Essence of the Latent Variable Models

Shin-ichi Mayekawa

950629,960823,011102,030618,050714

1 Models

1.1 Latent Variable Models with Continuous Latent Variables

1. Factor Analysis Model : continuous observed variable
   $\boldsymbol{\Lambda}$=$\{\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}\}$ (grand mean, factor pattern and error variances)
   $\boldsymbol{\rho}$=$\{\boldsymbol{\mu}_x$, $\boldsymbol{\Phi}\}$ (factor mean and factor dispersion)

$$\begin{aligned}
\boldsymbol{y} &\sim N(\boldsymbol{\mu}+\boldsymbol{\Lambda}\boldsymbol{x},\boldsymbol{\Psi}) && (1)\\
f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Lambda}) &= \frac{1}{\sqrt{2\pi|\boldsymbol{\Psi}|}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}+\boldsymbol{\Lambda}\boldsymbol{x})'\,\boldsymbol{\Psi}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}+\boldsymbol{\Lambda}\boldsymbol{x})\right) && (2)\\
\boldsymbol{x} &\sim N(\boldsymbol{\mu}_x,\boldsymbol{\Phi}) \qquad (\boldsymbol{\mu}_x=\boldsymbol{0},\boldsymbol{\Phi}=\boldsymbol{I}) && (3)\\
h(\boldsymbol{x}|\boldsymbol{\rho}) &= \frac{1}{\sqrt{2\pi|\boldsymbol{\Phi}|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)'\,\boldsymbol{\Phi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right). && (4)
\end{aligned}$$

2. IRT Model : binary observed variable
   $\boldsymbol{\Lambda}$=$\{\boldsymbol{\lambda}_j$,j=1,2, $\cdots$,p $\}$ (item parameters), $\boldsymbol{\rho}$=$\{\mu_x,\sigma_x^2$ $\}$ (ability mean and variance)

$$\begin{aligned}
\boldsymbol{y} &\sim \text{independent Bernoulli with prob } = P_j = \text{ICC of item j at } x && (5)\\
f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Lambda}) &= \prod_{j=1}^{p} P_j(x|\boldsymbol{\lambda}_j)^{y_j}(1-P_j(x|\boldsymbol{\lambda}_j))^{1-y_j} && (6)\\
\boldsymbol{x} &\sim N(\mu_x,\sigma_x^2) \qquad (\mu_x=0,\sigma_x^2=1) && (7)\\
h(\boldsymbol{x}|\boldsymbol{\rho}) &= \frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left(-\frac{1}{2}(x-\mu_x)^2/\sigma_x^2\right). && (8)
\end{aligned}$$

1.2 Latent Variable Models with Categorical Latent Variables

1. Classification Model : continuous observed variable
   $\boldsymbol{\Lambda}$=$\{\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_q$,q=1,2, $\cdots$, Q $\}$ (class means and dispersions),
   $\boldsymbol{\rho}$=$\{\rho_q$,q=1,2, $\cdots$, Q $\}$ (prior probabilities)

$$\begin{aligned}
\boldsymbol{y} &\sim N(\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_q), \quad \text{if x = q,} && (9)\\
f(\boldsymbol{y}|\boldsymbol{x}^*,\boldsymbol{\Lambda}) &= \prod_{q=1}^{Q}\left[\frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_q|}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_q)'\,\boldsymbol{\Sigma}_q^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_q)\right)\right]^{x_q^*} && (10)
\end{aligned}$$

2. Latent Class Analysis Model : binary observed variable
   $\boldsymbol{\Lambda}$=$\{\boldsymbol{\lambda}_{jq}, j=1,2,\ \cdots,p, q=1,2,\ \cdots,Q\}$ (class response probabilities) ,
   $\boldsymbol{\rho}$=$\{\rho_q$,q=1,2, $\cdots$, Q $\}$ (prior probabilities)

$$\begin{aligned}
\boldsymbol{y} &\sim \text{independent Bernoulli with prob} = \boldsymbol{\lambda}_{jq}, \quad \text{if x = q,} && (11)\\
f(\boldsymbol{y}|\boldsymbol{x}^*,\boldsymbol{\Lambda}) &= \prod_{q=1}^{Q}\left[\prod_{j=1}^{p}\boldsymbol{\lambda}_{jq}^{y_j}(1-\boldsymbol{\lambda}_{jq})^{1-y_j}\right]^{x_q^*} && (12)
\end{aligned}$$

2 Parameter Estimation using EM Algorithm

2.1 EM Algorithm

EM algorithm can be used to maximize the likelihood of the form

$$
\begin{aligned}
mL &= \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\Lambda},\boldsymbol{\rho}) && (13)\\
&= \prod_{i=1}^{N} \int_{\boldsymbol{X}} f(\boldsymbol{y}_{\mathrm{i}},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho}) d\boldsymbol{x} && (14)\\
&= \prod_{i=1}^{N} \int_{\boldsymbol{X}} f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{x},\boldsymbol{\Lambda})\ h(\boldsymbol{x}|\boldsymbol{\rho}) d\boldsymbol{x} && (15)
\end{aligned}
$$

or its logarithm

$$
\begin{aligned}
\ln mL &= \sum_{i=1}^{N} \ln f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\Lambda},\boldsymbol{\rho}) && (16)\\
&= \sum_{i=1}^{N} \ln \left(\int_{\boldsymbol{X}} f(\boldsymbol{y}_{\mathrm{i}},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho}) d\boldsymbol{x} \right) && (17)
\end{aligned}
$$

Outline

1. Find the initial value of the parameters $\boldsymbol{\Lambda}$ and $\boldsymbol{\rho}$
   $\boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0$
2. Calculate the log complete data likelihood
   $\ln L(\boldsymbol{\Lambda},\boldsymbol{\rho})$
3. Calculate the conditional distribution (posterior distribution) of $\boldsymbol{x}$ given $\boldsymbol{y}$.
   $h(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Lambda}_0,\boldsymbol{\rho}_0)$
4. Calculate the expectation of the log complete data likelihood with respect to $\boldsymbol{x}$ given $\boldsymbol{y}$ and the current value of the parameters
   $\mathcal{E} \ln L(\boldsymbol{\Lambda},\boldsymbol{\rho})$
5. Maximize $\mathcal{E} \ln L(\boldsymbol{\Lambda},\boldsymbol{\rho})$ with respect to the parameters
6. Evaluate $\mathcal{E} \ln L(\boldsymbol{\Lambda},\boldsymbol{\rho})$ with the updated parameters
7. If the increase of $\mathcal{E} \ln L(\boldsymbol{\Lambda},\boldsymbol{\rho})$ is negligible, stop iteration,
   else set $\boldsymbol{\Lambda}_0 = \boldsymbol{\Lambda}, \boldsymbol{\rho}_0 = \boldsymbol{\rho}$ and go back to step 2.

2.2 Continuous Latent Variables

Notation:

$\boldsymbol{y} = p \times 1$ observed variable
$\boldsymbol{x} = r \times 1$ continuous latent variable
$\boldsymbol{y}_{\mathrm{i}}$, i=1 $\cdots N$, observed data
$\boldsymbol{x}_{\mathrm{i}}$, i=1 $\cdots N$, missing data
$\boldsymbol{\Lambda}$, $\boldsymbol{\rho}$= parameters to be estimated

1. Model

$$
f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Lambda}) = \prod_{j=1}^{p} f_j(y_j|\boldsymbol{x},\boldsymbol{\Lambda}) \qquad \text{if locally independent.} \qquad (18)
$$

2. Prior

$$h(\boldsymbol{x}|\boldsymbol{\rho}) \tag{19}$$

3. Joint

$$f(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho}) = f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Lambda})h(\boldsymbol{x}|\boldsymbol{\rho}) \tag{20}$$

4. Marginal

$$f(\boldsymbol{y}|\boldsymbol{\Lambda},\boldsymbol{\rho}) = \int_{\boldsymbol{X}} f(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho})d\boldsymbol{x} \tag{21}$$

5. Posterior

$$\begin{aligned} h(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Lambda},\boldsymbol{\rho}) &= \frac{f(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho})}{f(\boldsymbol{y}|\boldsymbol{\Lambda},\boldsymbol{\rho})} & (22)\\ &\propto f(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho}) = f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Lambda})h(\boldsymbol{x}|\boldsymbol{\rho}) & (23)\end{aligned}$$

6. Marginal Likelihood

This is the likelihood on the basis of the marginal density of $\boldsymbol{y}$.

$$\begin{aligned} mL(\boldsymbol{\Lambda},\boldsymbol{\rho}) &= \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathtt{i}}|\boldsymbol{\Lambda},\boldsymbol{\rho}) & (24)\\ &= \prod_{i=1}^{N} \int_{\boldsymbol{X}} f(\boldsymbol{y}_{\mathtt{i}},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho})d\boldsymbol{x} & (25)\end{aligned}$$

7. Complete Data Likelihood

This is the likelihood on the basis of the joint density of $\boldsymbol{y}$ and $\boldsymbol{x}$.

$$\begin{aligned} L(\boldsymbol{\Lambda},\boldsymbol{\rho}) &= \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathtt{i}},\boldsymbol{x}_{\mathtt{i}}|\boldsymbol{\Lambda},\boldsymbol{\rho}) & (26)\\ &= \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathtt{i}}|\boldsymbol{x}_{\mathtt{i}},\boldsymbol{\Lambda})h(\boldsymbol{x}_{\mathtt{i}}|\boldsymbol{\rho}) & (27)\end{aligned}$$

8. Log Complete Data Likelihood

$$\begin{aligned} \ln L(\boldsymbol{\Lambda},\boldsymbol{\rho}) &= \sum_{i=1}^{N} \ln f(\boldsymbol{y}_{\mathtt{i}},\boldsymbol{x}_{\mathtt{i}}|\boldsymbol{\Lambda},\boldsymbol{\rho}) & (28)\\ &= \sum_{i=1}^{N} \ln f(\boldsymbol{y}_{\mathtt{i}}|\boldsymbol{x}_{\mathtt{i}},\boldsymbol{\Lambda}) + \sum_{i=1}^{N} \ln h(\boldsymbol{x}_{\mathtt{i}}|\boldsymbol{\rho}) & (29)\end{aligned}$$

9. Expected Log Complete Data Likelihood

This is the expectation of $\ln L$ with respect to the posterior distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$. The posterior distribution is treated as known with $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\rho}_0$ being the provisional values of the parameters.

$$\begin{aligned} \mathcal{E}\ln L(\boldsymbol{\Lambda},\boldsymbol{\rho}) &= \mathcal{E}(\ln L(\boldsymbol{\Lambda},\boldsymbol{\rho}) \mid \boldsymbol{Y}) & (30)\\ &= \sum_{i=1}^{N} \mathcal{E}(\ln f(\boldsymbol{y}_{\mathtt{i}},\boldsymbol{x}|\boldsymbol{\Lambda},\boldsymbol{\rho}) \mid \boldsymbol{y}_{\mathtt{i}}) & (31)\\ &= \sum_{i=1}^{N} \mathcal{E}(\ln f(\boldsymbol{y}_{\mathtt{i}}|\boldsymbol{x},\boldsymbol{\Lambda}) \mid \boldsymbol{y}_{\mathtt{i}}) + \sum_{i=1}^{N} \mathcal{E}(\ln h(\boldsymbol{x}|\boldsymbol{\rho}) \mid \boldsymbol{y}_{\mathtt{i}}) & (32)\end{aligned}$$

$$= \sum_{i=1}^{N} \int_{\boldsymbol{X}} \ln f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{x}, \boldsymbol{\Lambda}) h(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) d\boldsymbol{x}$$
$$+ \sum_{i=1}^{N} \int_{\boldsymbol{X}} \ln h(\boldsymbol{x}|\boldsymbol{\rho}) h(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) d\boldsymbol{x} \tag{33}$$

In the M-step, this is maximized with respect to $\boldsymbol{\Lambda}$ and $\boldsymbol{\rho}$. $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\rho}_0$ are treated as known.

10. Numerical Integration

The general numerical integration formula is:

$$I = \int_{\boldsymbol{X}} G(x)\ H(x) dx \quad \approx \quad \sum_{q=1}^{Q} G(x_q)\ w(x_q) \tag{34}$$

where x_q and $w(x_q)$ are, respectively, the quadrature points and the associated weights which depend on the form of H and the range of integration.

- Gauss-Legender Quadrature

$$G(x) := \ln f(\boldsymbol{y}_{\mathrm{i}}|x, \boldsymbol{\Lambda}) h(x|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0), \quad H(x) := 1 \tag{35}$$

- Gauss-Hermite Quadrature

$$G(x) := \ln f(\boldsymbol{y}_{\mathrm{i}}|x, \boldsymbol{\Lambda}) \frac{h(x|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0)}{\text{density of } N(\mu, \sigma^2)}, \quad H(x) := \text{density of } N(\mu, \sigma^2) \tag{36}$$

11. Fixed Point Apprximation

When

$$\int_{\boldsymbol{X}} H(\boldsymbol{x}) d\boldsymbol{x} = 1, \tag{37}$$

we may use the following approximation:

$$I = \int_{\boldsymbol{X}} G(\boldsymbol{x})\ H(\boldsymbol{x}) d\boldsymbol{x} \quad \approx \quad \sum_{q=1}^{Q} G(\boldsymbol{x}_q)\ H^*(\boldsymbol{x}_q) \tag{38}$$

where

$$H^*(\boldsymbol{x}_q) = \frac{H(\boldsymbol{x}_q)}{\sum_{q=1}^{Q} H(\boldsymbol{x}_q)}. \tag{39}$$

Therefore,

$$\mathcal{E} \ln L(\boldsymbol{\Lambda}, \boldsymbol{\rho}) \quad \approx \quad \sum_{i=1}^{N} \sum_{q=1}^{Q} \ln f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{x}_q, \boldsymbol{\Lambda}) h^*(\boldsymbol{x}_q|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) + \sum_{i=1}^{N} \sum_{q=1}^{Q} \ln h(\boldsymbol{x}_q|\boldsymbol{\rho}) h^*(\boldsymbol{x}_q|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \tag{40}$$

$$= \sum_{q=1}^{Q} \sum_{i=1}^{N} h^*(\boldsymbol{x}_q|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \ln f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{x}_q, \boldsymbol{\Lambda}) + \sum_{q=1}^{Q} \left(\sum_{i=1}^{N} h^*(\boldsymbol{x}_q|\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \right) \ln h(\boldsymbol{x}_q|\boldsymbol{\rho}) \tag{41}$$

2.3 Categorical Latent Variables

Notation:

$\boldsymbol{y} = p \times 1$ observed variable
$\boldsymbol{x} = Q \times 1$ discrete latent variable
$\boldsymbol{x}^* = q \times 1$ dummy expanded $\boldsymbol{x}$ variable (indicator)
$\boldsymbol{y}_{\mathrm{i}}$, i=1 $\cdots N$, observed data
$\boldsymbol{x}_{\mathrm{i}}$, i=1 $\cdots N$, missing data
$\boldsymbol{\Lambda}$, $\boldsymbol{\rho}$= parameters to be estimated

1. Model

$$f(\boldsymbol{y}|\boldsymbol{x}^*, \boldsymbol{\Lambda}) = \prod_{q=1}^{Q} (f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q))^{x_q^*} \tag{42}$$

$$= \prod_{q=1}^{Q} \left(\prod_{j=1}^{p} f_{jq}(y_j|\boldsymbol{\lambda}_q) \right)^{x_q^*} \tag{43}$$

if locally independent. (44)

The subscript q for f is not necessary if all the class shares the same form of pdf with different values of parameters.

2. Prior

$$h(\boldsymbol{x}^*|\boldsymbol{\rho}) = \prod_{q=1}^{Q} {\rho_q}^{x_q^*}, \quad \text{where} \quad \sum_{q=1}^{Q} \rho_q = 1 \tag{45}$$

3. Joint

$$f(\boldsymbol{y}, \boldsymbol{x}^*|\boldsymbol{\Lambda}, \boldsymbol{\rho}) = f(\boldsymbol{y}|\boldsymbol{x}^*, \boldsymbol{\Lambda}) h(\boldsymbol{x}^*|\boldsymbol{\rho}) \tag{46}$$

$$= \prod_{q=1}^{Q} (f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\ \rho_q)^{x_q^*} \tag{47}$$

4. Marginal

$$f(\boldsymbol{y}|\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \sum_{\boldsymbol{x}^*} f(\boldsymbol{y}, \boldsymbol{x}^*|\boldsymbol{\Lambda}, \boldsymbol{\rho}) \tag{48}$$

$$= \sum_{\boldsymbol{x}^*} \prod_{q=1}^{Q} (f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q)^{x_q^*} \tag{49}$$

$$= \sum_{q=1}^{Q} f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q \tag{50}$$

5. Posterior

$$h(\boldsymbol{x}^*|\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{\rho}) = \frac{f(\boldsymbol{y}, \boldsymbol{x}^*|\boldsymbol{\Lambda}, \boldsymbol{\rho})}{f(\boldsymbol{y}|\boldsymbol{\Lambda})} \tag{51}$$

$$= \frac{\prod_{q=1}^{Q} (f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q)^{x_q^*}}{\sum_{q=1}^{Q} f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q} = \prod_{q=1}^{Q} \left(\frac{f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q}{\sum_{q=1}^{Q} f_q(\boldsymbol{y}|\boldsymbol{\lambda}_q)\rho_q} \right)^{x_q^*} \tag{52}$$

$$= \prod_{q=1}^{Q} h_q(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{\rho})^{x_q^*} \tag{53}$$

6. Marginal Likelihood

This is the likelihood on the basis of the marginal density of $\boldsymbol{y}$.

$$mL(\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\Lambda}, \boldsymbol{\rho}) \tag{54}$$

$$= \prod_{i=1}^{N} \left(\sum_{q=1}^{Q} f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q)\ \rho_q \right) \tag{55}$$

7. Complete Data Likelihood

This is the likelihood on the basis of the joint density of $\boldsymbol{y}$ and $\boldsymbol{x}$.

$$L(\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \prod_{i=1}^{N} f(\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{x}_i^*|\boldsymbol{\Lambda}, \boldsymbol{\rho}) \tag{56}$$

$$= \prod_{i=1}^{N} \prod_{q=1}^{Q} (f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q)\ \rho_q)^{x_{iq}^*} \tag{57}$$

8. Log Complete Data Likelihood

$$\ln L(\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \sum_{i=1}^{N} \ln f(\boldsymbol{y}_{\mathrm{i}}, \boldsymbol{x}_i^*|\boldsymbol{\Lambda}, \boldsymbol{\rho}) \tag{58}$$

$$= \sum_{i=1}^{N} \sum_{q=1}^{Q} x_{iq}^* \ln f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q) + \sum_{i=1}^{N} \sum_{q=1}^{Q} x_{iq}^* \ln \rho_q \tag{59}$$

9. Expected Log Complete Data Likelihood

This is the expectation of $\ln L$ with respect to the posterior distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$. The posterior distribution is treated as known with $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\rho}_0$ being the provisional values of the parameters. Note that the posterior expectation of x_{iq}^* is $h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0)$.

$$\mathcal{E} \ln L(\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \mathcal{E}(\ln L \mid \boldsymbol{Y}) \tag{60}$$

$$= \mathcal{E} \left(\sum_{i=1}^{N} \sum_{q=1}^{Q} x_{iq}^* \ln f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q) + \sum_{i=1}^{N} \sum_{q=1}^{Q} x_{iq}^* \ln \rho_q \mid \boldsymbol{Y} \right) \tag{61}$$

$$= \sum_{i=1}^{N} \sum_{q=1}^{Q} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \ln f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q) + \sum_{i=1}^{N} \sum_{q=1}^{Q} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \ln \rho_q \tag{62}$$

$$= \sum_{q=1}^{Q} \sum_{i=1}^{N} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \ln f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q) + \sum_{q=1}^{Q} (\sum_{i=1}^{N} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0)) \ln \rho_q \tag{63}$$

In the M-step, this is maximized with respect to $\boldsymbol{\Lambda}$ and $\boldsymbol{\rho}$. $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\rho}_0$ are treated as known.

$$\text{the first term of } \mathcal{E} \ln L(\boldsymbol{\Lambda}, \boldsymbol{\rho}) = \sum_{q=1}^{Q} \sum_{i=1}^{N} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \ln f_q(\boldsymbol{y}_{\mathrm{i}}|\boldsymbol{\lambda}_q) \tag{64}$$

$$= \sum_{q=1}^{Q} \sum_{i=1}^{N} h_q(\boldsymbol{y}_i, \boldsymbol{\Lambda}_0, \boldsymbol{\rho}_0) \sum_{j=1}^{p} \ln f_{jq}(\boldsymbol{y}_{ij}|\boldsymbol{\lambda}_q) \tag{65}$$