := \frac{1}{\sqrt{\nu_s}} L^T(\boldsymbol{x}-\boldsymbol{\mu}_s)$ とおく. このとき, $\boldsymbol{x} = \sqrt{\nu_s}(L^T)^{-1}\boldsymbol{y} + \boldsymbol{\mu}_s$ で,

$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \sqrt{\nu_s} L^{-1}$.

Jacobian は, $J(\boldsymbol{y}) = \nu_s^{\frac{D}{2}} (\det \Lambda_s)^{-\frac{1}{2}}$.

$$(\boldsymbol{x}-\boldsymbol{\mu}_s)(\boldsymbol{x}-\boldsymbol{\mu}_s)^T = \sqrt{\nu_s}(L^T)^{-1}\boldsymbol{y}\left(\sqrt{\nu_s}(L^T)^{-1}\boldsymbol{y}\right)^T = \nu_s (L^T)^{-1}\boldsymbol{y}\boldsymbol{y}^T L^{-1}$$

であるから,

$$\mathcal{I}_1 = \nu_s^{\frac{D}{2}+1}(\det \Lambda_s)^{-\frac{1}{2}}(L^T)^{-1}\left(\int_{\mathbb{R}^D} yy^T(1+y^Ty)^{-\frac{\nu_s+D}{2}}dy\right)L^{-1}.$$

以下，積分

$$\mathcal{I}_2 = \int_{\mathbb{R}^D} yy^T(1+y^Ty)^{-\frac{\nu_s+D}{2}}dy$$

の (i,j)-成分を考える．yy^T は対称なので，$i \leq j$ としてよい．

$$(\mathcal{I}_2)_{ij} = \int_{\mathbb{R}^D} y_i y_j (1+y^Ty)^{-\frac{\nu_s+D}{2}}dy.$$

ここで，$\mathbb{Z}_D := (z_1, \cdots, z_D)$

$$:= (y_i, y_j, y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_{j-1}, y_{j+1}, \cdots, y_D)^T$$

$$\mathbb{Z}_k := (z_1, \cdots, z_k) \in \mathbb{R}^k \quad (k = 1, \cdots, D-1)$$

とすれば，

$$(\mathcal{I}_2)_{ij} = \int_{\mathbb{R}^D} z_1 z_2 (1+\mathbb{Z}_D^T\mathbb{Z}_D)^{-\frac{\nu_s+D}{2}}d\mathbb{Z}_D.$$

□ $i<j$ のとき．z_1, z_2 はそれぞれ独立．

平均のときの計算と同様にして，

$$(\mathcal{I}_2)_{ij} = \left(\prod_{k=1}^{D-2} B\left(\frac{\nu_s+D-k}{2}, \frac{1}{2}\right)\right)\int_{\mathbb{R}^2} z_1 z_2 (1+\mathbb{Z}_2^T\mathbb{Z}_2)^{-\frac{\nu_s+2}{2}}d\mathbb{Z}_2.$$

$$\int_{\mathbb{R}^2} z_1 z_2 (1+\mathbb{Z}_2^T\mathbb{Z}_2)^{-\frac{\nu_s+2}{2}}d\mathbb{Z}_2$$

$$= \int_{-\infty}^{\infty} z_1 \left(\int_{-\infty}^{\infty} z_2 (1+z_1^2+z_2^2)^{-\frac{\nu_s+2}{2}}dz_2\right)dz_1.$$

$$\int_{-\infty}^{\infty} z_2\left(1+z_1^2+z_2^2\right)^{-\frac{\nu_s+2}{2}} dz_2$$

$$= \int_{0}^{\infty} z_2\left(1+z_1^2+z_2^2\right)^{-\frac{\nu_s+2}{2}} dz_2 + \int_{-\infty}^{0} z_2\left(1+z_1^2+z_2^2\right)^{-\frac{\nu_s+2}{2}} dz_2 .$$

$1+z_1^2 =: Z_1$ とし，$t = Z_1\left(Z_1+z_2^2\right)^{-1} \in (0,1]$ と変数変換する．

$$z_2^2 = Z_1 \frac{1-t}{t} .$$

$$z_2 = \begin{cases} \sqrt{Z_1 \frac{1-t}{t}} & (z_2 \in [0,\infty)) \\ -\sqrt{Z_1 \frac{1-t}{t}} & (z_2 \in (-\infty,0)) \end{cases}$$

$$\therefore dz_2 = \begin{cases} -\frac{1}{2} Z_1^{\frac{1}{2}} t^{-\frac{3}{2}} (1-t)^{-\frac{1}{2}} dt & (z_2 \in [0,\infty)) \\ \frac{1}{2} Z_1^{\frac{1}{2}} t^{-\frac{3}{2}} (1-t)^{-\frac{1}{2}} dt & (z_2 \in (-\infty,0)) \end{cases}$$

ゆえに，

$$\int_{-\infty}^{\infty} z_2\left(1+z_1^2+z_2^2\right)^{-\frac{\nu_s+2}{2}} dz_2$$

$$= \frac{1}{2} Z_1^{-\frac{\nu_s}{2}} \int_0^1 t^{\frac{\nu_s-2}{2}} dt + \frac{1}{2} Z_1^{-\frac{\nu_s}{2}} \int_0^1 t^{\frac{\nu_s-2}{2}} dt$$

$$= Z_1^{-\frac{\nu_s}{2}} \int_0^1 t^{\frac{\nu_s}{2}-1} dt$$

$\nu_s > 1$ なので $\frac{\nu_s}{2}-1 \neq -1$

$$= Z_1^{-\frac{\nu_s}{2}} \left[\frac{2}{\nu_s} t^{\frac{\nu_s}{2}}\right]_0^1$$

$$= \frac{2}{\nu_s} Z_1^{-\frac{\nu_s}{2}} .$$

これより，

$$\int_{\mathbb{R}^2} z_1 z_2 \left(1+\mathbb{Z}_2^T\mathbb{Z}_2\right)^{-\frac{\nu_s+2}{2}} d\mathbb{Z}_2$$

$$= \frac{2}{\nu_s}\int_{-\infty}^{\infty} z_1\left(1+z_1^2\right)^{-\frac{\nu_s}{2}} dz_1$$

$1+z_1^2 =: u$ とおく.

$du = 2z_1$

$$= \frac{1}{\nu_s}\left(\int_1^{\infty} u^{-\frac{\nu_s}{2}} du + \int_{\infty}^{1} u^{-\frac{\nu_s}{2}} du\right).$$

・$\nu_s \neq 2$ のとき.

$$\int_1^{\infty} u^{-\frac{\nu_s}{2}} du$$

$$= \lim_{a\to\infty}\left[-\frac{2}{\nu_s-2}u^{-\frac{\nu_s-2}{2}}\right]_1^a$$

$$= \begin{cases} \dfrac{2}{\nu_s-2} & (\nu_s > 2) \\ -\infty & (1 < \nu_s < 2) \end{cases}$$

・$\nu_s = 2$ のとき.

$$\int_1^{\infty} u^{-\frac{\nu_s}{2}} du = \lim_{a\to\infty}\left[\log u\right]_1^a = \infty.$$

$$\therefore \int_1^{\infty} u^{-\frac{\nu_s}{2}} du = \begin{cases} \dfrac{2}{\nu_s-2} & (\nu_s > 2) \\ \infty & (\nu_s = 2) \\ -\infty & (1 < \nu_s < 2) \end{cases}$$

$$\therefore \int_{\mathbb{R}^2} z_1 z_2 \left(1+\mathbb{Z}_2^T\mathbb{Z}_2\right)^{-\frac{\nu_s+2}{2}} d\mathbb{Z}_2 = \begin{cases} 0 & (\nu_s > 2) \\ \text{indeterminate} & (1 < \nu_s \leq 2) \end{cases}$$

よって，$i<j$ のときは，

$$(\mathcal{I}_2)_{ij} = \begin{cases} 0 & (\nu_s > 2) \\ \text{indeterminate} & (1 < \nu_s \leq 2) \end{cases}$$

$\nu_s > 2$ でないと $\mathbb{V}[X]$ が定義されないので，以下では $\nu_s > 2$ とする．

□ $i=j$ のとき．常に $z_1 = z_2$．

平均のときの計算と同様にして，

$$(\mathcal{I}_2)_{ij} = \left(\prod_{k=1}^{D-1} B\left(\frac{\nu_s+D-k}{2}, \frac{1}{2}\right)\right)\int_{-\infty}^{\infty} z_1^2\left(1+z_1^2\right)^{-\frac{\nu_s+1}{2}} dz_1$$

偶関数

$$= 2\left(\prod_{k=1}^{D-1} B\left(\frac{\nu_s+D-k}{2}, \frac{1}{2}\right)\right)\int_{0}^{\infty} z_1^2\left(1+z_1^2\right)^{-\frac{\nu_s+1}{2}} dz_1.$$

ここで，1次元 t 分布のときに行った計算より，

$\nu_s > 2$ では，$\displaystyle\int_0^\infty t^2(1+t^2)^{-\frac{\nu_s+1}{2}} dt = \frac{1}{2} B\left(\frac{\nu_s-2}{2}, \frac{3}{2}\right).$

$$\therefore (\mathcal{I}_2)_{ij} = \left(\prod_{k=1}^{D-1} B\left(\frac{\nu_s+D-k}{2}, \frac{1}{2}\right)\right) B\left(\frac{\nu_s-2}{2}, \frac{3}{2}\right)$$

$$= \left(\prod_{k=1}^{D-1} \frac{\Gamma\left(\frac{\nu_s+D-k}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_s+D-k+1}{2}\right)}\right) \frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\nu_s+1}{2}\right)}$$

$$= \frac{1}{2} \frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\pi^{\frac{D}{2}}}{\Gamma\left(\frac{\nu_s+D}{2}\right)}$$

$$= \frac{1}{\nu_s-2} \frac{\nu_s-2}{2} \frac{\Gamma\left(\frac{\nu_s-2}{2}\right)\pi^{\frac{D}{2}}}{\Gamma\left(\frac{\nu_s+D}{2}\right)} = \frac{1}{\nu_s-2} \frac{\Gamma\left(\frac{\nu_s}{2}\right)\pi^{\frac{D}{2}}}{\Gamma\left(\frac{\nu_s+D}{2}\right)}.$$

以上より,

$$\mathcal{I}_2 = \frac{1}{\nu_s - 2} \frac{\Gamma\left(\frac{\nu_s}{2}\right) \pi^{\frac{D}{2}}}{\Gamma\left(\frac{\nu_s + D}{2}\right)} I_D$$

なので,

$$\mathcal{I}_1 = \frac{\nu_s^{\frac{D}{2}+1}}{\nu_s - 2} (\det \Lambda_s)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu_s}{2}\right) \pi^{\frac{D}{2}}}{\Gamma\left(\frac{\nu_s + D}{2}\right)} (L^T)^{-1} I_D L^{-1}$$

$$= \frac{\nu_s}{\nu_s - 2} \frac{\Gamma\left(\frac{\nu_s}{2}\right)}{\Gamma\left(\frac{\nu_s + D}{2}\right)} \frac{(\pi \nu_s)^{\frac{D}{2}}}{(\det \Lambda_s)^{\frac{1}{2}}} (L L^T)^{-1}$$

$$= \frac{\nu_s}{\nu_s - 2} C(\mu_s, \Lambda_s, \nu_s)^{-1} \Lambda_s^{-1}.$$

ゆえに,

$$\mathbb{V}[X] = C(\mu_s, \Lambda_s, \nu_s) \mathcal{I}_1$$

$$= \frac{\nu_s}{\nu_s - 2} \Lambda_s^{-1}. \qquad (\nu_s > 2).$$