

ECM Algorithm for Fitting of Mixtures of Multivariate Skew t -Distribution

Seung-Gu Kim^{1,a}

^aDepartment of Data and Information, Sangji University

Abstract

Cabral *et al.* (2012) defined a mixture model of multivariate skew t -distributions(STMM), and proposed the use of an ECME algorithm (a variation of a standard EM algorithm) to fit the model. Their estimation by the ECME algorithm is closely related to the estimation of the degree of freedoms in the STMM. With the ECME, their purpose is to escape from the calculation of a conditional expectation that is not provided by a closed form; however, their estimates are quite unstable during the procedure of the ECME algorithm. In this paper, we provide a conditional expectation as a closed form so that it can be easily calculated; in addition, we propose to use the ECM algorithm in order to stably fit the STMM.

Keywords: Multivariate skew t -distribution, mixture model, ECME algorithm, ECM algorithm, estimation of degree of freedom.

1. 서론

정규혼합모형(normal mixture model; NMM)이나 t -분포 혼합모형(TMM)은 근 20년 동안 많은 응용분야에서 큰 역할을 하여 왔으며, 특히 군집분석에서는 거의 표준적 도구로서 사용되고 있다. 그러나 정규분포나 t -분포의 대칭성 때문에 자료 공간에서 비대칭성을 보이는 자료에 대해서는 적합의 문제성을 드러냈다. 이를 극복하기 위해 최근 다변량 상황에서 치우친 정규분포 혼합모형(skew normal mixture model; SNMM)과 더 나아가 치우친 t -분포 혼합모형(skew t -distribution mixture model; STMM)이 개발되었다. 이에 대한 배경을 간략히 정리하면 다음과 같다.

Lin 등 (2007a)은 단변량 상황에서 Azzalini (1985)의 치우친 정규분포를 바탕으로 SNMM을 제안 하였으며, Azzalini와 Dalla-Valle (1996)은 다변량 SNMM을 제공하였다. 한편 t -분포와 관련해서는 Lin 등 (2007b)은 Azzalini과 Capitanio (2003)의 치우친 단변량 t -분포를 바탕으로 STMM을 개발하였다.

치우친 다변량 t -분포는 크게 두 가지 정의가 있는데, 하나는 Sahu 등 (2003)의 것이고 나머지 하나는 Azzalini와 Dalla-Valle (1996)의 것이다. 두 분포를 바탕으로 한 혼합모형을 각각 STMM-Sahu와 STMM-AzDv라 부르겠다. STMM-Sahu에 대해서는 Lin (2010)이 제안하면서 EM 구조 하에서 몬테카를로(EM-MC)에 의한 추정법을 제공하였고, Lee와 McLachlan (2011)는 ECM(expectation-conditional maximization) 알고리즘에 의한 추정법을 소개하였다. 한편 STMM-AzDv에 대해서는 Cabral 등 (2012)이 제안하면서 ECM 알고리즘 구현의 한 가지 방편인 ECME(expectation-conditional maximization either) 알고리즘에 의한 추정법을 제공하였다. 두 STMM 모형의 자료적합 능력 비교에

This research was supported by the Sangji research fund in 2011.

¹ Professor, Department of Data and Information, Sangji University, 83 Usan-Dong, Kang-Won do, Wonju 122-807, Korea.
E-mail: sgukim@sanji.ac.kr

대해 논란의 여지가 있지만, 저자의 경험으로는 후자의 방법이 전자의 그것보다 나쁘다고는 보여지지 않으며, 무엇보다도 처리시간이 상대적으로 매우 빠르다는 장점이 있다. 두 방법의 적합능력의 비교에 대해서는 4절에서 언급할 것이다.

본 논문에서는 후자의 방법 즉 Cabral 등의 STMM-AzDv를 다루는데, 그들의 방법으로 모수를 추정할 때, 특별히 자유도 추정할 때 문제점을 지적한다. 이에 대한 자세한 언급은 다음 절에서 그들의 모형과 추정 방법을 요약하면서 소개할 것이다. 3절에서는 ECM 알고리즘의 구현을 위한 조건부 기대값을 명시적 형태로 구하며, 4절에서는 모의자료와 실자료를 이용하여 제안된 ECM 알고리즘의 유효성을 살필 것이다. 5절에서는 결론과 토의의 정리한다.

2. 연구배경 및 문제제기

2.1. STM 모형

Cabral 등 (2012)은 FM-SNI(finite mixture of skew normal independent)라는 系의 하나의 구성원으로서 STMM을 소개하고 있다. 그들에 따르면 p -변량 관측치 $\mathbf{y}_j = (y_{1j}, \dots, y_{pj})^T$ 에 대한 g 개의 성분을 가지는 STMM은

$$f(\mathbf{y}_j; \Theta) = \sum_{i=1}^g \pi_i f_i(\mathbf{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\delta}_i, \nu_i), \quad j = 1, \dots, n \quad (2.1)$$

와 같으며, 여기서 $\pi_i (i = 1, \dots, g)$ 는 $\pi_1 + \dots + \pi_g = 1$ 을 만족하는 i 번째 성분비율이고, i 번째 성분밀도로서 다변량 치우친 t -밀도를

$$f_i(\mathbf{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\delta}_i, \nu_i) = 2t_p(\mathbf{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i) T\left(r_{ij} \sqrt{\frac{\nu_i + p}{\nu_i + D_{ij}}}; \nu_i + p\right) \quad (2.2)$$

와 같이 정의한다. 이때 $t_p(\cdot; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i)$ 는 위치모수와 척도모수가 $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ 그리고 자유도가 ν_i 인 p -변량 t -밀도이며, $T(\cdot; df)$ 는 자유도가 df 인 표준 단변량 t -분포함수를 나타낸다. 그리고 $D_{ij} = D(\mathbf{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = (\mathbf{y}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_i)$ 으로서 Mahalanobis 거리제곱을 나타내며, $r_{ij} = r(\mathbf{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\delta}_i) = \boldsymbol{\delta}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_i)$ 를 나타내며, $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{ip})^T$ 은 i 번째 성분밀도의 치우침(skewness) 모수벡터를 나타낸다. 그리고 Θ 는 모든 (중복되지 않는) 모수를 포함하는 벡터를 나타낸다.

식 (2.1)의 STMM은 만약 $\nu_i (i = 1, \dots, g) \rightarrow \infty$ 이면 SNMM이 되며, $\boldsymbol{\delta}_i (i = 1, \dots, g) \rightarrow \mathbf{0}$ 이면 TMM이 되고, 두 조건이 동시에 만족하면 NMM으로 축소되는 매우 신축적인 모형이다.

2.2. EM 알고리즘에 의한 추정

우도함수를 알기 쉽게 나타내기 위해 식 (2.1)에서 $\boldsymbol{\delta}_i^* = \boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\delta}_i / \sqrt{1 + \boldsymbol{\delta}_i^T \boldsymbol{\delta}_i}$ 및 $\boldsymbol{\Omega}_i = \boldsymbol{\Sigma}_i - \boldsymbol{\delta}_i^* \boldsymbol{\delta}_i^{*T}$ 와 같은 재모수화가 필요한데, 이것은 언제든지

$$\boldsymbol{\delta}_i = \frac{(\boldsymbol{\Omega}_i + \boldsymbol{\delta}_i^* \boldsymbol{\delta}_i^{*T})^{-\frac{1}{2}} \boldsymbol{\delta}_i^*}{\left[1 - \boldsymbol{\delta}_i^{*T} (\boldsymbol{\Omega}_i + \boldsymbol{\delta}_i^* \boldsymbol{\delta}_i^{*T})^{-\frac{1}{2}} \boldsymbol{\delta}_i^*\right]^{\frac{1}{2}}}, \quad \boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \boldsymbol{\delta}_i^* \boldsymbol{\delta}_i^{*T}$$

와 같이 되돌릴 수 있다. 이제 벡터 $\mathbf{Z}_j = (Z_{1j}, \dots, Z_{gj})^T$ 의 원소 Z_{ij} 를 관측치 \mathbf{y}_j 가 i 번째 성분으로부터 생성되었으면 1, 그렇지 않으면 0을 나타내는 지시변수라 하자. 이때 모형 (2.1)은 다음과 같은 위계적

구조로서 표현되는 것으로 알려져 있다. 즉,

$$\begin{aligned} \mathbf{Z}_j &\sim \text{multi}(1; \pi_1, \dots, \pi_g), \quad U_j | Z_{ij} = 1 \sim \text{gamma}\left(\frac{v_i}{2}, \frac{v_i}{2}\right), \quad T_j | U_j = u_j, Z_{ij} = 1 \sim \text{HN}(0, u_j^{-1}), \\ \mathbf{Y}_j | T_j = t_j, U_j = u_j, Z_{ij} = 1 &\sim N_p\left(\boldsymbol{\mu}_i + \boldsymbol{\delta}_i^* t_j, \frac{\boldsymbol{\Omega}_i}{u_j}\right). \end{aligned} \quad (2.3)$$

여기서 $\text{multi}(n; \pi_1, \dots, \pi_g)$ 은 시행횟수 n 및 성공확률 (π_1, \dots, π_g) 인 g -항 분포를 나타내며, $\text{gamma}(\alpha, \beta)$ 은 평균이 α/β 인 감마분포, $\text{HN}(\mu, \sigma^2)$ 는 평균과 분산이 (μ, σ^2) 인 단변량 절반정규분포 (half normal distribution) 그리고 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 는 평균과 분산이 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 인 p -변량 정규분포를 나타낸다.

식 (2.3)의 조건부 관계로부터 완비자료 $\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)^T$, $\mathbf{u} = (u_1, \dots, u_n)^T$, $\mathbf{t} = (t_1, \dots, t_n)^T$ 및 $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$ 의 로그-우도는

$$L_c(\boldsymbol{\Theta} | \mathbf{Z}, \mathbf{u}, \mathbf{t}, \mathbf{y}) = L_{c1}(\boldsymbol{\theta}) + L_{c2}(\boldsymbol{\nu}) \quad (2.4)$$

와 같이 두 개로 분리된 성분으로 나타낼 수 있는데, 여기서

$$L_{c1}(\boldsymbol{\theta}) \propto \sum_{i=1}^g \sum_{j=1}^n z_{ij} \left\{ \log \pi_i - \frac{1}{2} \log |\boldsymbol{\Omega}_i| - \frac{1}{2} u_j (\mathbf{y}_j - \boldsymbol{\mu}_i - \boldsymbol{\delta}_i^* t_j)^T \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_i - \boldsymbol{\delta}_i^* t_j) \right\}, \quad (2.5)$$

$$L_{c2}(\boldsymbol{\nu}) \propto \sum_{i=1}^g \sum_{j=1}^n z_{ij} \left\{ -\log \Gamma\left(\frac{v_i}{2}\right) + \frac{v_i}{2} \log\left(\frac{v_i}{2}\right) + \frac{v_i}{2} (\log u_j - u_j) \right\} \quad (2.6)$$

이며, $\boldsymbol{\theta}$ 는 $\{\pi_i\}, \{\boldsymbol{\mu}_i\}, \{\boldsymbol{\delta}_i^*\}$ 및 $\{\boldsymbol{\Omega}_i\}$ 를 포함하는 모수 벡터이고 $\boldsymbol{\nu} = (\nu_1, \dots, \nu_g)^T$ 그리고 $\boldsymbol{\Theta} = (\boldsymbol{\theta}^T, \boldsymbol{\nu}^T)^T$ 이다. 자유도 $\boldsymbol{\nu}$ 와 (미관측) 자료 $\log u_j$ 는 오직 $L_{c2}(\boldsymbol{\nu})$ 에만 포함되어 있음을 주목하자.

이제 EM 알고리즘의 $(k+1)$ 번째 단계의 E-스텝에서 완비자료의 조건부 기대값

$$\begin{aligned} Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(k)}) &= E_{\boldsymbol{\Theta}^{(k)}} [L_c(\boldsymbol{\Theta}) | \mathbf{y}] = E_{\boldsymbol{\theta}^{(k)}} [L_{c1}(\boldsymbol{\theta}) | \mathbf{y}] + E_{\boldsymbol{\nu}^{(k)}} [L_{c2}(\boldsymbol{\nu}) | \mathbf{y}] \\ &= Q_1(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)}) + Q_2(\boldsymbol{\nu} | \boldsymbol{\nu}^{(k)}) \end{aligned} \quad (2.7)$$

를 계산해야 한다. 이것은 식 (2.5)를 통해 알 수 있듯이 $Q_1(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)})$ 의 계산을 위해서 몇 가지의 변량에 대한 조건부 기대값을 계산해야 하는 문제인데, 이것들은 이중기대값과 확률분포에 관한 몇 가지 성질을 이용하여 $i = 1, \dots, g; j = 1, \dots, n$ 에 관해

$$\begin{aligned} \tau_{ij}^{(k+1)} &= E_{\boldsymbol{\Theta}^{(k)}} [Z_{ij} | \mathbf{y}_j] = \Pr_{\boldsymbol{\Theta}^{(k)}} \{Z_{ij} = 1 | \mathbf{y}_j\} = \frac{\pi_i^{(k)} f_i(\mathbf{y}_j; \boldsymbol{\mu}_i^{(k)}, \boldsymbol{\Sigma}_i^{(k)}, \boldsymbol{\delta}_i^{(k)}, \nu_i^{(k)})}{\sum_{h=1}^g \pi_h^{(k)} f_h(\mathbf{y}_j; \boldsymbol{\mu}_h^{(k)}, \boldsymbol{\Sigma}_h^{(k)}, \boldsymbol{\delta}_h^{(k)}, \nu_h^{(k)})}, \\ e_{1,ij}^{(k+1)} &= E_{\boldsymbol{\Theta}^{(k)}} [U_j | \mathbf{y}_j, Z_{ij} = 1] = b_{ij}^{(k)}(0) \frac{T\left(r_{ij}^{(k)} \sqrt{b_{ij}^{(k)}(1)}; \nu_i^{(k)} + p + 2\right)}{T\left(r_{ij}^{(k)} \sqrt{b_{ij}^{(k)}(0)}; \nu_i^{(k)} + p\right)}, \\ e_{2,ij}^{(k+1)} &= E_{\boldsymbol{\Theta}^{(k)}} [U_j T_j | \mathbf{y}_j, Z_{ij} = 1] = e_{1,ij}^{(k+1)} m_{ij}^{(k)} + M_i^{(k)} K_{ij}^{(k)}, \\ e_{3,ij}^{(k+1)} &= E_{\boldsymbol{\Theta}^{(k)}} [U_j T_j^2 | \mathbf{y}_j, Z_{ij} = 1] = e_{1,ij}^{(k+1)} \left(m_{ij}^{(k)}\right)^2 + m_{ij}^{(k)} M_i^{(k)} K_{ij}^{(k)} + \left(M_i^{(k)}\right)^2 \end{aligned}$$

과 같이 명시적 형태로 얻게 된다. 여기서 $b_{ij}^{(k)}(s) = (\nu_i^{(k)} + p + 2s)/(\nu_i^{(k)} + D_{ij}^{(k)})$ ($s = 0, 1$)을 나타내며, $D_{ij}^{(k)} = D(\mathbf{y}_j; \boldsymbol{\mu}_i^{(k)}, \boldsymbol{\Sigma}_i^{(k)})$, $M_i^{(k)} = (1 + \boldsymbol{\delta}_i^{(k)T} \boldsymbol{\Omega}_i^{(k)-1} \boldsymbol{\delta}_i^{(k)})^{-1/2}$, $m_{ij}^{(k)} = (M_i^{(k)})^2 \boldsymbol{\delta}_i^{(k)T} \boldsymbol{\Omega}_i^{(k)-1} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k)})$ 를 나타내

고, $r_{ij}^{(k)} = \delta_i^{*(k)T} \Sigma_i^{(k)-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{(k)}) = m_{ij}^{(k)} / M_i^{(k)}$ 이며,

$$K_{ij} = \frac{\Gamma\left(\frac{v_i^{(k)} + p}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{v_i^{(k)} + p}{2}\right)} \frac{(v_i^{(k)} + D_{ij}^{(k)})^{(v_i^{(k)} + p)/2}}{(v_i^{(k)} + D_{ij}^{(k)} + (r_{ij}^{(k)})^2)^{(v_i^{(k)} + p + 1)/2}} T^{-1}\left(r_{ij}^{(k)} \sqrt{b_{ij}^{(k)}(0)}; v_i^{(k)} + p\right)$$

을 나타낸다.

반면, 식 (2.6)–(2.7)을 통해 알 수 있듯이 $Q_2(\boldsymbol{\nu}|\boldsymbol{v}^{(k)})$ 의 계산을 위해

$$e_{4,ij}^{(k+1)} = E_{\Theta^{(k)}}[\log U_j | \mathbf{y}_j, Z_{ij} = 1] \quad (2.8)$$

를 구해야 한다. 그런데 이 조건부 기대값은 STMM-Sahu 모형에 대해서 Lin (2010)에 의해 결코 계산이 용이하지 않은 형태이기는 하지만 이미 알려져 있다. 그러나 STMM-AzDv 모형에서는 저자가 아는 한 아직 알려지지 않았다. 다만 계산하기 쉽지 않은 형태일 것이란 짐작은 할 수 있다.

그래서 Cabral 등 (2012)은 CM-스텝(conditional maximization-step)에서 $\boldsymbol{\theta}$ 에 대하여 $\mathbf{0} = \partial Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) / \partial \boldsymbol{\theta}$ 를 통해

$$\begin{aligned} \pi_i^{(k+1)} &= \frac{\sum_{j=1}^n \tau_{ij}^{(k+1)}}{n}, \quad \boldsymbol{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(k+1)} (e_{1,ij}^{(k+1)} \mathbf{y}_j - e_{2,ij}^{(k+1)} \boldsymbol{\delta}_i^{*(k)})}{\sum_{j=1}^n \tau_{ij}^{(k+1)} e_{1,ij}^{(k+1)}}, \\ \boldsymbol{\delta}_i^{*(k+1)} &= \frac{\sum_{j=1}^n \tau_{ij}^{(k+1)} e_{2,ij}^{(k+1)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k+1)})}{\sum_{j=1}^n \tau_{ij}^{(k+1)} e_{3,ij}^{(k+1)}} \quad \text{및} \\ \boldsymbol{\Omega}_i^{(k+1)} &= \frac{1}{\sum_{j=1}^n \tau_{ij}^{(k+1)}} \sum_{j=1}^n \tau_{ij}^{(k+1)} \left\{ e_{1,ij}^{(k+1)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k+1)}) (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k+1)})^T - e_{2,ij}^{(k+1)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k+1)}) \boldsymbol{\delta}_i^{*(k+1)T} \right. \\ &\quad \left. - e_{2,ij}^{(k+1)} \boldsymbol{\delta}_i^{*(k+1)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(k+1)})^T + e_{3,ij}^{(k+1)} \boldsymbol{\delta}_i^{*(k+1)} \boldsymbol{\delta}_i^{*(k+1)T} \right\}, \quad i = 1, \dots, g \end{aligned}$$

를 얻은 반면 $\boldsymbol{\nu}^{(k+1)}$ 을 얻기 위해 $Q_2(\boldsymbol{\nu}|\boldsymbol{v}^{(k)})$ 를 최대화 하는 대신 불완비자료 \mathbf{y} 에 대한 로그-우도 $L(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\nu}|\mathbf{y})$ 를 직접 최대화하는 방법을 택하였다. 즉,

$$\boldsymbol{\nu}^{(k+1)} = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \left[\sum_{j=1}^n \log \left\{ \sum_{i=1}^g \pi_i^{(k+1)} f_i(\mathbf{y}_j | \boldsymbol{\mu}_i^{(k+1)}, \boldsymbol{\Sigma}_i^{(k+1)}, \boldsymbol{\delta}_i^{(k+1)}, \boldsymbol{\nu}_i) \right\} \right] \quad (2.9)$$

을 구함으로써, 식 (2.8)의 $E_{\Theta^{(k)}}[\log U_j | \mathbf{y}_j, Z_{ij} = 1]$ 를 계산할 필요가 없게 하였다. 이러한 방법을 ECME 알고리즘이라 하는데, Liu와 Rubin (1994)에 의해 제안된 방법으로서 원래는 ECM 알고리즘의 수렴속도를 개선할 목적으로 소개되었다.

2.3. 자유도 추정에 대한 문제

앞서 설명한 바를 정리하면 다음과 같다. $(k+1)$ -스텝에서 ECM 알고리즘으로 자유도 $\boldsymbol{\nu}$ 를 추정하기 위해 $Q_2(\boldsymbol{\nu}|\boldsymbol{v}^{(k)})$ 를 최대화하려면 $\log U_j$ 의 조건부 기대값을 계산해야 한다. Cabral 등 (2012)은 이를 회피하기 위해 식 (2.9)을 통해 불완비자료 \mathbf{y} 에 대한 로그-우도 $L(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\nu}|\mathbf{y})$ 를 최대화하는 ECME 알고리즘을 이용하였다.

그러나 McLachlan과 Peel (2000, 234–235)에서 지적하였듯이 TMM의 경우, $\nu_1 = \dots = \nu_g$ 를 가정하지 않는 한 주어진 자료 \mathbf{y} 에 따라서는 $\boldsymbol{\nu}$ 의 로그-우도 $L(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\nu}|\mathbf{y})$ 의 표면은 $\boldsymbol{\nu}$ 의 최대값 부근의 꽤 넓

은 영역에서 거의 평평한 상태를 보여준다. 따라서 이 평평하고 넓은 영역 내에서 어떠한 값도 $\nu^{(k+1)}$ 가 될 가능성이 높다. Lin (2010) 역시 STMM에서 비슷한 특성을 지적하였다. 확실치는 않으나 Cabral 등 (2012)은 식 (2.9)의 $\nu^{(k+1)}$ 를 얻기 위해 심플렉스(symplex) 탐색기법을 이용한 듯 보인다. 이러한 상황에서 탐색기법은 그릇된 해를 제공할 가능성이 높은 것으로 판단된다. 그래서인지는 몰라도 그들의 모의실험에서는 등자유도 상황에서의 결과만을 보여주고 있다. 사실 많은 응용문제에서 등자유도는 무리한 가정일 수 있다.

EM 알고리즘의 이점 중 하나는 미관측 결측자료를 포함함으로써 일종의 대리함수(surrogate function)인 $Q(\Theta|\Theta^{(k)})$ 를 통해 모수 공간에서 최대/최소의 상태를 두드러지게 하여 안정되게 최우추정치로 유도하게 하는데 있다. 저자는 ECME가 종종 이런 특성을 회손할 수 있다고 본다. 따라서 본 연구에서는 STMM 적합을 위해 $e_{4,ij}^{(k+1)}$ 를 계산하고 $Q_2(\nu|\nu^{(k)})$ 를 최대화하는 ECM 알고리즘의 사용을 제안한다. 이에 대해서는 다음 절에서 소개한다.

3. ECM 알고리즘에 의한 자유도 추정

3.1. $\log U$ 의 조건부 기대값 계산

우선 식 (2.6)으로부터

$$Q_2(\nu|\nu^{(k)}) \propto \sum_{i=1}^g \sum_{j=1}^n \tau_{ij}^{(k+1)} \left\{ -\log \Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} \log\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} [e_{4,ij}^{(k+1)} - e_{1,ij}^{(k+1)}] \right\} \quad (3.1)$$

이다. 식 (3.1)에서 우리는 $e_{4,ij}^{(k+1)}$ 를 구해야 하는데, 그 결과는 다음과 같다.

정리 1. 식 (2.3)의 조건부 분포 구조 하에서

$$\begin{aligned} e_{4,ij}^{(k+1)} &= E_{\Theta^{(k)}} [\log U_j | \mathbf{y}_j, Z_{ij} = 1] \\ &= S_{1,ij}(\Theta^{(k)}) + S_{2,ij}(\Theta^{(k)}), \quad i = 1, \dots, g; j = 1, \dots, n \end{aligned} \quad (3.2)$$

이다. 여기서

$$S_{1,ij}(\Theta^{(k)}) = e_{1,ij}^{(k+1)} - \left(\frac{\nu_i^{(k)} + p}{\nu_i^{(k)} + D_{ij}^{(k)}} \right) + \psi\left(\frac{\nu_i^{(k)} + p}{2}\right) - \log\left(\frac{\nu_i^{(k)} + D_{ij}^{(k)}}{2}\right), \quad (3.3)$$

$$S_{2,ij}(\Theta^{(k)}) = 2 \left[\frac{d}{d\nu_i} \log T\left(r_{ij} \sqrt{\frac{\nu_i + p}{\nu_i + D_{ij}}}; \nu_i + p \right) \right]_{\Theta = \Theta^{(k)}} \quad (3.4)$$

을 나타낸다.

증명: 부록에 수록하였음. □

식 (3.3)에서 $\psi(x) = d \log \Gamma(x)/dx$ 로서 digamma 함수를 나타낸다. 식 (3.3)의 $S_{1,ij}(\Theta^{(k)})$ 는 Lin (2010)이 유도한 STMM-Sahu 모형에서의 결과 정확히 일치한다. 반면 추정치 $\nu_i^{(k)}$ 에서 단변량 로그 t -cdf의 (2배의) 기울기로서 해석될 수 있는 $S_{2,ij}(\Theta^{(k)})$ 는 Lin의 것과 현저히 다르다. 다음 절의 실험에서 $S_{2,ij}(\Theta^{(k)})$ 내의 미분은 수치적으로 계산하였는데, 구체적으로 말하자면 $[(d/dx)f(x)]_{x=x_0} \approx \{f(x_0 + \Delta x) - f(x_0)\}/\Delta x$ 에서 $\Delta x = 10^{-10}$ 를 넣어 계산하였다. 애초에 식 (3.2) 조건부 기대값은 계산이 용이하지 않을 것이라 짐작하였으나, 실제로는 STMM-Sahu 모형에서의 그것과는 다르게 현실에서 수치미분을 이용하여 매우 정확하고 빠르게 계산될 수 있다.

3.2. 자유도의 추정

이제 식 (3.1)의 $Q_2(\nu|\nu^{(k)})$ 를 최대화 하는

$$\nu^{(k+1)} = \operatorname{argmax}_{\nu} Q_2(\nu|\nu^{(k)}) \quad (3.5)$$

를 구해야 하는데, 이를 위해 심플렉스 탐색기법을 사용할 수 있지만 이 기법은 대용량 다변량 자료일 때 많은 처리시간을 요한다. 그래서 본 연구에서는 (많은 실험을 통해 결과값에서 거의 차이가 없음을 확인한) 처리시간이 좀 더 빠른 Gauss-Newton 알고리즘을 사용하였다. 우선

$$0 = \frac{\partial Q_2(\nu|\nu^{(k)})}{\partial \nu_i} = \psi\left(\frac{\nu_i}{2}\right) - \log\left(\frac{\nu_i}{2}\right) - 1 - \bar{e}_i^{(k+1)} \stackrel{\text{def}}{=} h(\nu_i; \theta^{(k+1)}) \quad (3.6)$$

라 하자. 단, $\bar{e}_i^{(k+1)} = \sum_{j=1}^n \tau_{ij}^{(k+1)}(e_{4,ij}^{(k+1)} - e_{1,ij}^{(k+1)}) / \sum_{j=1}^n \tau_{ij}^{(k+1)}$ 을 나타낸다. 이때

$$h'(\nu_i; \theta^{(k+1)}) = \frac{1}{2} \psi'\left(\frac{\nu_i}{2}\right) - \frac{1}{\nu_i} \quad (3.7)$$

이다. 여기서 $\psi'(x)$ 는 trigamma 함수이며, “ ’ ”은 ν 에 관한 1차 미분을 표시한다. 이제 초기치로서 $\nu_i^{(k:0)} = \nu_i^{(k)}$ 라 주고

$$\nu_i^{(k:\ell+1)} = \left[\nu_i^{(k:\ell)} - \frac{h(\nu_i^{(k:\ell)}; \theta^{(k+1)})}{h'(\nu_i^{(k:\ell)}; \theta^{(k+1)})} \right]_+ , \quad i = 1, \dots, g \quad (3.8)$$

를 $\ell = N$ 번의 충분한 반복 후 $\nu_i^{(k:N)}$ 을 $\nu_i^{(k+1)}$ 로 정한다. 여기서 $[x]_+$ 은 $x \leq 0$ 이면 $x = \epsilon (> 0)$ 으로의 치환을 의미하는데, 자유도에 대한 양의 추정을 위한 일환이다. 다음 절 실험에서는 $\epsilon = 0.01$ 으로 하였다. 만약 $\nu_1 = \dots = \nu_g = \nu$ 로서 등자유도를 가정한다면,

$$h(\nu; \theta^{(k+1)}) = \psi\left(\frac{\nu}{2}\right) - \log\left(\frac{\nu}{2}\right) - 1 - \bar{e}^{(k+1)} \quad \text{및} \quad h'(\nu; \theta^{(k+1)}) = \frac{1}{2} \psi'\left(\frac{\nu}{2}\right) - \frac{1}{\nu} \quad (3.9)$$

라 하고,

$$\nu^{(k:\ell+1)} = \left[\nu^{(k:\ell)} - \frac{h(\nu^{(k:\ell)}; \theta^{(k+1)})}{h'(\nu^{(k:\ell)}; \theta^{(k+1)})} \right]_+ \quad (3.10)$$

을 반복 실행한다. 단, $\bar{e}^{(k+1)} = \sum_{i=1}^g \sum_{j=1}^n \tau_{ij}^{(k+1)}(e_{4,ij}^{(k+1)} - e_{1,ij}^{(k+1)}) / n$ 을 나타낸다.

4. 수치실험

4.1. 모의 실험

본 실험에서는 다음의 모수를 가지는 2-변량 2-성분 혼합모형

$$\begin{aligned} \pi &= 0.6, \quad \mu_1 = (0, 0)^T, \quad \mu_2 = (5, 4)^T, \quad \Sigma_1 = \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 5 \end{bmatrix}, \\ \delta_1 &= [0, 0]^T, \quad \delta_2 = [5, 3]^T, \quad \nu_1 = 50, \quad \nu_2 = 5 \end{aligned} \quad (4.1)$$

을 고려하였다. 두 성분밀도에서 각각 $n_1 = 300$ 및 $n_2 = 200$ 개의 자료를 생성하여 ECM 알고리즘으로 모수들을 추정하였다. 그리고 이러한 $n = n_1 + n_2 = 500$ 크기의 표본 세트를 100번 재생성하여 추정치들의 평균 및 표준편차 (괄호 내)를 얻은 결과는 다음과 같았다.

$$\begin{aligned} \bar{\pi} &= 0.6009 (0.0060), \\ \bar{\mu}_1 &= \begin{bmatrix} 0.6937 (0.2025) \\ -0.9612 (0.4198) \end{bmatrix}, & \bar{\mu}_2 &= \begin{bmatrix} 5.0204 (0.1817) \\ 3.9719 (0.3929) \end{bmatrix}, \\ \bar{\Sigma}_1 &= \begin{bmatrix} 4.4148 (0.4237), & 2.6693 (0.4966) \\ - & 9.8820 (1.1658) \end{bmatrix}, & \bar{\Sigma}_2 &= \begin{bmatrix} 1.0377 (0.2810), & 0.2230 (0.4125) \\ - & 5.3544 (1.5430) \end{bmatrix}, \\ \bar{\delta}_1 &= \begin{bmatrix} 0.3903 (0.1261) \\ 0.3946 (0.1961) \end{bmatrix}, & \bar{\delta}_2 &= \begin{bmatrix} 5.4028 (3.1123) \\ 3.3050 (2.0187) \end{bmatrix}, \\ \bar{\nu}_1 &= 37.6533 (27.4534), & \bar{\nu}_2 &= 5.8913 (1.8740). \end{aligned} \quad (4.2)$$

대체적으로 추정치들의 평균은 원 모수에 근접해 있는 것으로 보인다. 그러나 큰 자유도 추정치의 변이가 꽤 크게 나타났다. 이론적으로 자유도 $\nu = \infty$ 일 때 치우친 t -분포는 치우친 정규분포가 된다. 그러나 이것은 현실에서 대략 30 이상이면 치우친 정규분포의 적합과 거의 구분하기 어렵기 때문에 나타나는 어찌면 자연스러운 현상으로 보여진다.

4.2. AIS 자료 적합

AIS(Australian Institute of Sport) 자료 (Cook와 Weisberg, 1994)는 102명의 남성과 100명의 여성 운동선수들의 여러 특성을 조사한 자료로서 우리는 이 중에 두 가지 특성변량(BMI, Bfat)을 사용할 것이다. 이 자료는 두 자료집단의 비대칭성 때문에 치우친 분포 모형의 적합능력을 평가하는 벤치마킹 사례로 자주 이용된다.

Table 1에 STMM-AzDv 모형에 대한 뉘자유도 상황 하에서의 ECM 알고리즘과 등자유도 가정 하에서의 ECME 알고리즘으로 추정된 결과들을 수록하였다. 아울러 STMM-Sahu 모형에 대한 두 가지 방법 즉 Lee와 McLachlan (2011)의 ECM과 Lin (2010)의 EM-MC에 의한 결과를 추가하였는데, 이것은 Lee와 McLachlan이 보고한 결과를 그대로 인용한 것이다 (두 STMM 모형에서 Σ_i 는 다르게 정의되므로 분산 및 공분산 추정치를 비교할 필요는 없다). 그리고 그 추정치들을 바탕으로 Figure 1(아래쪽 두 그림)에 적합 등고선을 그렸다.

Table 1의 로그-우도, AIC 및 BIC 기준으로 볼 때 두 STMM 모형에서 ECM에 의한 적합이 좀 더 나아보이며, 그 중에서도 제안된 STMM-AzDv 모형 하에서 (이자유도를 가정한) ECM 알고리즘이 가장 우수한 것으로 평가된다. 특히 혼합비율 π 의 추정치가 정확한데, STMM-Sahu 모형 기반 방법들은 오히려 여성의 비율이 남성 그것보다 크다고 그릇되게 추정하고 있다. 그 이유는 Figure 1을 보면 대략 이해할 수 있다. STMM-AzDv 모형은 대체적으로 남성자료(Δ)를 치우친 분포로 그리고 여성자료(\circ)를 치우침이 없는 분포로 적합하려는 경향이 있는 반면, STMM-Sahu 모형은 그 반대로 보인다. 그래서 STMM-Sahu 모형 기반 방법들은 남성 자료의 상당수를 여성쪽의 성분분포로 흡수하면서 할당오류율을 높이게 한다. 특히 본 논문의 ECM 알고리즘은 여성자료에 대해 거의 순수한 2-변량 정규분포로 적합하면서 (치우침 모수 $\hat{\delta}_1 = 0.86$ 과 $\hat{\delta}_2 = 0.52$ 로서 매우 작고, $\hat{\nu}_2 = 63.01$ 인데 보통 자유도가 30 이상이면 정규분포와 큰 차이가 없다), 성분분포의 밀도를 고려하며 두 집단의 경계를 다른 기법에 비해 상대적으로 더 정확하게 나누는 것을 그림을 통해 볼 수 있다. 이것은 가장 작은 할당오류율 0.0495 (= 10개/202개)을 제공한다. 이러한 구현은 등자유도를 가정하면 어려운 것으로 보인다.

Table 1: Comparison of Various methods for AIS data. Etime indicates the execution time measured in second.)

모수	STMM-AzDv				STMM-Sahu			
	ECM(equal DF)		ECME(equal DF)		ECM(unequal DF)		EM-MC(unequal DF)	
	male	female	male	female	male	female	male	female
π_i	0.501	0.499	0.597	0.403	0.42	0.58	0.41	0.59
μ_{i1}	21.81	20.36	22.37	20.18	22.55	19.95	22.47	19.89
μ_{i2}	6.15	15.26	6.13	13.87	7.25	15.82	7.30	15.50
σ_{i1}^2	5.95	7.85	5.74	7.08	3.30	3.83	3.23	2.96
σ_{i2}^2	9.26	34.58	26.35	41.05	2.15	25.60	2.14	25.80
$\sigma_{i,12}$	5.96	11.85	9.18	12.94	1.35	6.19	1.34	6.17
δ_{i1}	2.63	0.86	2.23	0.14	0.64	3.55	0.72	2.72
δ_{i2}	5.72	0.52	10.72	1.22	1.23	1.82	1.13	2.22
ν_i	3.20	63.01		6.85	25.92	17.65	25.93	23.98
$L(\Theta)$	-1075.2		-1084.5		-1078.9		-1088.1	
AIC	2184.3		2201.0		2191.8		2208.0	
BIC	2240.6		2253.9		2248.1		2264.2	
Error Rate	0.0495		0.0941		0.0891		0.0891	
Etime	0.62		14.61		-		-	

한편, ECM 알고리즘의 100회 반복 당 처리시간은 상용 PC(Pentium dual-core, 2.5GHz)를 사용하여 0.62초 걸렸다. 반면, 등자유도를 가정한 ECME는 거의 24배인 14.61초가 걸렸으며, 이자유도를 가정하면 이보다 약 2.5배 더 오래 걸렸다.

Figure 2는 등자유도 가정을 하지 않았을 때 알고리즘 진행에 따른 ECME의 자유도 추정의 문제점을 분명하게 보여주고 있다. (ν_1, ν_2) 공간에서 ECM 알고리즘 (첫번째 그림)은 어떤 정상점(stationary point)을 향해 단조적으로 수렴 중인 반면, MCEM (두번째 그림)은 불규칙적인 진행 경로를 보이면서 발산한다. 이것은 로그-우도의 증분 상황 (세번째 그림)에 그대로 반영된다.

5. 결론 및 토의

본 연구에서는 Cabral 등 (2012)의 STMM-AzDv 모형 적합을 위한 ECM 알고리즘의 사용을 제안하였다. 이를 위해서는 $\log U_j$ 의 조건부 기대값을 얻어야 한다. 본 논문에서 이 조건부 기대값을 직접 계산이 가능한 수준의 명시적 형태로 제공하였다. ECM 알고리즘은 기존의 Cabral 등 (2012)의 ECME 알고리즘과는 다르게 이자유도 상황 하에서도 안정적이고 신속하게 최우추정을 할 수 있음 보였으며, AIS 자료를 통해 등자유도를 가정하지 않았을 때 적합이 좀 더 정확할 수 있음을 보였다.

그러나 4절의 모의실험에서 확인한 바와 같이 자유도 추정치가 클수록 산포가 큰데, 좀 더 안정적인 추정법을 고려해야 할 필요가 있다. 예를 들어 벌점화 최우추정(penalized maximum likelihood)이 도움이 될 것으로 사료된다. 식 (3.2)의 조건부 기대값에서 성분 $\sum_j \tau_{ij}^{(k)} S_{2,ij}(\Theta^{(k)})$ 은 $\sum_j \tau_{ij}^{(k)} S_{1,ij}(\Theta^{(k)})$ 에 비해 무시할 만큼 작다는 것을 수치 조사로 확인하였다. 따라서 $e_{4,ij}^{(k+1)} \approx S_{1,ij}(\Theta^{(k)})$ 로 하여도 거의 비슷한 결과를 얻을 수 있을 것으로 예상된다. 마지막으로 STMM-AzDv의 식별성(identifiability)이 이론적으로 확실하게 규명되어 있지 않아 논란의 여지가 있다. 만약 식별성을 만족하지 않는다면 이에 대한 대안적 방법이 요구된다.

부록:

정리 1을 증명하기 전에 먼저 다음 성질을 알아 두자. $g(u; \alpha, \beta)$ 를 $\text{gamma}(\alpha, \beta)$ 분포의 밀도이고 양

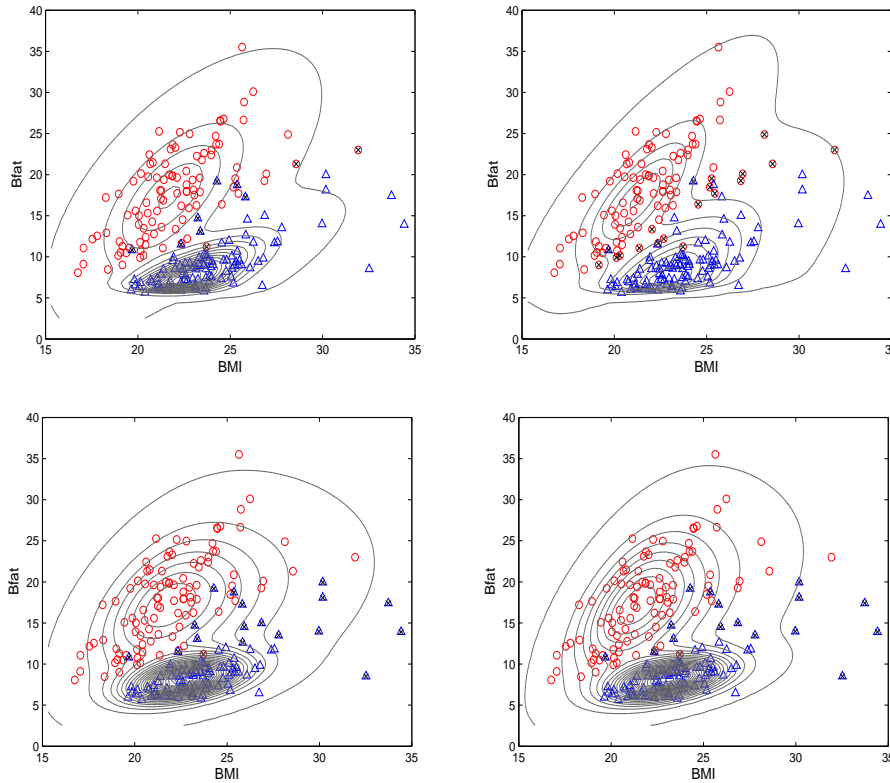


Figure 1: Results of Fit by STMM (upper: STMM-AzDv model (left: ECM, right: ECME), lower: STMM-Sahu model (left: ECM, right: EM-MC), \circ : female, \triangle : male, \times : miss-allocations, contours: fitted model)

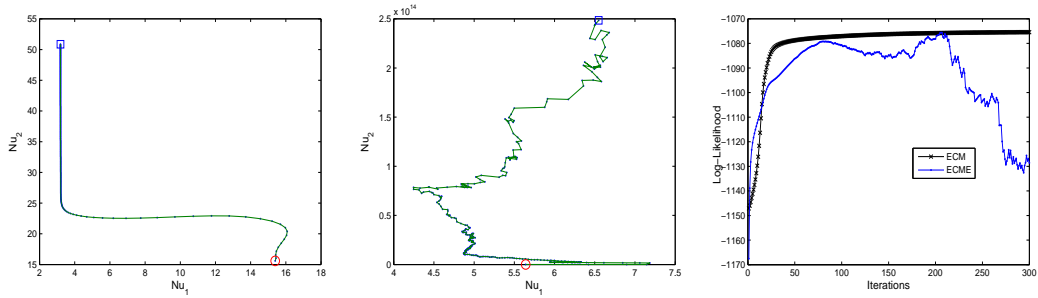


Figure 2: Estimates of ν of ECM and ECME during Iteration (1st: ECM, 2nd: ECM, and 3rd: sequence of log-likelihoods, \circ : beginning point, \square : running point)

의 두 함수 $a(\nu)$ 과 $b(\nu)$ 이 ν 에 관한 1차 미분가능함수라 할 때,

$$\begin{aligned} \frac{d}{d\nu} \log g(u; a(\nu), b(\nu)) &= \frac{d}{d\nu} \log \left\{ \Gamma^{-1}(a(\nu)) b(\nu)^{a(\nu)} u^{a(\nu)-1} \exp[-b(\nu)u] \right\} \\ &= -a'(\nu)\psi[a(\nu)] + a'(\nu) \log[b(\nu)] + \frac{a(\nu)}{b(\nu)} b'(\nu) + a'(\nu) \log(u) - b'(\nu)u \end{aligned}$$

이므로, 양변에 조건부 기대값을 취하면

$$E \left[\frac{d}{dv} \log g(U; a(v), b(v)) \mid \mathbf{y} \right] = -a'(v)\psi(a(v)) + a'(v) \log b(v) + \frac{a(v)}{b(v)} b'(v) + a'(v)E[\log(U) \mid \mathbf{y}] - b'(v)E[U \mid \mathbf{y}] \quad (\text{A.1})$$

이다.

증명: {정리 1의 증명} 앞으로 $Z_{ij} = 1$ 인 상황 하에서 전개하되, 표기의 간편함을 위해 아래첨자 i, j 와 반복횟수를 나타내는 위첨자 (k)는 떼고 설명하겠다. 이렇게 하여도 일반성을 잃지 않는다. 먼저 식 (2.3)의 관계로부터 조건부 밀도

$$h(u \mid \mathbf{y}) = \frac{g(u)f(\mathbf{y} \mid u)}{\int_0^\infty g(u)f(\mathbf{y} \mid u)du} = g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \Phi\left(u^{\frac{1}{2}}r\right) T^{-1}(w(\nu; \mathbf{y})r; \nu+p) \quad (\text{A.2})$$

을 얻을 수 있다. 여기서 $w(\nu; \mathbf{y}) = \sqrt{(\nu+p)/(\nu+D)}$ 를 나타낸다. 이때

$$\begin{aligned} 0 &= \frac{d}{dv} \int_0^\infty h(u \mid \mathbf{y}) du = \int_0^\infty \frac{d}{dv} h(u \mid \mathbf{y}) du \quad (\text{Leibnitz rule에 의해}) \\ &= \int_0^\infty \Phi\left(u^{\frac{1}{2}}r\right) \frac{d}{dv} \left[g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) T^{-1}(w(\nu; \mathbf{y})r; \nu+p) \right] du \\ &= \int_0^\infty \Phi\left(u^{\frac{1}{2}}r\right) \left[\frac{d}{dv} g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \right] T^{-1}(w(\nu; \mathbf{y})r; \nu+p) du \\ &\quad + \int_0^\infty \Phi\left(u^{\frac{1}{2}}r\right) g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \left[\frac{d}{dv} T^{-1}(w(\nu; \mathbf{y})r; \nu+p) \right] du \\ &\stackrel{\text{let}}{=} S_1^*(\nu) + S_2^*(\nu) \end{aligned} \quad (\text{A.3})$$

이다. 그런데

$$\begin{aligned} S_1^*(\nu) &= \int_0^\infty \left[\frac{d}{dv} g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \right] \Phi\left(u^{\frac{1}{2}}r\right) T^{-1}(w(\nu; \mathbf{y})r; \nu+p) du \\ &= \int_0^\infty \left[\frac{d}{dv} \log g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \right] g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \Phi\left(u^{\frac{1}{2}}r\right) T^{-1}(w(\nu; \mathbf{y})r; \nu+p) du \\ &= \int_0^\infty \left[\frac{d}{dv} \log g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \right] h(u \mid \mathbf{y}) du \\ &= E_{\Theta} \left[\frac{d}{dv} \log g\left(U; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \mid \mathbf{y} \right] \quad \left(\text{식 (A.1)에 } a(v) = \frac{\nu+p}{2} \text{ 및 } b(v) = \frac{\nu+D}{2} \text{를 대입하면} \right) \\ &= \frac{1}{2} \left[-\psi\left(\frac{\nu+p}{2}\right) + \log\left(\frac{\nu+D}{2}\right) + \left(\frac{\nu+p}{\nu+D}\right) + E_{\Theta}(\log U \mid \mathbf{y}) - E_{\Theta}(U \mid \mathbf{y}) \right] \end{aligned} \quad (\text{A.4})$$

이다. 그리고 t -cdf $T(w(\nu; \mathbf{y})r; \nu+p)$ 에는 U 의 실현값 u 가 포함되어 있지 않음을 유의하면서

$$\begin{aligned} S_2^*(\nu) &= \int_0^\infty \left[\frac{d}{dv} T^{-1}(w(\nu; \mathbf{y})r; \nu+p) \right] g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \Phi\left(u^{\frac{1}{2}}r\right) du \\ &= -\frac{\frac{d}{dv} T(w(\nu; \mathbf{y})r; \nu+p)}{T(w(\nu; \mathbf{y})r; \nu+p)} \int_0^\infty g\left(u; \frac{\nu+p}{2}, \frac{\nu+D}{2}\right) \Phi\left(u^{\frac{1}{2}}r\right) T^{-1}(w(\nu; \mathbf{y})r; \nu+p) du \\ &= -\frac{d}{dv} [\log T(w(\nu; \mathbf{y})r; \nu+p)] \int_0^\infty h(u \mid \mathbf{y}) du = -\frac{d}{dv} [\log T(w(\nu; \mathbf{y})r; \nu+p)] \end{aligned} \quad (\text{A.5})$$

임을 알 수 있다. 이제 두 결과 (A.4)와 (A.5)를 (A.3)에 대입하여 $E_{\Theta}(\log U|y)$ 를 즉시 얻을 수 있다. 그리고 $Z_{ij} = 1$ 의 조건을 취하여 성분첨자 i 와 관측치 첨자 j 를 붙이면서 $\Theta = \Theta^{(k)}$ 를 대입하면 정리 1의 결과를 얻게 된다. \square

References

- Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian Journal of Statistics*, **33**, 561–574.
- Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew normal distribution, *Biometrika*, **83**, 715–726.
- Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution, *Journal of the Royal Statistical Society, series B* **65**, 367–389.
- Cabral, C. S., Lachos, V. H. and Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distribution, *Computational Statistics and Data Analysis*, **56**, 126–142.
- Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*, **56**, Wiley, New York.
- Lee, S. and McLachlan, G. J. (2011). On the fitting of mixtures of multivariate skew t -distributions via the EM algorithm, *Technical Report of University of Queensland*, Available from: http://arxiv.org/PS_cache/arxiv/pdf/1109/1109.4706v1.pdf.
- Lin, T.-I. (2010). Robust mixture modeling using multivariate skew t distributions, *Statistics and Computing*, **20**, 343–356.
- Lin, T.-I., Lee, J.-C. and Hsieh, W. J. (2007b). Robust mixture modeling using the skew t distributions, *Statistics and Computing*, **17**, 81–92.
- Lin, T.-I., Lee, J.-C. and Yen, S. Y. (2007a). Finite mixture modeling using the skew normal distributions, *Statistica Sinica*, **17**, 909–927.
- Liu, C. and Rubin, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with fast monotonic convergence, *Biometrika*, **81**, 633–784.
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, Wiley, New York.
- Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distribution with application to Bayesian regression model, *The Canadian Journal of Statistics*, **31**, 129–150.