13.32

(13.108)、(13.77) より

$$Q(\theta, \theta^{old}) = -\frac{1}{2}\ln|P_0| - E_{p(z|x,\theta^{old})}\left[\frac{1}{2}(z_1-\mu_0)^T P_0^{-1}(z_1-\mu_0)\right] + C$$

, C は μ_0, P_0 に依らない項

μ_0 についての最尤推定は 2.3.4節と同じく μ_0 の微分が 0 となる値である。

$$0 = \frac{\partial}{\partial \mu_0} Q(\theta, \theta^{old}) = \frac{\partial}{\partial \mu_0}\left\{-E\left[\frac{1}{2}(z_1-\mu_0)^T P_0^{-1}(z_1-\mu_0)\right]\right\}$$

$$= -E\left[\frac{\partial}{\partial \mu_0}\left\{\frac{1}{2}(z_1-\mu_0)^T P_0^{-1}(z_1-\mu_0)\right\}\right]$$

← 積分が存在するなら即ち期待値が存在するなら微分と積分を交換してよい
ここでは期待値が存在する場合のみ考えることとする

$$\therefore\ 0 = E\left[\frac{\partial}{\partial \mu_0}\left\{(z_1-\mu_0)^T P_0^{-1}(z_1-\mu_0)\right\}\right]$$

$$= E\left[\frac{\partial}{\partial \mu_0}\left(z_1^T P_0^{-1} z_1 - z_1^T P_0^{-1}\mu_0 - \mu_0^T P_0^{-1} z_1 + \mu_0^T P_0^{-1}\mu_0\right)\right]$$

$$= E\left[-2P_0^{-1} z_1 + 2P_0^{-1}\mu_0\right]$$

$$\left\{\begin{aligned} &\frac{\partial}{\partial \mu_0}\mu_0^T P_0^{-1} z_1 = P_0^{-1} z_1 \quad \leftarrow \text{(C.19) 分母レイアウト} \\ &\frac{\partial}{\partial \mu_0} z_1^T P_0^{-1}\mu_0 = \overset{\swarrow \text{(C.19)}}{(z_1^T P_0^{-1})^T} = P_0^{-1} z_1 \\ &\frac{\partial}{\partial \mu_0}\mu_0^T P_0^{-1}\mu_0 = \overset{\swarrow \text{分母レイアウト}}{(P_0^{-T} + P_0^{-1})\mu_0} = \overset{\swarrow P_0^{-1}\text{対称なので}}{2P_0^{-1}\mu_0} \end{aligned}\right.$$

$$= -2P_0^{-1}E[z_1] + 2P_0^{-1}\mu_0$$

よって

$$\mu_0 = E[z_1] \quad \cdots (13.110)$$

を得る。

$\mathbf{P}_0$についての最尤推定は

$$0 = \frac{\partial}{\partial \mathbf{P}_0} Q(\theta, \theta^{old}) = \frac{\partial}{\partial \mathbf{P}_0}\left\{ -\frac{1}{2}\ln|\mathbf{P}_0| - \mathrm{E}\left[\frac{1}{2}(\mathbf{z}_1-\boldsymbol{\mu}_0)^T \mathbf{P}_0^{-1}(\mathbf{z}_1-\boldsymbol{\mu}_0)\right]\right\}$$

(C.28)

$$= -\frac{1}{2}(\mathbf{P}_0^{-1})^T - \mathrm{E}\left[\frac{1}{2}\frac{\partial}{\partial \mathbf{P}_0}(\mathbf{z}_1-\boldsymbol{\mu}_0)^T \mathbf{P}_0^{-1}(\mathbf{z}_1-\boldsymbol{\mu}_0)\right]$$

期待値が存在すると仮定する
このとき微分と積分は交換しても同じ結果になる

$$= -\frac{1}{2}(\mathbf{P}_0^{-1})^T - \mathrm{E}\left[\frac{1}{2}\left\{-(\mathbf{P}_0^{-1})^T(\mathbf{z}_1-\boldsymbol{\mu}_0)(\mathbf{z}_1-\boldsymbol{\mu}_0)^T(\mathbf{P}_0^{-1})^T\right\}\right]$$

$$= -\frac{1}{2}(\mathbf{P}_0^{-1})^T - \mathrm{E}\left[\frac{1}{2}\left\{-(\mathbf{P}_0^{-1})^T(\mathbf{z}_1\mathbf{z}_1^T - \mathbf{z}_1\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_0\mathbf{z}_1^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T)(\mathbf{P}_0^{-1})^T\right\}\right]$$

$$= -\frac{1}{2}\mathbf{P}_0^{-1} + \frac{1}{2}\mathbf{P}_0^{-1}\mathrm{E}\left[\mathbf{z}_1\mathbf{z}_1^T - \mathbf{z}_1\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_0\mathbf{z}_1^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T\right]\mathbf{P}_0^{-1}$$

$\mathbf{P}_0^{-1}$対称なので

$$\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial X_{ij}} = \frac{\partial \mathbf{a}^T\mathbf{X}^{-1}}{\partial X_{ij}}\mathbf{b} + \mathbf{a}^T\mathbf{X}^{-1}\underbrace{\frac{\partial \mathbf{b}}{\partial X_{ij}}}_{0}$$

(C.20)

$$= \left(\underbrace{\frac{\partial \mathbf{a}^T}{\partial X_{ij}}}_{0}\mathbf{X}^{-1} + \mathbf{a}^T\frac{\partial \mathbf{X}^{-1}}{\partial X_{ij}}\right)\mathbf{b} = \mathbf{a}^T\frac{\partial \mathbf{X}^{-1}}{\partial X_{ij}}\mathbf{b}$$

(C.20)

$$= \mathbf{a}^T\left(-\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial X_{ij}}\mathbf{X}^{-1}\right)\mathbf{b}$$

(C.21)

シングルエントリ

$$= \mathbf{a}^T\left(-\mathbf{X}^{-1}\mathbf{J}^{ij}\mathbf{X}^{-1}\right)\mathbf{b}$$

$$= -(\mathbf{a}^T\mathbf{X}^{-1})\mathbf{J}^{ij}(\mathbf{X}^{-1}\mathbf{b})$$

$$= -\left\{(\mathbf{a}^T\mathbf{X}^{-1})^T(\mathbf{X}^{-1}\mathbf{b})^T\right\}_{ij}$$

$$= -\left\{(\mathbf{X}^{-1})^T\mathbf{a}\mathbf{b}^T(\mathbf{X}^{-1})^T\right\}_{ij}$$

$$\mathbf{a}^T\mathbf{J}^{12}\mathbf{b} = (a_1, a_2)\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix} = a_1b_2 = (\mathbf{a}\mathbf{b}^T)_{12}$$

よって

$$0 = -\mathbf{P}_0^{-1} + \mathbf{P}_0^{-1}\mathrm{E}\left[\mathbf{z}_1\mathbf{z}_1^T - \mathbf{z}_1\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_0\mathbf{z}_1^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T\right]\mathbf{P}_0^{-1}$$

$$\therefore\ \mathbf{I} = \mathbf{P}_0^{-1}\mathrm{E}\left[\mathbf{z}_1\mathbf{z}_1^T - \mathbf{z}_1\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_0\mathbf{z}_1^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T\right]$$

よって

$$\mathbf{P}_0 = \mathrm{E}\left[\mathbf{z}_1\mathbf{z}_1^T - \mathbf{z}_1\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_0\mathbf{z}_1^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T\right]$$

(13.110)より $\boldsymbol{\mu}_0 = \mathrm{E}[\mathbf{z}_1]$

$$= \mathrm{E}[\mathbf{z}_1\mathbf{z}_1^T] - \mathrm{E}[\mathbf{z}_1]\mathrm{E}[\mathbf{z}_1^T] - \mathrm{E}[\mathbf{z}_1]\mathrm{E}[\mathbf{z}_1^T] + \mathrm{E}[\mathbf{z}_1]\mathrm{E}[\mathbf{z}_1^T]$$

$$= \mathrm{E}[\mathbf{z}_1\mathbf{z}_1^T] - \mathrm{E}[\mathbf{z}_1]\mathrm{E}[\mathbf{z}_1^T] \quad \cdots (13.111)$$

を得る。