$\boxed{10.1}$ $\mathcal{L}(q) = \int dz\, q(z) \ln\left(\frac{p(x,z)}{q(z)}\right)$

$\mathrm{KL}(q\|p) = -\int dz\, q(z) \ln\left(\frac{p(z|x)}{q(z)}\right).$

$$\rightarrow \mathcal{L}(q) + \mathrm{KL}(q\|p) = \int dz\, q(z) \ln\left(\frac{p(x,z)}{p(z|x)}\right) + \int dz \left(q(z)\ln q(z) - q(z)\ln q(z)\right)$$

$$= \int dz\, q(z) \ln p(x).$$

$$= \ln p(x) \qquad \left(\int dz\, q(z) = 1\right)$$

$$\therefore \underline{\underline{\ln p(x) = \mathcal{L}(q) + \mathrm{KL}(q\|p)}} \quad \blacksquare$$

$\boxed{10.2}$ $\mathbb{E}[z_i] = m_i \quad (i=1,2)$

(10.13) $m_1 = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(\mathbb{E}[z_2] - \mu_2)$

(10.15) $m_2 = \mu_2 - \Lambda_{22}^{-1}\Lambda_{21}(\mathbb{E}[z_1] - \mu_1) \rightarrow m_2 = \mu_2 - \Lambda_{22}^{-1}\Lambda_{21}(\mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(m_2 - \mu_2) - \mu_1)$

$$= \mu_2 + \Lambda_{22}^{-1}\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}(m_2 - \mu_2)$$

$$\rightarrow (m_2 - \mu_2) = \Lambda_{22}^{-1}\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}(m_2 - \mu_2) \quad \cdots\cdots (*)$$

If. $\Lambda_{22}^{-1}\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} = I$. $\Lambda_{22} = \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}$. then $\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{12}\Lambda_{11}^{-1}\Lambda_{12} \end{pmatrix}$ $\because \Lambda_{12} = \Lambda_{21}$.

$(\Lambda_{11}\ \Lambda_{12}) \xrightarrow{\times \Lambda_{11}^{-1}\Lambda_{12}} (\Lambda_{12}\ \Lambda_{12}\Lambda_{11}^{-1}\Lambda_{12})$ this means. matrix. Λ is singular.

In this case, we cannot define. $p(z) = \mathcal{N}(z|\mu, \Lambda^{-1})$ Hence. $\Lambda_{22}^{-1}\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} \neq I$.

We can conclude. $m_2 = \mu_2$ from $(*)$, Similary $m_1 = \mu_1$.

$$\Longrightarrow \underline{\underline{\mathbb{E}[z_i] = \mu_i}} \quad \blacksquare$$

$\boxed{10.3}$ (10.5) $q(z) = \prod_{i=1}^{M} q_i(z_i)$

$$\mathrm{KL}(p\|q) = -\int dz\, p(z)\left(\sum_{i=1}^{M} \ln q_i(z_i)\right) + (\text{const.}) = -\int dz\, p(z). \left(\ln q_i(z_i) + \overbrace{\ln \prod_{j\neq i} q_j(z_j)}^{\text{const for } q_i(z_i).}\right) + (\text{const.})$$

$$= -\int \prod_{i=1}^{M} p(z) \ln q_i(z_i).\, dz_i.$$

$$= -\int dz_i \ln q_i(z_i) \int p(z) \prod_{j\neq i} dz_j$$

for i, use Lagrange multipier.

$$\mathcal{L} = \mathrm{KL}(p\|q) + \lambda\left(\int dz_i\, q_i(z_i) - 1\right)$$

$$\frac{\partial \mathcal{L}}{\partial q_i(z_i)} = -\frac{1}{q_i}\int p(z) \prod_{j\neq i} dz_j + \lambda = 0$$

$$\rightarrow \lambda q_i(z_i) = \int p(z) \prod_{j\neq i} dz_j. \quad \cdots\cdots (*)$$

$\int dz_i.$ $\lambda = \int p(z) \prod_{j=1} dz_j = \int p(z) dz = 1 \quad \therefore \lambda = 1.$

$$(*) \Longrightarrow \underline{\underline{q_i(z_i) = \int p(z) \prod_{j\neq i} dz_j.}} \quad \blacksquare.$$

$\boxed{10.4}$ $p(x)$: some fixed distribution.

$q(x) = \mathcal{N}(x|\mu, \Sigma)$

$$KL(p\|q) = -\int dx\, p(x) \ln \frac{q(x)}{p(x)} = +\int dx\, p(x) \ln p(x) - \int dx\, p(x) \ln \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$= (\text{const. for } \mu, \Sigma) - \int dx\, p(x) \left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$= (\text{const. for } \mu, \Sigma) + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}\int dx\, p(x)(x-\mu)^T\Sigma^{-1}(x-\mu)$$

$$\frac{\partial}{\partial\mu}\left(KL(p\|q)\right) = \Sigma^{-1}\int dx\, p(x)(x-\mu) = 0 \quad \therefore \underline{\underline{\mu = \int dx\, p(x)\, x}}$$

$\frac{\partial}{\partial\Sigma}\left(KL(p\|q)\right)$. First of all, $\frac{\partial}{\partial\Sigma_{ij}} \log\det\Sigma = \frac{1}{\det\Sigma}\frac{\partial}{\partial\Sigma_{ij}}\det\Sigma = \frac{1}{\det\Sigma}\det\Sigma\,(\Sigma^{-1})_{ji} = (\Sigma^{-1})_{ji}$

$$\frac{\partial}{\partial\Sigma_{ij}}\left(KL(p\|q)\right) = \frac{1}{2}(\Sigma^{-1})_{ji} - \frac{1}{2}\int dx\, p(x)(x-\mu)_j \Sigma^{-2}_{ji}(x-\mu)_i = 0.$$

$$\times\, \Sigma^2_{ji} \qquad \underline{\underline{\Sigma_{ji} = \int dx\, p(x)(x-\mu)_j(x-\mu)_i}} \quad \blacksquare$$

cofactor matrix $C_{ij} = (-1)^{ij} M_{ij}$

$\det(V) = \sum_{i=1}^{n} v_{ij} C_{ij}$ fixed j.

$\left(\det(V) = \sum_{j=1}^{n} v_{ij} C_{ij} \text{ fixed } i\right)$

$\mathrm{adj}(V) = C^T$

$V^{-1} = \frac{1}{\det(V)}\mathrm{adj}(V)$

$\frac{d}{dv_{ij}}\log(\det(V)) = \frac{1}{\det(V)}C_{ij}$

$= (V^{-1})_{ji}$

$\boxed{10.5}$ $q(z,\theta) = q_z(z)\, q_\theta(\theta)$ & $q_\theta(\theta) = \delta(\theta - \theta_0)$

$$KL(q\|p) = -\iint dz\,d\theta\, q_z(z) q_\theta(\theta) \ln\left(\frac{p(z,\theta|X)}{q_z(z) q_\theta(\theta)}\right)$$

$$= -\int dz\, q_z(z) \ln\left(\frac{p(z,\theta_0|X)}{q_z(z) q_\theta(\theta_0)}\right) \qquad \int d\theta\, \delta(\theta-\theta_0) f(\theta) := f(\theta_0)$$

$$= -\int dz\, q_z(z) \ln\left(\frac{p(z|\theta_0, X)}{q_z(z)}\right) + (\text{const. for } z)$$

Minimize this $KL(q\|p)$, we get $q_z(z) = p(z|\theta_0, X)$ ~ E step of the EM algorithm ~

To find $q_\theta(\theta)$,

$$\int dz\,d\theta\, q_z(z) q_\theta(\theta) \ln\left(\frac{p(X,\theta,z)}{q_\theta(\theta) q_z(z)}\right) = \int d\theta\, q_\theta(\theta)\, \mathbb{E}_{q_z(z)}\left[\ln p(X,\theta,z)\right] - \underline{\int d\theta\, q_\theta(\theta) \ln q_\theta(\theta)} + (\text{const. for } \theta)$$

$$\underset{q_\theta(\theta) = \delta(\theta-\theta_0)}{=} \mathbb{E}_{q_z(z)}\left[\ln p(X,\theta_0,z)\right] + (\text{const. for } \theta_0)$$

contribution from ONLY point. θ_0 (not distributed) → constant.

Maximize $\mathbb{E}_{q_z(z)}\left[\ln p(X,\theta_0,z)\right]$. ~ M step of the EM algorithm ~ $\blacksquare$

$\boxed{10.6}$ (10.19) $D_\alpha(p\|q) = \frac{4}{1-\alpha^2}\left(1 - \int dx\, p(x)^{\frac{1+\alpha}{2}} q(x)^{\frac{1-\alpha}{2}}\right)$

Take $\alpha \to 1$ limit. $\to$ $\alpha = 1+\epsilon$ & take $\epsilon \to 0$ limit.

$$p(x)^{\frac{1+\alpha}{2}} = p(x)^{1+\frac{\epsilon}{2}} = p(x)\, p(x)^{\frac{\epsilon}{2}} = p(x) \cdot e^{\frac{\epsilon}{2}\ln p(x)}.$$

$$\left(\begin{aligned} f(\epsilon) = e^{\frac{\epsilon}{2}\ln p(x)} &= f(0) + f'(0)\,\epsilon + O(\epsilon^2) \\ &= 1 + \frac{\ln p(x)}{2}\,\epsilon + O(\epsilon^2) \end{aligned} \right)$$

Using this result,

$$D_\alpha(p\|q) = \frac{4}{1-(1+\epsilon)^2}\left(1 - \int dx\, p(x)\left(1+\frac{\epsilon}{2}\ln p(x)\right)\left(1-\frac{\epsilon}{2}\ln q(x)\right)\right) + O(\epsilon^2)$$

$$= + \frac{4}{2\epsilon\left(1+\frac{\epsilon}{2}\right)} \int dx\, \frac{\epsilon}{2}\, p(x)\left(\ln p(x) - \ln q(x)\right) + O(\epsilon^2).$$

$$= + \int dx\, p(x) \ln \frac{p(x)}{q(x)}$$

$$= \underline{-\int dx\, p(x) \ln \frac{q(x)}{p(x)}}$$

$\boxed{10.7}$ (10.25) $\ln q^*_\mu(\mu) = -\frac{\mathbb{E}[\tau]}{2}\left\{\lambda_0(\mu-\mu_0)^2 + \sum_{n=1}^{N}(x_n-\mu)^2\right\} + (\text{const.})$

Complete the square.

$$\ln q^*_\mu(\mu) = -\frac{\mathbb{E}[\tau]}{2}\cdot\left(\begin{aligned} &\lambda_0\mu^2 - 2\mu_0\lambda_0\mu + \lambda_0\mu_0^2 \\ &N\mu^2 - 2\sum_{n=1}^{N}x_n\mu + \sum_{n=1}^{N}x_n^2 \end{aligned}\right) + (\text{const.})$$

$$= -\frac{\mathbb{E}[\tau]}{2}(\lambda_0+N)\left(\mu^2 - 2\cdot\frac{\mu_0\lambda_0 + \sum_{n=1}^{N}x_n}{\lambda_0+N}\mu\right) + (\text{const.})$$

$$= -\underbrace{\frac{\mathbb{E}[\tau]}{2}(\lambda_0+N)}_{\lambda_N}\left(\mu - \underbrace{\frac{\mu_0\lambda_0 + \sum_{n=1}^{N}x_n}{\lambda_0+N}}_{\mu_N.}\right)^2 + (\text{const.})$$

Hence,

$$\underline{\mu_N = \frac{\mu_0\lambda_0 + \sum_{n=1}^{N}x_n}{\lambda_0+N}}, \quad \underline{\lambda_N = (\lambda_0+N)\,\mathbb{E}[\tau]}$$

(10.28) $\ln q^*_\tau(\tau) = (a_0-1)\ln\tau - b_0\tau + \frac{1}{2}\ln\tau + \frac{N}{2}\ln\tau - \frac{\tau}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right] + (\text{const.})$ $\cdots$ (*).

gamma distribution. $f(x) = \frac{1}{\beta^\alpha\Gamma(\alpha)}\exp\left(-\frac{x}{\beta}\right)x^{\alpha-1}$

$$\to \ln f(x) = -\alpha\ln\beta - \ln\Gamma(\alpha) \underbrace{-\frac{x}{\beta}}_{②} + \underbrace{(\alpha-1)\ln x}_{①}.$$

Rearrange (*),

$$\ln q_\tau^*(\tau) = \underbrace{\left(a_0 - \frac{1}{2} + \frac{N}{2}\right)\ln\tau}_{①} \underbrace{- \tau\left(b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right]\right)}_{③} + (\text{const.})$$

① : $\alpha = a_0 + \frac{N+1}{2}$

③ : $\beta = \left(b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right]\right)^{-1}$

Note that the definition of gamma distribution in PRML is $\mathrm{Gam}(\lambda \mid a, b) = \frac{1}{\Gamma(\alpha)} b^a \lambda^{a-1} e^{-b\lambda}$,

$$\underline{a_N = a_0 + \frac{N+1}{2}}$$

$$\underline{b_N = b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right]}$$

$\boxed{10.8}$. (10.29) $a_N = a_0 + \frac{N+1}{2}$

(10.30) $b_N = b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right]$

$$\left(\begin{array}{ll} (B.26) & \mathrm{Gam}(\tau \mid a, b) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} e^{-b\tau} \\ (B.27) & \mathbb{E}[\tau] = \frac{a}{b} \\ (B.28) & \mathrm{var}[\tau] = \frac{a}{b^2} \end{array}\right)$$

$$\mathbb{E}[\tau] = \frac{a_N}{b_N} = \frac{a_0 + \frac{1}{2} + \frac{N}{2}}{b_0 + \frac{1}{2}\mathbb{E}_\mu[\lambda_0(\mu-\mu_0)^2] + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2\right]}$$

Picking up only $O(N)$ dependence terms. (this tends to be exact when $N \to \infty$)

$$\mathbb{E}[\tau] = \frac{N}{\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2\right]} = \underline{\left(\mathbb{E}_\mu\left[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2\right]\right)^{-1}}$$

inverse of the maximum likelihood estimator.

$$\underline{\mathrm{var}[\tau] = \frac{a_N}{b_N^2} \xrightarrow{N\to\infty} 0.} \quad \because \; b_N \text{ is } O(N) \text{ and } b_N^2 \text{ is } O(N^2).$$

$\boxed{10.9}$. Derive (10.33) $\frac{1}{\mathbb{E}[\tau]} = \left(\overline{x^2} - \overline{x}^2\right) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2$

Standard result: $\mathbb{E}[\tau] = \frac{a_N}{b_N}$ for the mean of a gamma distribution.

$$\frac{1}{\mathbb{E}[\tau]} = \frac{b_N}{a_N} = \frac{b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right]}{a_0 + \frac{N+1}{2}} \quad \leftarrow \quad \begin{array}{ll} (10.29) & a_N = a_0 + \frac{N+1}{2} \\ (10.30) & b_N = b_0 + \frac{1}{2}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right] \end{array}$$

prior. $\mu_0 = a_0 = b_0 = \lambda_0 = 0$.

$$\frac{1}{\mathbb{E}[\tau]} = \frac{1}{N+1}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2\right] = \frac{1}{N+1}\mathbb{E}_\mu\left[\sum_{n=1}^{N}(x_n^2 - 2\mu x_n + \mu^2)\right]$$

$$= \frac{N}{N+1}\left(\overline{x^2} - 2\overline{x}\,\mathbb{E}[\mu] + \mathbb{E}[\mu^2]\right) \quad \text{where } \overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$\cdots (*)$

Using (10.26) $\mu_N = \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N} \xrightarrow{\mu_0=\lambda_0=0} \bar{x}$ $\therefore$ $\mathbb{E}[\mu] = \bar{x}$

(10.27) $\lambda_N = (\lambda_0 + N)\mathbb{E}[\tau] \xrightarrow{\lambda_0=0} N\mathbb{E}[\tau]$ $\mathrm{Var}[\mu] = \mathbb{E}[\mu^2] - (\mathbb{E}[\mu])^2$ $\therefore$ $\mathbb{E}[\mu^2] = \bar{x}^2 + \frac{1}{N\mathbb{E}[\tau]}$

Substituting these results into (*),

$$\frac{1}{\mathbb{E}[\tau]} = \frac{N}{N+1}\left(\overline{x^2} - 2\bar{x}\cdot\bar{x} + \bar{x}^2 + \frac{1}{N\mathbb{E}[\tau]}\right)$$

$$= \frac{N}{N+1}\left(\overline{x^2} - \bar{x}^2\right) + \frac{1}{N+1}\cdot\frac{1}{\mathbb{E}[\tau]}$$

$$\rightarrow \underline{\frac{1}{\mathbb{E}[\tau]} = \left(\overline{x^2} - \bar{x}^2\right) \equiv \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})^2}$$

$\boxed{10.10}$ Derive (10.34) $\ln p(X) = \mathcal{L} - \sum_m\sum_Z q(Z|m)q(m)\ln\left(\frac{p(Z,m|X)}{q(Z|m)q(m)}\right)$ where $\mathcal{L} = \sum_m\sum_Z q(Z|m)q(m)\ln\left(\frac{p(Z,X,m)}{q(Z|m)q(m)}\right)$ ..... (*)

$$\sum_m\sum_Z q(Z|m)q(m)\ln\left(\frac{p(Z,X,m)}{q(Z|m)q(m)}\right) - \sum_m\sum_Z q(Z|m)q(m)\ln\left(\frac{p(Z,m|X)}{q(Z|m)q(m)}\right)$$

$$= \sum_m\sum_Z q(Z,m)\left(\ln\left(\frac{p(Z,X,m)}{p(Z,m|X)}\right) - \ln q(Z,m) + \ln q(Z,m)\right)$$

$$= \sum_m\sum_Z q(Z,m)\ln\frac{p(Z,m|X)p(X)}{p(Z,m|X)}$$

$$= \sum_m\sum_Z q(Z,m)\ln p(X)$$

$$= \underline{\ln p(X)}$$ Then (*) is proved.

$\boxed{10.11}$ $\mathcal{L} = \sum_m\sum_Z q(Z|m)q(m)\ln\left(\frac{p(Z,X,m)}{q(Z|m)q(m)}\right)$

$$= \sum_m\sum_Z q(Z|m)q(m)\left[\ln p(Z,X|m) + \ln p(m) - \ln q(Z|m) - \ln q(m)\right]$$

$$= \sum_m q(m)\left(\ln p(m) - \ln q(m)\right) + \sum_m\sum_Z q(Z|m)q(m)\left(\ln p(Z,X|m) - \ln q(Z|m)\right)$$

$$= \sum_m q(m)\left[\ln p(m) + \sum_Z q(Z|m)\{\ln p(Z,X|m) - \ln q(Z|m)\} - \ln q(m)\right]$$ $\mathcal{L}_m = \sum_Z q(Z|m)\ln\left\{\frac{p(Z,X|m)}{q(Z|m)}\right\}$

$$= \sum_m q(m)\left[\ln\{p(m)*\exp(\mathcal{L}_m)\} - \ln q(m)\right]$$

Using Lagrange multiplier.

$$\mathcal{L}' = \mathcal{L} + \lambda\left(\sum_m q(m) - 1\right)$$

$$\frac{\partial\mathcal{L}'}{\partial q(m)} = \ln\{p(m)*\exp(\mathcal{L}_m)\} - \ln q(m) - 1 + \lambda = 0. \quad \therefore \ln q(m) = (-1+\lambda) + \ln\{p(m)e^{\mathcal{L}_m}\}$$

Hence,

$$q(m) = e^{-1+\lambda}e^{\ln p(m)e^{\mathcal{L}_m}} = e^{-1+\lambda}p(m)\exp(\mathcal{L}_m) \propto \underline{p(m)\exp(\mathcal{L}_m)}$$

$\boxed{10.12}$ (10.41) $p(X,Z,\pi,\mu,\Lambda) = p(X|Z,\mu,\Lambda)\,p(Z|\pi)\,p(\pi)\,p(\mu|\Lambda)\,p(\Lambda)$

general result. (10.9) $\ln q_j^*(Z_j) = \mathbb{E}_{i\neq j}[\ln p(X,Z)] + (\text{const.})$

From the general result,

$$\ln q^*(Z) = \mathbb{E}_{\pi,\mu,\Lambda}[\ln p(X,Z,\pi,\mu,\Lambda)]$$
$$= \mathbb{E}_{\pi,\mu,\Lambda}[\ln p(Z|\pi) + \ln p(X|Z,\mu,\Lambda)] + (\text{const. for } Z)$$

Note that $\left\{\begin{array}{ll} p(Z|\pi) = \prod_{n=1}^{N}\prod_{k=1}^{K} \pi_k^{z_{nk}} & (10.37) \\ p(X|Z,\mu,\Lambda) = \prod_{n=1}^{N}\prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k,\Lambda_k^{-1})^{z_{nk}} & (10.38) \end{array}\right\}$, then

$$\ln q^*(Z) = \mathbb{E}_{\pi,\mu,\Lambda}\left[\sum_{n=1}^{N}\sum_{k=1}^{K}\{z_{nk}\ln\pi_k + z_{nk}\ln\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1})\}\right] + (\text{const.})$$
$$= \underbrace{\sum_{n=1}^{N}\sum_{k=1}^{K}}_{z_{nk}}\left[\mathbb{E}_\pi[\ln\pi_k] - \frac{D}{2}\ln(2\pi) + \frac{1}{2}\mathbb{E}_\Lambda[\ln|\Lambda_k|] - \frac{1}{2}\mathbb{E}_{\mu_k,\Lambda_k}[(x_n-\mu_k)^T\Lambda_k(x_n-\mu_k)]\right] + (\text{const.})$$
$$:= \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\ln\rho_{nk} + (\text{const.})$$

Taking the exponential,

$$q^*(Z) \propto \prod_{n=1}^{N}\prod_{k=1}^{K}\rho_{nk}^{z_{nk}}$$

Consider the normalization.

$\sum_Z q^*(Z) = 1$. note that $\sum_k z_{nk} = 1$ for $^\forall n$. $\rightarrow$ define the factor for each $\rho_{nk}^{z_{nk}}$ with fixed n.

$$\rightarrow \sum_Z \prod_{k=1}^{K}\rho_{nk}^{z_{nk}} = \sum_Z \rho_{n1}^{z_{n1}}\cdots\rho_{nK}^{z_{nK}}$$

$\sum_Z$ picks up only one $z_{ni}=1$ once. (others 0) then

$$\rightarrow \sum_Z\prod_{k=1}^{K}\rho_{nk}^{z_{nk}} = \rho_{n1} + \rho_{n2} + \cdots \rho_{nK} = \sum_{k=1}^{K}\rho_{nk}$$

Using this factor, $\rho_{nk}^{z_{nk}} \longrightarrow \left(\frac{\rho_{nk}}{\sum_{k=1}^{K}\rho_{nk}}\right)^{z_{nk}}$ $\leftarrow$ out side of the bracket because we don't need the normalization for $z_{nk}=0$ case.

Finally, we get.

$$\underline{\underline{q^*(Z) = \prod_{n=1}^{N}\prod_{k=1}^{K}\left(\frac{\rho_{nk}}{\sum_{k'=1}^{K}\rho_{nk'}}\right)^{z_{nk}}}}$$

■

$\boxed{10.13}$ (10.54) $\ln q^*(\pi, \mu, \Lambda) = \ln p(\pi) + \sum_{k=1}^{K} \ln p(\mu_k, \Lambda_k) + \mathbb{E}_z[\ln p(z|\pi)] + \sum_{k=1}^{K}\sum_{n=1}^{N} \mathbb{E}[z_{nk}] \ln \mathcal{N}(x_n|\mu_k, \Lambda_k^{-1}) + (\text{const.})$

Factorization $q(\pi, \mu, \Lambda) = q(\pi)\prod_{k=1}^{K} q(\mu_k, \Lambda_k)$, we can separate the above eq. $\ln q^*(\pi, \mu, \Lambda) = \ln q^*(\pi) + \sum_{k=1}^{K} \ln q^*(\mu_k, \Lambda_k)$

Then,

$$\ln q^*(\mu_k, \Lambda_k) = \ln p(\mu_k, \Lambda_k) + \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \ln \mathcal{N}(x_n|\mu_k, \Lambda_k^{-1}) + (\text{const for } \mu_k, \Lambda_k)$$

Note that. (10.40) $p(\mu, \Lambda) = p(\mu|\Lambda)p(\Lambda) = \prod_{k=1}^{K} \mathcal{N}(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1})\, \mathcal{W}(\Lambda_k | W_0, \nu_0).$, so.

$$\ln q^*(\mu_k, \Lambda_k) = \ln \mathcal{N}(\mu_k | m_0, (\beta_0\Lambda_k)^{-1}) + \ln \mathcal{W}(\Lambda_k | W_0, \nu_0) + \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \ln \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1}) + (\text{const.})$$

$$\mathcal{W}(\Lambda, W, \nu) = B(W, \nu)|\Lambda|^{\frac{\nu - D - 1}{2}} \exp\left(-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)\right)$$

(inverse Wishar distribution is the conjugate prior of covariance matrix of multivariate gaussian)

$$= -\frac{\beta_0}{2}(\mu_k - m_0)^T \Lambda_k (\mu_k - m_0) + \frac{1}{2}\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}(\Lambda_k W_0^{-1}) + \frac{\nu_0 - D - 1}{2}\ln|\Lambda_k| - \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[z_{nk}](x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$
$$+ \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[z_{nk}] \ln|\Lambda_k| + (\text{const.}) \quad \cdots (*)$$

Due to the product rule, $\ln q^*(\mu_k, \Lambda_k) = \ln q^*(\mu_k|\Lambda_k) + \ln q^*(\Lambda_k)$, we can identify

$$\ln q^*(\mu_k|\Lambda_k) = -\frac{1}{2}\mu_k^T\left(\beta_0 + \sum_{n=1}^{N}\mathbb{E}[z_{nk}]\right)\Lambda_k \mu_k + \mu_k^T \Lambda_k \left(\beta_0 m_0 + \sum_{n=1}^{N}\mathbb{E}[z_{nk}] x_n\right) + (\text{const.})$$
$$= -\frac{1}{2}\mu_k^T(\beta_0 + N_k)\Lambda_k \mu_k + \mu_k^T \Lambda_k (\beta_0 m_0 + N_k \bar{x}_k) + (\text{const.}) \quad \cdots\cdots (**)$$

where (10.50) $\mathbb{E}[z_{nk}] = r_{nk}$, (10.51) $N_k = \sum_{n=1}^{N} r_{nk}$, (10.52) $\bar{x}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk} x_n$ are used.

Now we find (**) depends μ_k quadratically $\Longrightarrow$ $\ln q^*(\mu_k|\Lambda_k)$ is a Gaussian distribution.

$$\ln q^*(\mu_k|\Lambda_k) = -\frac{1}{2}(\beta_0 + N_k)\left[\mu_k^T \Lambda_k \mu_k - 2\frac{\beta_0 m_0 + N_k \bar{x}_k}{\beta_0 + N_k}\mu_k^T \Lambda_k\right] + (\text{const.})$$
$$= -\frac{1}{2}(\beta_0 + N_k)\left(\mu_k - \frac{\beta_0 m_0 + N_k \bar{x}_k}{\beta_0 + N_k}\right)^T \Lambda_k \left(\mu_k - \frac{\beta_0 m_0 + N_k \bar{x}_k}{\beta_0 + N_k}\right) + (\text{const.})$$

$\therefore$ $q^*(\mu_k|\Lambda_k) = \mathcal{N}(\mu_k | m_k, \beta_k \Lambda_k)$ where $\beta_k = \beta_0 + N_k$, $m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k \bar{x}_k)$ //

Next, $\ln q^*(\Lambda_k)$

$$\ln q^*(\Lambda_k) = \ln q^*(\Lambda_k, \mu_k) - \ln q^*(\mu_k|\Lambda_k)$$
$$= -\frac{\beta_0}{2}(\mu_k - m_0)^T \Lambda_k (\mu_k - m_0) + \frac{\nu_0 - D}{2}\ln|\Lambda_k| + \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[z_{nk}]\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}[\Lambda_k W_0^{-1}] - \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[z_{nk}](x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$
$$+ \frac{1}{2}\ln|\Lambda_k| + \frac{\beta_k}{2}(\mu_k - m_k)^T \Lambda_k (\mu_k - m_k)$$

$$= \frac{(\nu_0 + N_k + 1) - D}{2} \ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W_0^{-1} + \beta_0(\mu_k - m_0)(\mu_k - m_0)^T + \sum_{n=1}^{N} \mathbb{E}[z_{nk}](x_n - \mu_k)(x_n - \mu_k)^T$$

$$- \beta_k(\mu_k - m_k)(\mu_k - m_k)^T\Big\}\Big].$$

Use $\beta_k = \beta_0 + N_k$, $m_k = \frac{1}{\beta_0 + N_k}(\beta_0 m_0 + N_k \bar{x}_k)$

$$:= \frac{\nu_k - D}{2}\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W_0^{-1} + \cancel{\beta_0 \mu_k \mu_k^T} - 2\cancel{\beta_0 \mu_k} m_0^T + \beta_0 m_0 m_0 + \cancel{N_k \mu_k} \mu_k^T - 2\cancel{\sum \mathbb{E}[z] \mu_k} x_k^T + \underline{\underline{\sum_{n=1}^{N} \mathbb{E}[z] x_k x_k^T}} - \cancel{(\beta_0 + N_k)} \mu_k \mu_k^T + 2\beta_0 \mu_k m_0^T \cancel{+ 2 N_k} \mu_k \bar{x}_k - \frac{1}{\beta_0 + N_k}(\beta_0 m_0 + N_k \bar{x}_k)(\beta_0 m_0^T + N_k \bar{x}_k^T)\Big\}\Big]$$

$$\sum_{n=1}^{N} \mathbb{E}[z_{nk}](x_n - \bar{x}_n)(x_n - \bar{x}_n)^T + N_k \bar{x}_n \bar{x}_n^T$$

$$= N_k S_k + N_k \bar{x}_n \bar{x}_n^T$$

$$= \frac{\nu_k - D}{2}\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W_0^{-1} + N_k S_k + \underbrace{\beta_0 m_0 m_0^T + N_k \bar{x}_n \bar{x}_n^T - \frac{\beta_0^2 m_0 m_0^T + 2\beta_0 N_k m_0 \bar{x}_0^T + N_k^2 \bar{x}_k \bar{x}_k^T}{\beta_0 + N_k}}_{\frac{1}{\beta_0 + N_k}(\beta_0 N_k m_0 m_0^T + \beta_0 N_k \bar{x}_n \bar{x}_n - 2\beta_0 N_k m_0 \bar{x}_0^T)}\Big\}\Big]$$

$$= \frac{\beta_0 N_k}{\beta_0 + N_k}(m_0 - \bar{x}_k)(m_0^T - \bar{x}_n^T).$$

$$= \frac{\nu_k - D}{2}\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}\left[\Lambda_k W_k^{-1}\right]$$

where $\nu_k = \nu_0 + N_k + 1$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(m_0 - \bar{x}_k)(m_0^T - \bar{x}_n^T)$$

From this observation, get. $\underline{q^*(\Lambda_k) = \mathcal{W}(\Lambda_k | W_k, \nu_k)}$ ∎

$\boxed{10.14}$ (10.59) $q^*(\mu_k, \Lambda_k) = \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1})\mathcal{W}(\Lambda_k | W_k, \nu_k)$

$$\mathbb{E}_{\mu_k, \Lambda_k}[(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)] = \int d\mu_k d\Lambda_k \; q^*(\mu_k, \Lambda_k)(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$

$$= \int d\mu_k d\Lambda_k \; \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1})\mathcal{W}(\Lambda_k | W_k, \nu_k)(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$

$$= \int d\Lambda_k \; \mathcal{W}(\Lambda_k | W_k, \nu_k) \cdot \int d\mu_k \; \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) \cdot (x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$

First,

$$\int d\mu_k \; \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1})(x_n - \mu_k)_i \Lambda_{k,ij}(x_n - \mu_k)_j$$

$$= \int d\mu_k \; \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) \quad \Lambda_{k,ij}\left(x_{n,i} x_{n,j} - 2x_{n,i}\mu_{k,j} + \mu_{k,i}\mu_{k,j}\right)$$

$$= \Lambda_{k,ij}\left(x_{n,i}x_{n,j} - 2x_{n,i}m_{k,j} + \beta_k^{-1}\Lambda_{k,ij}^{-1} + m_{k,i}m_{k,j} \right)$$

$\left(\text{where } E[x_i x_j] - E[x_i]E[x_j] = \Sigma_{ij} \right)$

$$= \beta_k^{-1}D + (x_n - m_k)^T \Lambda_k (x_n - m_k)$$

$$\Lambda_{k,ij}\Lambda_{k,ji}^{-1} = Tr[\Lambda\Lambda^{-1}] = Tr[I] = D$$

Then,

$$\int d\Lambda_k \, W(\Lambda_k|W_k, \nu_k) \left\{ \beta_k^{-1}D + (x_n - m_k)^T \Lambda_k (x_n - m_k) \right\}$$

$$= \beta_k^{-1}D \underbrace{\int d\Lambda_k W(\Lambda_k|W_k,\nu_k)}_{\to 1} + (x_n - m_k)^T \underbrace{\left\{ \int d\Lambda_k \, W(\Lambda_k|W_k,\nu_k)\Lambda_k \right\}}_{\to \nu_k W_k \quad (B.80)} (x_n - m_k)$$

$$= \underline{\beta_k^{-1} \cdot D + \nu_k (x_n - m_k)^T \cdot W_k (x_n - m_k)}$$

$\boxed{10.15}$ Dirichlet distribution. (B.17) $\mathbb{E}[\mu_k] = \frac{\alpha_k}{\hat{\alpha}}$ where $\hat{\alpha} = \sum_{k=1}^{K} \alpha_k$

$$\mathbb{E}[\pi_k] = \frac{\alpha_k}{\hat{\alpha}} = \frac{\alpha_0 + N_k}{\sum_{k=1}^{K}(\alpha_0 + N_k)}$$

(10.58) $\alpha_k = \alpha_0 + N_k$

$$= \underline{\frac{\alpha_0 + N_k}{K\alpha_0 + N}}$$

$N \equiv \sum_{k=1}^{K} N_k$

$\boxed{10.16}$ (10.70) $\mathcal{L} = \mathbb{E}[\ln p(X|Z,\mu,\Lambda)] + \mathbb{E}[\ln p(Z|\pi)] + \mathbb{E}[\ln p(\pi)] + \mathbb{E}[\ln p(\mu,\Lambda)] - \mathbb{E}[\ln q(Z)] - \mathbb{E}[\ln q(\pi)] - \mathbb{E}[\ln q(\mu,\Lambda)]$

(10.38) $p(X|Z,\mu,\Lambda) = \prod_{n=1}^{N}\prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Lambda_k^{-1})^{z_{nk}}$

$$\mathbb{E}[\ln p(X|Z,\mu,\Lambda)] = \mathbb{E}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk} \left\{ \frac{1}{2}\ln|\Lambda_k| - \frac{D}{2}\ln(2\pi) - (x_n - \mu_k)\Lambda_k(x_n - \mu_k) \right\} \right]$$

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}[z_{nk}] \left\{ \mathbb{E}[\ln|\Lambda_k|] - D\ln(2\pi) - \mathbb{E}[(x_n - \mu_k)\Lambda_k(x_n - \mu_k)] \right\}$$

$$= \underline{\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk} \left\{ \ln\tilde{\Lambda}_k - D\ln(2\pi) - D\beta_k^{-1} - \nu_k Tr[S_k W_k] - \nu_k(\bar{x}_k - m_k)^T W_k(\bar{x}_k - m_k) \right\}}$$

Here (10.64) & (10.65) are used.

(10.37) $p(Z|\pi) = \prod_{n=1}^{N}\prod_{k=1}^{K} \pi_k^{z_{nk}}$

$$\mathbb{E}[\ln p(Z|\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}[z_{nk}]\mathbb{E}[\ln \pi_k] = \underline{\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk} \ln\tilde{\pi}_k}$$

$\boxed{10.17}$. (10.39) $p(\pi) = \mathrm{Dir}(\pi|\alpha_0) = C(\alpha_0)\prod_{k=1}^{K}\pi_k^{\alpha_0-1}$

$$\mathbb{E}[\ln p(\pi)] = \mathbb{E}\left[\ln C(\alpha_0) + \sum_{k=1}^{K}(\alpha_0-1)\ln\pi_k\right] = \underline{\underline{\ln C(\alpha_0) + (\alpha_0-1)\sum_{k=1}^{K}\ln\tilde{\pi}_k}}$$ ▨

(10.40) $p(\mu,\Lambda) = \prod_{k=1}^{K}\mathcal{N}(\mu_k|m_0,(\beta_0\Lambda_k)^{-1})\,\mathcal{W}(\Lambda_k|W_0,\nu)$

$$\mathbb{E}[\ln p(\mu,\Lambda)] = \sum_{k=1}^{K}\mathbb{E}\left[\ln\mathcal{N}(\mu_k|m_0,(\beta_0\Lambda_k)^{-1}) + \ln\mathcal{W}(\Lambda_k|W_0,\nu)\right]$$

$$= \sum_{k=1}^{K}\mathbb{E}\left[\frac{D}{2}\ln\left(\frac{\beta_0}{2\pi}\right) + \ln|\Lambda_k| - \beta_0(\mu_k-m_0)^T\Lambda_k(\mu_k-m_0) + \ln B(W_0,\nu) + \frac{(\nu-D-1)}{2}\ln|\Lambda_k| - \frac{1}{2}\mathrm{Tr}(W_0^{-1}\Lambda_k)\right]$$

$$= \frac{1}{2}\sum_{k=1}^{K}\left\{D\ln\left(\frac{\beta_0}{2\pi}\right) + \ln\tilde{\Lambda}_k - \frac{\beta_0}{\beta_k}D - \beta_0\nu_k(m_k-m_0)^T W_k(m_k-m_0)\right\}$$

$$+ K\ln B(W_0,\nu) + \frac{\nu-D-1}{2}\sum_{k=1}^{K}\ln\tilde{\Lambda}_k - \frac{1}{2}\sum_{k=1}^{K}\nu_k\mathrm{Tr}[W_0^{-1}W_k]$$

where we have used the result of $\boxed{10.14}$.

(10.48) $q^*(Z) = \prod_{n=1}^{N}\prod_{k=1}^{K} r_{nk}^{z_{nk}}$

$$\mathbb{E}[\ln q(Z)] = \sum_{n=1}^{N}\sum_{k=1}^{K}\mathbb{E}[z_{nk}]\,\mathbb{E}[\ln r_{nk}] = \underline{\underline{\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\ln r_{nk}}}$$

∵ r_{nk} is NOT dependent on parameters

(10.57) $q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$

$$\mathbb{E}[\ln q(\pi)] = \underline{\underline{\ln C(\alpha) + \sum_{k=1}^{K}(\alpha_k-1)\ln\tilde{\pi}_k}}$$

(Here we've used $\mathbb{E}[\ln p(\pi)]$ result)

(10.59) $q^*(\mu_k,\Lambda_k) = \mathcal{N}(\mu_k|m_k,(\beta_k\Lambda_k)^{-1})\,\mathcal{W}(\Lambda_k|W_k,\nu_k)$; $q(\mu,\Lambda) = \prod_{k=1}^{K} q(\mu_k,\Lambda_k)$

$$\mathbb{E}[\ln q(\mu,\Lambda)] = \sum_{k=1}^{K}\left\{\frac{D}{2}\ln\left(\frac{\beta_k}{2\pi}\right) + \ln\tilde{\Lambda}_k - \frac{D}{2}\right\} + \sum_{k=1}^{K}\mathbb{E}[\ln\mathcal{W}(\Lambda_k|W_k,\nu_k)]$$

← ($m_0 \to m_k$, $W_0 \to W_k$ for $\mathbb{E}[\ln p(\mu,\Lambda)]$)

$$= \underline{\underline{\sum_{k=1}^{K}\left\{\frac{D}{2}\ln\left(\frac{\beta_k}{2\pi}\right) - \frac{D}{2} + \ln\tilde{\Lambda}_k\right\} - \sum_{k=1}^{K}\mathrm{H}[\underbrace{\mathcal{W}(\Lambda_k|W_k,\nu_k)}_{q(\Lambda_k)}]}}$$

$(\mathrm{H}[p] := -\int dx\, p\ln p)$

Hence, we've derived (10.73) ~ (10.77). ▨

$\boxed{10.18}$ (10.42) $q(Z,\pi,\mu,\Lambda) = q(Z)\, q(\pi,\mu,\Lambda) = q(Z)\, q(\pi) \prod_{k=1}^{K} q(\mu_k, \Lambda_k)$

(10.48) $q^*(Z) = \prod_{n=1}^{N}\prod_{k=1}^{K} r_{nk}^{z_{nk}}$

(10.57) $q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$

(10.59) $q^*(\mu_k,\Lambda_k) = \mathcal{N}(\mu_k|m_k, (\beta_k\Lambda_k)^{-1})\,\mathcal{W}(\Lambda_k|W_k,\nu_k)$

(10.70) $\mathcal{L} = \underset{②}{\mathbb{E}[\ln p(X|Z,\mu,\Lambda)]} + \underset{①}{\mathbb{E}[\ln p(Z|\pi)]} + \underset{①}{\mathbb{E}[\ln p(\pi)]} + \underset{②}{\mathbb{E}[\ln p(\mu,\Lambda)]} - \mathbb{E}[\ln q(Z)] - \underset{①}{\mathbb{E}[\ln q(\pi)]} - \underset{②}{\mathbb{E}[\ln q(\mu,\Lambda)]}$.

First, see parameter π.

①: $\mathbb{E}[\ln p(Z|\pi)] + \mathbb{E}[\ln p(\pi)] - \mathbb{E}[\ln q(\pi)]$

$$= \sum_Z \iiint d\pi\, d\mu\, d\Lambda.\ q(Z)\, q(\pi) \prod_{k=1}^{K} q(\mu_k,\Lambda_k) \left[\sum_{k=1}^{K}\sum_{n=1}^{N} z_{nk}\ln(\pi_k) + \ln C(\alpha_0) + \sum_{k=1}^{K}(\alpha_0-1)\ln(\pi_k) - \ln C(\alpha) - \sum_{k=1}^{K}(\alpha-1)\ln(\pi_k)\right]$$

$$= \sum_Z \int d\pi\ q(Z)\, q(\pi)\left[\sum_{k=1}^{K}\sum_{n=1}^{N} z_{nk}\ln(\pi_k) + \ln C(\alpha_0) + \sum_{k=1}^{K}(\alpha_0-1)\ln(\pi_k) - \ln C(\alpha) - \sum_{k=1}^{K}(\alpha-1)\ln(\pi_k)\right]$$

$$= \int d\pi.\ q(\pi)\left[\sum_{k=1}^{K}\sum_{n=1}^{N} r_{nk}\ln(\pi_k) + \ln\left(\frac{C(\alpha_0)}{C(\alpha)}\right) + \sum_{k=1}^{K}(\alpha_0-\alpha)\ln(\pi_k)\right]$$

$$= \sum_{k=1}^{K}\int d\pi\ q(\pi)\left[\Big(\underbrace{\sum_{n=1}^{N} r_{nk}}_{N_k} + \alpha_0 - \alpha\Big)\ln(\pi_k)\right] + (\text{const. for } \pi)$$

Note that $\mathcal{L} \leq 0$. To maximize this function, the parameter α should be. $\underline{\underline{\alpha = \alpha_0 + N_k}}$

same as (10.58)

②: $\mathbb{E}[\ln p(X|Z,\mu,\Lambda)] + \mathbb{E}[\ln p(\mu,\Lambda)] - \mathbb{E}[\ln q(\mu,\Lambda)]$

$$= \sum_Z \iiint d\pi\, d\mu\, d\Lambda\ q(Z)\, q(\pi)\prod_{k=1}^{K} q(\mu_k,\Lambda_k)\Big[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\ln\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1}) + \sum_{k=1}^{K}\big(\ln\mathcal{N}(\mu_k|m_0,(\beta_0\Lambda_k)^{-1}) + \ln\mathcal{W}(\Lambda_k|W_0,\nu_0)\big) - \sum_{k=1}^{K}\big(\ln\mathcal{N}(\mu_k|m,(\beta\Lambda_k)^{-1}) + \ln\mathcal{W}(\Lambda_k|W,\nu)\big)\Big]$$

$$= \iint d\mu\, d\Lambda \prod_{k=1}^{K} q(\mu_k,\Lambda_k) \sum_{k=1}^{K}\left[\begin{array}{ll}\sum_{n=1}^{N}\mathbb{E}[z_{nk}]\ln\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1}) & \\ \ln\mathcal{N}(\mu_k|m_0,(\beta_0\Lambda_k)^{-1}) & + \ln\mathcal{W}(\Lambda_k|W_0,\nu_0) \\ -\ln\mathcal{N}(\mu_k|m,(\beta\Lambda_k)^{-1}) & - \ln\mathcal{W}(\Lambda_k|W,\nu)\end{array}\right]$$

$$= {''} \left[\begin{array}{llll} N_k\frac{1}{2}\ln|\Lambda_k| & -\sum_{n=1}^{N}\mathbb{E}[z_{nk}(x_n-\mu_k)^T\Lambda_k(x_n-\mu_k)] & & \\ \frac{\beta_0}{2}\ln|\Lambda_k| & -\frac{\beta_0}{2}(\mu_k-m_0)^T\Lambda_k(\mu_k-m_0) & +\ln B(W_0,\nu_0) + \frac{\nu_0-D-1}{2}\ln|\Lambda_k| & -\frac{1}{2}\mathrm{Tr}[W_0^{-1}\Lambda_k] \\ -\frac{\beta}{2}\ln|\Lambda_k| & +\frac{\beta}{2}(\mu_k-m)^T\Lambda_k(\mu_k-m) & -\ln B(W,\nu) - \frac{\nu-D-1}{2}\ln|\Lambda_k| & +\frac{1}{2}\mathrm{Tr}[W^{-1}\Lambda_k]\end{array}\right]$$

$$= {''}\left(\left(\frac{N_k+\beta_0-\beta}{2}\right)\ln|\Lambda_k| + \frac{\nu_0-\nu}{2}\ln|\Lambda_k| + \frac{1}{2}\mathrm{Tr}[(W^{-1}-W_0^{-1})\Lambda_k] + \frac{1}{2}\mathrm{Tr}\left[\begin{array}{l}\beta\Lambda_k(\mu_k-m)(\mu_k-m)^T \\ -\beta_0\Lambda_k(\mu_k-m_0)(\mu_k-m_0)^T \\ -\sum_{n=1}^{N}\mathbb{E}[z_{nk}]\Lambda_k(x_n-\mu_k)(x_n-\mu_k)^T\end{array}\right]\right) + (\text{const.})$$

Not correct.

$$\underline{\underline{\beta = \beta_0 + N_k}}$$

$$\frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W^{-1} - W_0^{-1} + \beta\mu_k\mu_k^T - 2\beta\mu_k m^T + \beta m m^T$$
$$- \beta_0\mu_k\mu_k^T + 2\beta_0\mu_k m_0^T + \beta_0 m_0 m_0^T$$
$$- \sum_{n=1}^{N}\mathbb{E}[z_{nk}]\,x_n x_n^T + 2\sum_{n=1}^{N}\mathbb{E}[z_{nk}]\,x_n\mu_k^T - N_k\mu_k\mu_k^T\Big\}\Big]$$

$$= \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W^{-1} - W_0^{-1} + \cancel{N_k\mu_k\mu_k^T} - 2\beta_0\mu_k(m - m_0)^T + \beta_0(m m^T - m_0 m_0^T) - N_k S_k - N_k\bar{x}_n\bar{x}_n^T\Big\}\Big]$$
$$- 2N_k\mu_k m^T + N_k m m^T \qquad + 2N_k\bar{x}_n\mu_k^T$$
$$- \cancel{N_k\mu_k\mu_k^T}$$

$$= \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{W^{-1} - W_0^{-1} + 2\mu_k\big[(-\beta_0 - N_k)m^T + \beta_0 m_0^T + N_k\bar{x}_n^T\big] + (\beta_0 + N_k)m m^T - \beta_0 m_0 m_0^T$$
$$- N_k S_k - N_k\bar{x}_n\bar{x}_n^T$$

$$\downarrow$$

$$\underline{\underline{m = \frac{\beta_0 m_0^T + N_k\bar{x}_n^T}{\beta_0 + N_k}}}$$

$$= \frac{1}{2}\mathrm{Tr}\Big[\Lambda_k\Big\{\underbrace{W^{-1} - W_0^{-1} + \beta\, m m^T - \beta_0 m_0 m_0^T - N_k S_k - N_k\bar{x}_n\bar{x}_n^T}\Big\}\Big]$$

$$\underline{\underline{W^{-1} = W_0^{-1} + N_k S_k + N_k\bar{x}_n\bar{x}_n^T - \beta\, m m^T + \beta_0 m_0 m_0^T.}}$$

Finally, we got

$$\begin{cases} \beta = \beta_0 + N_k. \\ m = \dfrac{1}{\beta}\cdot(\beta_0 m_0 + N_k\bar{x}_n) \\ W^{-1} = W_0^{-1} + N_k S_k + N_k\bar{x}_n\bar{x}_n^T - \beta m m^T + \beta_0 m_0 m_0^T. \\ \nu_0 = \nu. \end{cases}$$

NOT CORRECT. where is my mistake ?

$\boxed{10.19}$ (10.80) $p(\hat{x}|X) \simeq \sum_{k=1}^{K} \iiint d\pi\, d\mu\, d\Lambda_k\ \pi_k \mathcal{N}(\hat{x}|\mu_k, \Lambda_k^{-1})\, g(\pi)\, g(\mu_k, \Lambda_k)$ where $g(\pi) = \mathrm{Dir}(\pi|\alpha)$, $g(\mu,\Lambda) = \mathcal{N}(\mu|m, (\beta\Lambda)^{-1})\, \mathcal{W}(\Lambda|W,\nu)$

$$p(\hat{x}|X) = \sum_{k=1}^{K} \iiint d\pi\, d\mu_k\, d\Lambda_k\ \pi_k \left(C(\alpha) \prod_{k'=1}^{K} \pi_{k'}^{\alpha_{k'}-1}\right) \cdot \mathcal{N}(\hat{x}|\mu_k, \Lambda_k^{-1})\, \mathcal{N}(\mu|m, (\beta\Lambda)^{-1}) \cdot \mathcal{W}(\Lambda_k|W_k, \nu_k)$$

$$\frac{\alpha_k}{\sum_k \alpha_k}$$

$$\frac{|\Lambda|}{\sqrt{2\pi}} \cdot \frac{|\beta\Lambda|}{\sqrt{2\pi}} \cdot e^{-(\hat{x}-\mu)^T \Lambda_k (\hat{x}-\mu)}\, e^{-(m-\mu)^T \beta\Lambda_k (m-\mu)}$$

$$= \ '' \ \exp\begin{cases} -\hat{x}^T\Lambda_k\hat{x} & -\beta m^T \Lambda_k m \\ +2\hat{x}^T\Lambda_k\mu & +2\beta m^T\Lambda_k \mu \\ -\hat{\mu}^T\Lambda_k\hat{\mu} & -\beta\mu^T\Lambda_k\mu \end{cases} \to 2(\hat{x}^T+\beta m^T)\Lambda_k\mu, \quad -(1+\beta)\mu^T\Lambda_k\mu$$

$$= \ '' \ \exp (1+\beta)\left\{-\mu^T\Lambda_k\mu + 2\left(\frac{\beta m^T+\hat{x}^T}{1+\beta}\right)\Lambda_k\mu - \hat{x}^T\frac{\Lambda_k}{1+\beta}\hat{x} - m^T\frac{\beta\Lambda_k}{1+\beta}m\right\}$$

$$= \ '' \ \exp(1+\beta)\left\{-\left(\mu^T - \left(\frac{\beta m^T+\hat{x}^T}{1+\beta}\right)\right)\Lambda_k\left(\mu - \frac{\beta m^T+\hat{x}^T}{1+\beta}\right) + \frac{1}{(1+\beta)^2}(\beta m^T+\hat{x}^T)\Lambda_k(\beta m^T+\hat{x}^T)\right.$$
$$\left. - \frac{1}{1+\beta}\hat{x}^T\Lambda_k\hat{x} - \frac{\beta}{1+\beta}m^T\Lambda_k m\right\}$$

$$\frac{1}{(1+\beta)^2}\left\{\beta^2 m^T\Lambda_k m^T - \beta(1+\beta)m^T\Lambda_k m\right\}$$
$$\frac{1}{(1+\beta)^2}\left\{\hat{x}^T\Lambda_k\hat{x} - (1+\beta)\hat{x}^T\Lambda_k\hat{x}\right\}$$
$$\frac{1}{(1+\beta)^2}\left\{2\beta m^T\Lambda_k\hat{x}\right\}$$

$$= \frac{|\Lambda|}{\sqrt{2\pi}}\frac{|\beta\Lambda|}{\sqrt{2\pi}}\left(\begin{array}{l} \exp(1+\beta)\left\{-\left(\mu^T - \frac{\beta m^T+\hat{x}^T}{1+\beta}\right)\Lambda_k\left(\mu - \frac{\beta m^T+\hat{x}^T}{1+\beta}\right)\right\} \\ \exp(1+\beta)\left\{-\frac{\beta}{(1+\beta)^2}m^T\Lambda_k m - \frac{\beta}{(1+\beta)^2}\hat{x}^T\Lambda_k\hat{x}^T + \frac{2\beta}{(1+\beta)^2}m^T\Lambda_k\hat{x}\right\}\end{array}\right)$$

$$\exp\left(\frac{\beta}{1+\beta}\right)\left\{-(m^T-\hat{x}^T)\Lambda_k(m^T-\hat{x})\right\}$$

うまくいってない。

Doing μ_k integration,

$$p(\hat{x}|X) = \sum_{k=1}^{K}\frac{\alpha_k}{\hat{\alpha}}\int d\Lambda_k \cdot \frac{|\Lambda|}{\sqrt{2\pi}}\frac{|\beta\Lambda|}{\sqrt{2\pi}}\frac{\sqrt{2\pi}}{|(1+\beta)\Lambda|}\, e^{\frac{\beta}{1+\beta}\{-(m^T-\hat{x}^T)\Lambda_k(m-\hat{x})\}} \cdot \mathcal{W}(\Lambda_k|W_k,\nu_k)$$

$$B(W_k,\nu_k)\,|\Lambda_k|^{\frac{\nu-D-1}{2}} \cdot \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(W_k^{-1} + \frac{2\beta}{1+\beta}(m-\hat{x})(m-\hat{x})^T\right)\Lambda_k\right\}$$

Doing Λ_k integration

$$p(\hat{x}|X) = \sum_{k=1}^{K}\frac{\alpha_k}{\hat{\alpha}}\frac{B(W_k,\nu_k)}{B(\tilde{W}_k,\nu_k)} = \sum_{k=1}^{K}\frac{\alpha_k}{\hat{\alpha}}|W_k|^{\frac{\nu}{2}} \cdot |\tilde{W}_k|^{\frac{\nu}{2}}$$

$(\hat{x}-m_k)^T\Lambda_k$

(2.161) $\mathrm{St}(x|\mu,\Lambda,\nu) = \int_0^\infty d\eta\ \mathcal{N}(x|\mu,(\eta\Lambda)^{-1})\ \mathrm{Gamma}\left(\eta\middle|\frac{\nu}{2},\frac{\nu}{2}\right)$

m_k

$\frac{(\nu_k+1-D)\beta_k}{1+\beta_k}W_k$

ν_k+1-D

$\mathrm{Gam}(\tau|a,b) = \frac{1}{\Gamma(a)}b^a\tau^{a-1}e^{-b\tau}$

$\mathcal{W}(\Lambda,W,\nu) = B\Lambda^{\frac{\nu-D-1}{2}}e^{-\frac{1}{2}\mathrm{Tr}[W^{-1}\Lambda]}$

$|\nu(\beta\Lambda : \Lambda^{-1}$

$\frac{1}{\beta\Lambda}\left(\frac{1}{\beta\Lambda}+\frac{1}{\beta\Lambda}\right)^{-1}\frac{1}{\Lambda} = \frac{\beta}{1+\beta}\Lambda\frac{1}{\beta\Lambda}\frac{1}{\Lambda} = \frac{1}{1+\beta}$

$\boxed{10.19}$ (10.80) $p(\hat{x}|X) \simeq \sum_{k=1}^{K} \iiint d\pi\, d\mu\, d\Lambda_k\; \pi_k \mathcal{N}(\hat{x}|\mu_k, \Lambda_k^{-1})\, g(\pi)\, g(\mu_k, \Lambda_k)$ where $g(\pi) = \mathrm{Dir}(\pi|\alpha)$, $g(\mu, \Lambda) = \mathcal{N}(\mu|m, (\beta\Lambda)^{-1}) \mathcal{W}(\Lambda|W, \nu)$

$$p(\hat{x}|X) = \sum_{k=1}^{K} \iiint \underbrace{d\pi}\, d\mu_k\, d\Lambda_k\; \underbrace{\pi_k \left(C(\alpha) \prod_{k'=1}^{K} \pi_{k'}^{\alpha_0 - 1}\right)}_{\frac{\alpha_k}{\sum_{k=1}^{K} \alpha_k}} \cdot \mathcal{N}(\hat{x}|\mu_k, \Lambda_k^{-1})\, \mathcal{N}(\mu|m, (\beta\Lambda)^{-1}) \cdot \mathcal{W}(\Lambda_k|W_k, \nu_k)$$

$$= \sum_{k=1}^{K} \frac{\alpha_k}{\hat{\alpha}} \iint d\mu_k\, d\Lambda_k\; \underline{\mathcal{N}(\hat{x}|\mu_k, \Lambda_k^{-1})\, \mathcal{N}(\mu_k|m_k, (\beta_k\Lambda_k)^{-1})}\, \mathcal{W}(\Lambda_k|W_k, \nu_k)$$

$\mathcal{N}(x|m_1, \Sigma_1)\, \mathcal{N}(x|m_2, \Sigma_2) = C\mathcal{N}(x|m_c, \Sigma_c)$

where $C = \mathcal{N}(m_1|m_2, \Sigma_1 + \Sigma_2)$

$m_c = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1} m_1 + \Sigma_2^{-1} m_2)$

$\Sigma_c = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$.

$$= \mathcal{N}(\hat{x}|m_k, \Lambda_k^{-1} + (\beta_k\Lambda_k)^{-1})\, \mathcal{N}\left(\mu_k \middle| (\Lambda_k + \beta_k\Lambda_k)^{-1}(\Lambda_k\hat{x} + \beta\Lambda_k m), (\Lambda_k + \beta_k\Lambda_k)^{-1}\right)$$

Doing μ_k integral,

$$p(\hat{x}|X) = \sum_{k=1}^{K} \frac{\alpha_k}{\hat{\alpha}} \underline{\int d\Lambda_k.\; \mathcal{N}(\hat{x}|m_k, \Lambda_k^{-1} + (\beta_k\Lambda_k)^{-1})\, \mathcal{W}(\Lambda_k|W_k, \nu_k)}$$

$\int d\Lambda\; \mathcal{N}(x|m, (\beta\Lambda)^{-1})\, \mathcal{W}(\Lambda|W, \nu)$

$= \mathrm{St}(\hat{x}|m_k, \beta(\nu - D + 1) W)$ NOT PROVED YET.

$$= \sum_{k=1}^{K} \frac{\alpha_k}{\hat{\alpha}} \int d\Lambda_k\; \mathcal{N}\left(\hat{x}\middle|m_k, \left(\frac{\beta_k}{1+\beta_k}\Lambda_k\right)^{-1}\right) \mathcal{W}(\Lambda_k|W_k, \nu_k).$$

$$= \sum_{k=1}^{K} \frac{\alpha_k}{\hat{\alpha}} \mathrm{St}\left(\hat{x}\middle|m_k, \frac{(\nu - D + 1)\beta_k}{1+\beta_k} W_k\right) \quad \blacksquare$$

$\boxed{10.20}$ (10.59) $g^*(\mu_k, \Lambda_k) = \mathcal{N}(\mu_k|m_k, (\beta_k\Lambda_k)^{-1})\, \mathcal{W}(\Lambda_k|W_k, \nu_k)$

Marginalizing over μ_k,

$$g^*(\Lambda_k) = \mathcal{W}(\Lambda_k|W_k, \nu_k)$$

(10.63) $\nu_k = \nu_0 + N_k + 1$ for large N. $\nu_k \to N_k$.

(10.62). $W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^T$ for large N $W_k \to N_k^{-1} S_k^{-1}$

Then, $\mathbb{E}[\Lambda_k] \equiv \nu_k W_k \xrightarrow{\text{large } N} S_k^{-1}$ maximum likelihood value.

Consider the posterior distribution.

entropy. $H[\Lambda] = -\ln B(W,\nu) - \frac{\nu-D-1}{2}\mathbb{E}[\ln|\Lambda|] + \frac{\nu D}{2}$

check this entropy behavior in the limit of $N_k \to \infty$.

since $\begin{pmatrix} W_k \to N_k^{-1} S_k^{-1} \\ \nu_k \to N_k \end{pmatrix}$, $-\ln B(W_k,\nu_k) \to -\frac{N_k}{2}(D\ln N_k + \ln|S_k| - D\ln 2) + \sum_{i=1}^{D} \ln\Gamma\left(\frac{N_k+1-i}{2}\right)$

stirling's approx. , $\ln\Gamma\left(\frac{N_k+1-i}{2}\right) \simeq \frac{N_k}{2}(\ln N_k - \ln 2 - 1)$

Then, $-\ln B(W_k,\nu_k) \to \frac{-N_k D}{2}(\ln N_k - \ln 2 - \ln N_k + \ln 2 + 1) - \frac{N_k}{2}\ln|S_k| = -\frac{N_k}{2}(\ln|S_k| + D)$ ①

$\mathbb{E}[\ln|\Lambda|] \to D\ln\frac{N_k}{2} + D\ln 2 - D\ln N_k - \ln|S_k| = -\ln|S_k|$ ②

(Here we've used (10.241) $\psi(x) = \ln x + O\left(\frac{1}{x}\right)$)

Using ①, ②,

$H[\Lambda] \to 0$ when $N_k \to \infty$

Next, consider $q^*(\mu_k|\Lambda_k) = \mathcal{N}(\mu_k|m_k, \beta_k\Lambda_k)$

mean: (10.61) $m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k\bar{x}_k) = \frac{1}{\beta_0 + N_k}(\beta_0 m_0 + N_k\bar{x}_k) \xrightarrow{N_k\to\infty} \bar{x}_k$ maximum likelihood value.

precision : $\beta_k\Lambda_k \xrightarrow{N_k\to\infty} \beta_k\nu_k W_k \longrightarrow N_k S_k^{-1}$ large value for large N_k ∴ sharply peaked around its mean.

Next, (10.57) $q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$

$\mathbb{E}[\pi_k] = \frac{\alpha_k}{\hat{\alpha}} \xrightarrow{N_k\to\infty} \frac{N_k}{N}$. maximum likelihood value.

Using these results, we can express. (NOTE that. distribution is sharply peaked over μ_k, Λ_k

$$p(\hat{x}|D) = \sum_{k=1}^{K}\frac{\alpha_k}{\alpha}\iint d\mu_k\, d\Lambda_k\, \mathcal{N}(\hat{x}|\mu_k, \Lambda_k)\, q(\mu_k,\Lambda_k)$$

$\delta(\mu_k - \bar{x}_k)\cdot\delta(\Lambda_k - W_k)$

$$\simeq \sum_{k=1}^{K}\frac{N_k}{N}\mathcal{N}(\hat{x}|\bar{x}_k, W_k)$$

■.

A mixture of Gaussians.

$\boxed{10.21}$ For example, gaussian mixture model $\prod_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Lambda_k)$

parameter interchanging like $(\mu_1, \Lambda_1) \leftrightarrow (\mu_2, \Lambda_2)$ remains model unchanged.

$K=2$ case. $\{(\mu_1,\Lambda_1), (\mu_2,\Lambda_2)\}$. $\{(\mu_2,\Lambda_2), (\mu_1,\Lambda_1)\}$ give same settings.

$K=3$ case. $\{(1,2,3)\}$ $\{(1,3,2)\}$ $\{(2,1,3)\}$ $\{(2,3,1)\}$ $\{(3,1,2)\}$ $\{(3,2,1)\}$ give same settings.

clearly, K components case, $\underrightarrow{K!}$ patterns give the same settings

---

$\boxed{10.22}$ K components gaussian mixture model. $\rightarrow$ $K!$ equivalent modes.

Suppose an approximate posterior q is localized in the neighbourhood of one of the modes.

Full posterior distribution : mixture of $K!$ such q distribution. (equal mixing coefficients).

Lower bound. (10.70)

$$\mathcal{L} = \sum_{z} \iiint d\pi \, d\mu \, d\Lambda \; q(z,\pi,\mu,\Lambda) \ln\left(\frac{p(X,z,\pi,\mu,\Lambda)}{q(z,\pi,\mu,\Lambda)}\right)$$

$$= \mathbb{E}[\ln p(X|z,\mu,\Lambda)] + \mathbb{E}[\ln p(z|\pi)] + \mathbb{E}[\ln p(\pi)] + \mathbb{E}[\ln p(\mu,\Lambda)] - \mathbb{E}[\ln q(z)] - \mathbb{E}[\ln q(\pi)] - \mathbb{E}[\ln q(\mu,\Lambda)].$$

Especially,

$$-\mathbb{E}[\ln q(\mu,\Lambda)] = -\mathbb{E}\left[\ln \prod_{k=1}^{K} q(\mu_k,\Lambda_k)\right].$$

From the result of $\boxed{10.21}$, we have $K!$ patterns that lead the same result.

Then we have to take account this factor.

$$-\mathbb{E}[\ln q(\mu,\Lambda)] \longmapsto -\mathbb{E}\left[\ln \frac{\prod_{k=1}^{K} q(\mu_k,\Lambda_k)}{K!}\right] = -\mathbb{E}\left[\ln \prod_{k=1}^{K} q(\mu_k,\Lambda_k)\right] + \underrightarrow{\ln K!}$$

Hence. ELBO includes $+\ln K!$ log-factorial term.

$\boxed{10.23}$ Treat the mixing coefficients as parameters.

$$\mathbb{E}[\ln p(Z|\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk} \ln \pi_k.$$

Using the Lagrange multiplier method,

$$L' = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk} \ln \pi_k + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right).$$

$$\rightarrow \frac{\partial L'}{\partial \pi_k} = \sum_{n=1}^{N}\frac{r_{nk}}{\pi_k} + \lambda. = 0 \quad \therefore \lambda = -\frac{N_k}{\pi_k} \quad \cdots (*)$$

Summing over k in both sides $(\lambda\pi_k = -N_k)$

$$\lambda = -N. \quad \cdots (**)$$

Finally, we gather $(*)$ and $(**)$, then.

$$\underline{\underline{\pi_k = \frac{N_k}{N}}}$$

■ This is the same form as (10.83)

$\boxed{10.24}$ Discuss about singularities for MAP solution of the Bayesian model.

Singularities arise when $r_{nk}=1$, $\mu_k = x_k$, $|\Lambda_k| \to \infty$. Is this possible?

$$(10.40) \quad p(\mu, \Lambda) = \prod_{k=1}^{K} \mathcal{N}(\mu_k|m_0, (\beta_0\Lambda)^{-1})\,\mathcal{W}(\Lambda_k|W_0, \nu_0)$$

$$(10.38) \quad p(X|Z,\mu,\Lambda) = \prod_{n=1}^{N}\prod_{k=1}^{K}\mathcal{N}(x_n|\mu_k, \Lambda_k^{-1})^{z_{nk}}.$$

$$p(X, Z, \pi, \mu, \Lambda) \propto p(X|Z,\mu,\Lambda)\,p(\mu,\Lambda) \quad \text{(including } \Lambda \text{ depending terms only).}$$

Then,

$$\mathbb{E}_{q(z)}[\ln p(X|Z,\mu,\Lambda)\,p(\mu,\Lambda)]$$

$$= \mathbb{E}_{q(z)}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}.\ln\mathcal{N}(x_n|\mu_k, \Lambda_k^{-1}) + \sum_{k=1}^{K}\left\{\ln\mathcal{N}(\mu_k|m_0, (\beta_0\Lambda)^{-1}) + \ln\mathcal{W}(\Lambda_k|W_0, \nu_0)\right\}\right]$$

$k=k$ terms

$$\supset \frac{1}{2}\Big\{\sum_{n=1}^{N} r_{nk}\left(\ln|\Lambda_k| - (x_n-\mu_k)^T\Lambda_k(x_n-\mu_k)\right)$$

$$+ \ln|\Lambda_k| - \beta_0(\mu_k - m_0)^T\Lambda_k(\mu_k - m_0) + (\nu_0 - D - 1)\ln|\Lambda_k| - \mathrm{Tr}[W_0^{-1}\Lambda_k]\Big\} + (\text{const. for } \Lambda_k).$$

$$= \frac{1}{2}\Big\{(\nu_0 + N_k - D)\ln|\Lambda_k| - \mathrm{Tr}\Big[\Big\{W_0^{-1} + \beta_0(\mu_k - m_0)(\mu_k - m_0)^T + \underbrace{\sum_{n=1}^{N} r_{nk}(x_n - \mu_k)(x_n-\mu_k)^T}_{= N_k S_k.}\Big\}\Lambda_k\Big]\Big\} + (\text{const.})$$

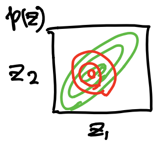
Using $\frac{\partial}{\partial A}\mathrm{Tr}[AB] = B^T$, $\frac{\partial}{\partial A}\ln|A| = (A^{-1})^T$, differentiate the above eq.

$$0 = (\nu_0 + N_k - D)\cdot(\Lambda_k^{-1})^T - \{W_0^{-1} + \beta_0(\mu_k - m_0)(\mu_k - m_0)^T + N_k S_k\}^T.$$

$$\underline{\underline{\therefore \Lambda_k^{-1} = \frac{1}{\nu_0 + N_k - D}\left(W_0^{-1} + \beta_0(\mu_k - m_0)(\mu_k - m_0)^T + N_k S_k\right)}}$$

■. Λ_k^{-1} never becomes 0. because of W_0^{-1} (positive definite)

$\boxed{10.25}$. Fig 10.2



two component case.

under estimate in $\nearrow$ direction. $q(z) = \prod_i \mathcal{N}(z_i|\mu_i, \sigma_i^2)$

(10.48) $q^*(Z) = \prod_{n=1}^{N}\prod_{k=1}^{K} r_{nk}^{z_{nk}}$ where $r_{nk} = \frac{\rho_{nk}}{\sum_{j=1}^{K}\rho_{nj}}$, $\ln\rho_{nk} = \mathbb{E}[\ln\pi_k] + \frac{1}{2}\mathbb{E}[\ln|\Lambda_k|] - \frac{1}{2}\mathbb{E}[(x_n-\mu_k)^T\Lambda_k(x_n-\mu_k)] + C.$

$$r_{nk} \propto \tilde{\pi}_k \tilde{\Lambda}_k^{1/2} \exp\left\{-\frac{D}{2\beta_k} - \frac{\nu_k}{2}(x_n - m_k)^T W_k (x_n - m_k)\right\}$$

$$\left(\begin{array}{l}\beta_k = \beta_0 + N_k \\ W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^T\end{array}\right)$$

$$r_{nk} \propto \exp\left\{-\frac{D}{2\beta_k} - \frac{\nu_k}{2}(x_n - m_k)^T W_k (x_n - m_k)\right\}$$

for the direction of $\max_k\{N_k\}$ $\beta_k \nearrow$, $W_k^{-1} \nearrow$ then $r_{nk} \searrow$ centerize !!

So the variance is smaller than real one. ☑

If. N: fixed, $K \nearrow$ $\beta_k \searrow$, $W_k^{-1} \searrow$, each distribution broaden. $\rightarrow$ smaller contribution

Also $\mathbb{E}[\pi_k] = \frac{\alpha_0 + N_k}{K\alpha_0 + N}$; this goes to smaller when N: fixed $K \nearrow$

Hence contribution of each component gets to be smaller when the # of components get larger.

$\therefore$ Tends to be under-estimated ☑