한국통계학회논문집
2009, 16권, 4호, 573–578

벌점가능추정치의 일치성에 대하여

안성만[1,a]

[a]국민대학교 경영대학

요 약

본 논문에서는 과대적합모형을 가정하여 성분의 수에 대한 벌점항목을 추가한 Ahn (2001)의 벌점가능도에 의한 추정치가 일치성을 달성하는 것에 대한 증명을 제시하였다. Wald (1949)의 조건을 만족하는 지를 보이기 위하여 과대적합모형를 활용하였다. 또한 모수가 한계점으로 수렴할 경우 우도값이 무한대로 간다는 문제점을 해결하였다.

주요용어: 벌점가능도, 일치성, 정규혼합모형.

1. 서론

유한 혼합모형(finite mixture model)은 여러 분야에서 적용되는 모형이다. 특히 정규혼합모형은 성분이 정규분포로서 다음과 같이 정의된다.

$$h(x;\pi,\mu,\sigma) = \sum_{j=1}^{g} \pi_j N(x;\mu_j,\sigma_j). \tag{1.1}$$

이 모형에서의 모수는 π_j, μ_j 그리고 σ_j, $j = 1, 2, \ldots, g$ 등이며 일반적으로 성분의 수 g는 주어진 것으로 본다. 이모형에서의 모수의 추정은 최대우도 추정법을 사용하고 있으며 해를 찾는 대표적인 방법은 EM알고리듬이다. 정규혼합모형에서 최대우도추정법을 사용할 때의 문제점은 모수가 한계점으로 수렴할 경우 우도값이 무한대로 간다는 점이다 (Cheng과 Liu, 2001; Ciuperca 등, 2003). 예를 들면, 정규분포의 평균이 데이터 하나와 일치하고 분산이 0으로 수렴하면 우도값은 무한대가 된다. 이러한 문제의 해결을 위하여 몇몇 저자들은 분산에 제약을 두거나 (Hathaway, 1985), 벌점가능도를 사용하기도 하였다 (Ciuperca 등, 2003). 한편 최대우도 추정을 사용한 경우 추정치의 일치성에 대한 연구도 진행되어 왔다. 일치성에 대한 연구는 몇몇 연구자 (Cramér, 1946; Wald, 1949)에 의하여 시작되었으며, 특히 유한혼합모형에 대한 연구는 Redner (1981)가 발표한 바 있다. 그러나 Redner는 모수가 한계점으로 수렴하는 경우의 문제점을 수용하지 못하여서, 추정치가 일치성을 달성하는 조건을 제시하는 연구 (Feng과 McCulloch, 1996; Cheng과 Liu, 2001)가 발표되기도 하였고, 혹은 벌점가능도를 사용하여 그 추정치가 일치성을 만족한다는 것을 보이기도 하였다 (Ciuperca 등, 2003; Chen 등, 2008). 본 연구에서는 Ahn (2001)이 제안한 벌점가능도 모형을 이용하여 추정치가 일치성을 만족하는 것을 보이도록 하겠다. Ahn (2001)이 제안한 벌점가능도는 Ciuperca 등 (2003)과 Chen 등 (2008)이 사용한 σ에 대한 벌점항목과 달리 π에 대한 벌점항목을 사용하였다. 그러한 점에서 기존의 벌점가능도 항목과 큰 차이가 있으며, 또한 그럼에도 불구하고 σ가 한계점으로 수렴하는 경우의 문제점도 함께 해결하는 장점이 있음을 보이도록 하겠다.

본 연구는 2009년 국민대학교 교내학술연구비 지원으로 이루어졌음.
[1] (136-702) 서울시 성북구 정릉동 861-1, 국민대학교 경영대학, 부교수. E-mail: sahn@kookmin.ac.kr

2. 벌점가능도

$x_1, x_2, \ldots, x_n$를 독립적이고 같은 분포를 따르는 1차원 확률변수라고 하고, x_1의 분포는 정규혼합분포를 따르고 모수 θ의 값은 모르는 상태이다. 모수 공간을 다음과 같이 정의한다.

$$\Omega = \left\{ \theta = (\pi_1, \ldots, \pi_g, \mu_1, \ldots, \mu_g, \sigma_1, \ldots, \sigma_g) : \sum_{j=1}^{g} \pi_j = 1,\ \pi_j \geq 0,\ \sigma_j \geq 0,\ j = 1, \ldots, g \right\}. \tag{2.1}$$

Ω를 모수의 집합으로 정의하고 θ_0를 실제 모수라고 하자. 또한 σ-유한척도 μ가 존재하며 μ에 대하여 절대연속인 확률척도 μ_θ가 존재한다 가정한다. Ahn (2001)이 제안한 로그 벌점가능도는 다음과 같다.

$$\log f(x;\theta) = \sum_{i=1}^{n} \log h(x_i;\theta) + \lambda n \sum_{j=1}^{g} (\alpha_j - 1) \log \pi_j. \tag{2.2}$$

유한혼합모형에서의 모수추정에서는 반복적절차를 이용하는 것이 일반적이다. 이러한 반복적 절차 중 하나가 EM알고리즘이며, Dempster 등 (1977)에 따르면, 각 데이터 $x_1, x_2, \ldots, x_n$은 관측되지 않는 또 다른 데이터 $z_1, z_2, \ldots, z_n$과 대응된다고 가정한다. 여기서 z_i는 인디케이터 벡터($z_i = (z_{i1}, z_{i2}, \ldots, z_{ig})^T$)로서 $z_{ij} = 1$이라면 x_i가 j번째 성분에 의해 생성되었다는 것을 의미한다. 식 (2.2)에 대한 EM알고리즘의 M-step에서 π_j의 추정치는 다음과 같다 (Ahn, 2001).

$$\hat{\pi}_j = \frac{\frac{1}{n}\sum_{i=1}^{n} z_{ij} + \frac{1}{n}(\alpha_j - 1)}{1 + \frac{1}{n}(\sum_{j=1}^{g} \alpha_j - g)}. \tag{2.3}$$

그리고 벌점가능도를 사용하는 데 있어서 다음과 같은 가정을 하였다.

가정 1: $\pi_j > \epsilon$ ($\epsilon > 0$인 임의의 작은 수).

가정 2: 과대적합모형- 현재의 모형은 과대적합모형이다. 그래서 반복적 방법으로 모수를 추정하는 과정에서 π_j가 가정 1에서 ϵ보다 작은 값이 되는 경우는 존재하지 않는 성분으로 가정한다. 따라서 해당성분을 제거할 수 있다.

가정 3: 벌점항목- 벌점가능도에서 벌점항목으로 추가된 항에서 π_j는 다음을 만족한다.

$$\sum_{j=1}^{g} (\alpha_j - 1) \log \pi_j < \sum_{j=1}^{g} (\alpha_j - 1) \log \pi_{0j}, \tag{2.4}$$

여기서 π_{0j}는 실제 모수의 값이다. $\sum_{j=1}^{g} \pi_j = 1$ 인 조건에서 $\sum_{j=1}^{g} (\alpha_j - 1) \log \pi_j$는 $\pi_j = (\alpha_j - 1)/\{\sum_{j=1}^{g} (\alpha_j - 1)\}$ 일때 최소값($(\alpha_j - 1) < 0$인 경우)을 가지게 된다. 여기서 α_j를 같은 값으로 가정하면 $\pi_1 = \cdots = \pi_g = 1/g$인 경우에 최소값이 되므로, 위의 가정은 실제모수의 값은 π_j가 서로 같지 않다는 것을 의미한다. 따라서 식 (2.4)에 의하면 실제 모수는 현재 모수의 값에서 한 개 이상의 π_j가 0으로 더 가까운 값임을 의미하며, 이는 가정 1에서의 과대적합모형을 해결하기 위한 하나의 방법으로서, 이 벌점가능도의 주요 특징이라고 하겠다.

가정 4: Ω에서의 거리를 다음과 같이 정의하고 (Chen 등, 2008),

$$d(\theta, \theta') = \sum_{j=1}^{g} \arctan|\pi_j - \pi'_j| + \sum_{j=1}^{g} \arctan|\mu_j - \mu'_j| + \sum_{j=1}^{g} \arctan|\sigma_j - \sigma'_j|, \tag{2.5}$$

함수 f와 f^*를 다음과 같이 정의하자.

$$f(x;\theta,\rho) = \sup f(x;\theta') : d(\theta,\theta') < \rho, \quad \theta' \in \Omega, \tag{2.6}$$

$$f^*(x;\theta,\rho) = \max(1, f(x;\theta,\rho)). \tag{2.7}$$

그러면 다음의 기대값은 유한하다.

$$E_0 \log f^*(x;\theta,\rho) = \int \log f^*(x;\theta,\rho) du_{\theta_0}. \tag{2.8}$$

가정 4는 앞에서 언급한 우도값이 무한대가 되는 문제를 고려하지 않고 Redner (1981)가 증명한 바 있다. 이 논문의 모형에서는 아래의 보조정리 1에서 이 문제가 일어나지 않는다는 것을 보일 것이므로 이 가정이 만족하는 것을 가정하기로 한다.

3. 일치성

일치성을 증명하기 위하여 우선 몇 가지 보조정리를 제안하기로 한다.

보조정리 1. 임의의 성분 j에 대하여 $n \to \infty$이고 $\mu_j \to x_i$이고 $\sigma_j \to 0$이면 $\pi_j \to \epsilon(\epsilon < 0)$이다 (단, $\lambda = -1/\sum_{j=1}^{g}(\alpha_j - 1)$).

증명: z_{ij}는 데이터 i를 성분 j가 만들었을 확률을 의미하므로 $n \to \infty$이고 $\sigma_j \to 0$이면 $z_{ij} \to 1$이고 $z_{kj} \to 0(k \neq i)$이 된다. 따라서 $1/n \sum z_{ij} \to 0$이 되며, $\pi_j \to \{\lambda(\alpha_j - 1)\}/\{1 + \lambda \sum_{j=1}^{g}(\alpha_j - 1)\}$이 된다. 앞의 가정에서 $(\alpha_j - 1) < 0)$이므로 $\lambda = -1/\{\sum_{j=1}^{g}(\alpha_j - 1)\}$을 만족하면 $\pi_j \to \epsilon(\epsilon < 0)$이다. □

Ahn (2001)에 의하면 α와 λ를 각각 0.1과 0.002로 사용하였으므로, $\lambda < 0.11$의 조건을 만족하며 그리고 성분의 수(g)를 10으로 한 경우에 $\epsilon \approx -0.0018$정도의 값이 된다. π_j가 음수의 값으로 수렴하는 것이 이상해 보이지만, 그렇게 됨으로써 필요 없는 성분으로 간주되어 모형에서 제거될 수 있게 되는 것이다.

보조정리 2. $\theta \neq \theta_0$인 경우에

$$E_0 \log f(x,\theta) < E_0 \log f(x,\theta_0). \tag{3.1}$$

증명: $v = \log f(x,\theta) - \log f(x,\theta_0)$라고 하면

$$E_0[e^v] = E\left[\frac{f(x,\theta)}{f(x,\theta_0)}\right] = \int h(x,\theta)\left[\frac{\prod_{j=1}^{g}\pi_j^{(\alpha_j-1)}}{\prod_{j=1}^{g}\pi_{0j}^{(\alpha_j-1)}}\right]^{\lambda n} dx = \left[\prod_{j=1}^{g}\left(\frac{\pi_j}{\pi_{0j}}\right)^{(\alpha_j-1)}\right]^{\lambda n} \tag{3.2}$$

이다. 따라서

$$\log E_0[e^v] = \lambda n \sum_{j=1}^{g}(\alpha_j - 1)\log \pi_j - \sum_{j=1}^{g}(\alpha_j - 1)\log \pi_{0j} < 0 \tag{3.3}$$

이다. 마지막 부등호는 가정 3에 근거한다. 로그함수가 오목하므로 Jensen 부등식에 의하여 $E_0[v] \leq \log E_0[e^v] < 0$이 성립하므로 식 (3.1)가 성립한다. □

보조정리 3. 다음이 성립한다.

$$\lim_{\rho \to 0} E_0 \log f(x;\theta,\rho) = E_0 \log f(x;\theta). \tag{3.4}$$

증명: $f^*(x;\theta) = \max(1, f(x;\theta))$라고 정의했을 때, $\lim_{\rho\to 0} \log f^*(x;\theta,\rho) = \log f^*(x;\theta)$이 성립한다. 그러므로 가정4와 Wald (1949)의 보조정리 2에 따라서 다음이 성립한다.

$$\lim_{\rho \to 0} E_0 \log f^*(x;\theta,\rho) = E_0 \log f^*(x;\theta). \tag{3.5}$$

또한, $f^{**}(x;\theta) = \min(1, f(x;\theta))$이고 $f^{**}(x;\theta,\rho) = \min(1, f(x;\theta,\rho))$로 정의하면, Wald (1949)의 보조정리 2에 따라서 다음이 성립한다.

$$\lim_{\rho\to 0} E_0 \log f^{**}(x;\theta,\rho) = E_0 \log f^{**}(x;\theta). \tag{3.6}$$

식 (3.5)와 (3.6)에 의해서 식 (3.4)가 성립한다. □

이제 벌점최대우도추정치의 일치성에 대한 증명을 하겠다. 이 증명은 Wald (1949)의 증명의 구조를 따른다.

정리 1. 닫힌 집합 S는 Ω의 부분집합으로서 실제 모수인 θ_0를 포함하지 않는다. 그러면 다음이 성립한다.

$$P\left(\lim_{n\to\infty} \sup_{\theta\in S} \frac{\prod_{i=1}^n f(x_i;\theta)}{\prod_{i=1}^n f(x_i;\theta_0)} = 0\right) = 1. \tag{3.7}$$

증명: 우선 보조정리 1에 의하여 $n \to \infty$이고 $\mu_j \to x_i$이고 $\sigma \to 0$인 경우에 우도값이 무한대로 가는 경우는 해당하는 성분이 삭제되기 때문에, 모수가 한계점으로 수렴할 때 우도값이 무한대가 되는 문제는 없다. 한편 집합 S의 한 원소인 θ에 대하여 다음 식을 만족하는 양수 ρ_θ를 할당한다고 하자. 보조정리 2와 3에 의하여 ρ_θ는 존재한다.

$$E_0 \log f(x,\theta,\rho_\theta) < E_0 \log f(x,\theta_0). \tag{3.8}$$

거리를 식 (2.5)와 같이 정의하면, 집합 S는 컴팩트집합이 되므로 (Chen 등, 2008), $S(\theta,\rho)$를 중심이 θ이고 지름이 ρ인 구라고 할때, $\cup_{i=1}^h S(\theta_i,\rho_{\theta_i})$를 만족하는 $\theta_1,\theta_2,\ldots,\theta_h$가 존재한다. 그러면 다음이 성립한다.

$$0 \le \sup_{\theta\in S} \prod_{i=1}^n f(x_i,\theta) \le \sum_{j=1}^h \prod_{i=1}^n f(x_i,\theta_j,\rho_{\theta_j}). \tag{3.9}$$

따라서 식 (3.7)은 다음의 식과 동등하다.

$$P\left(\lim_{n\to\infty} \sum_{i=1}^n (\log f(x_i,\theta_j,\rho_{\theta_j}) - \log f(x_i,\theta_0)) = -\infty\right) = 1, \quad j = 1,\ldots,h. \tag{3.10}$$

식 (3.8)과 대수의 법칙에 의해서 식 (3.10)이 성립함을 알 수 있다. □

정리 2. $\bar\theta_n(x_1,x_2,\ldots,x_n)$는 $x_1,x_2,\ldots,x_n$의 함수로서 다음을 만족한다고 하자.

$$\frac{f(x_1,\bar\theta_n)\cdots f(x_n,\bar\theta_n)}{f(x_1,\theta_0)\cdots f(x_n,\theta_0)} \ge c > 0, \quad \forall x_1,\ldots,x_n,\ \forall n. \tag{3.11}$$

그러면 $P(\lim_{n\to\infty}\bar\theta_n = \theta_0) = 1$이다.

증명: $d(\bar\theta_n,\theta_0) > \epsilon(\epsilon > 0)$ 를 만족하는 $\bar\theta_n$의 극한점인 $\bar\theta$가 존재한다면, 다음을 만족하는 수많은 n이 존재한다.

$$\sup_{d(\theta-\theta_0)\ge\epsilon} f(x_1,\theta)\cdots f(x_n,\theta) \ge f(x_1,\theta_0)\cdots f(x_n,\theta_0). \tag{3.12}$$

그러면 다음을 만족하는 수많은 n이 존재한다.

$$\frac{\sup_{d(\theta-\theta_0)\ge\epsilon} f(x_1,\theta)\cdots f(x_n,\theta)}{f(x_1,\theta_0)\cdots f(x_n,\theta_0)} \ge c > 0. \tag{3.13}$$

그런데 정리 1에 의하면 식 (3.13)은 확률이 0인 사건이다. 그러므로 $P(\lim_{n\to\infty}\bar\theta_n = \theta_0) = 1$이다. □

4. 결론

정규혼합모형에서는 최대우도방식을 이용하여 모수를 추정할 때, 정규분포의 평균이 데이터 하나와 일치하고 분산이 0으로 수렴하면 우도값은 무한대가 된다는 문제가 있다. 이를 해결하기 위한 방법으로 벌점항목을 우도에 추가한 벌점가능도를 이용하는 방안이 제시된 바 있다. (Ciuperca 등, 2003; Chen 등, 2008) 본 논문에서는 과대적합모형을 가정하여 성분의 수에 대한 벌점항목을 추가한 Ahn (2001)의 벌점가능도에 의한 추정치가 일치성을 달성하는 것에 대한 증명을 제시하였다. Ahn (2001)이 제안한 벌점가능도는 Ciuperca 등 (2003)과 Chen 등 (2008)이 사용한 σ에 대한 벌점항목과 달리 π에 대한 벌점항목을 사용한다. 그러한 점에서 기존의 벌점가능도 항목과 큰 차이가 있으며, 그렇기 때문에 논문의 보조정리 2에서 해당 벌점가능도 모형이 식 (3.1)을 만족함을 보였다. 일반적으로 벌점가능도추정치에 대한 일치성은 최대우도추정치의 일치성에 대한 Wald (1949)의 증명을 근거로 하며, 벌점항목을 추가한 벌점가능도의 기대값이 Wald의 조건(위의 보조정리 2)을 만족하는 지를 확인하면 된다. 본 논문의 보조정리 2에서는 Wald의 조건을 만족하는 지를 보이기 위하여 과대적합모형(가정2)을 활용하였다. 또한 보조정리 1을 통하여 모수가 한계점으로 수렴할 경우 우도값이 무한대로 간다는 문제점을 해결하였다.

참고 문헌

Ahn, S. M. (2001). A penalized likelihood method for model complexity reduction in Gaussian Mixture density, *The Korean Communications in Statistics*, **8**, 173–184.

Chen, J., Tan, X. and Zhang, R. (2008). Consistency of penalized MLE for normal mixtures in mean and variance, *Statistica Sinica*, **18**, 443–465.

Cheng, R. C. H. and Liu, W. B. (2001). The consistency of estimators in finite mixtures models, *Scandinavian Journal of Statistics*, **28**, 603–616.

Ciuperca, G., Ridolfi, A. and Idier, J. (2003). Penalized maximum likelihood estimator for normal mixtures, *Scandinavian Journal of Statistics*, **30**, 45–59.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B*, **39**, 1–38.

Feng, Z. D. and McCulloch, C. E. (1996). Using bootstrap likelihood rations in finite mixture models, *Journal of the Royal Statistical Society. Series B*, **58**, 609–617.

Hathaway, R. J. (1985). A constrained formulation of maximum likelihood estimation for normal mixture distributions, *The Annals of Statistics*, **13**, 795–800.

Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions, *The Annals of Statistics*, **9**, 225–228.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *The Annals of Mathematical Statistics*, **20**, 595–601.

2009년 5월 접수; 2009년 5월 채택

Note on the Consistency of a Penalized Maximum Likelihood Estimate

Ahn, SungMahn[1,a]

[a]College of Business Administration, Kookmin University

Abstract

We prove the consistency of a penalized maximum likelihood estimate proposed by Ahn (2001). The PMLE not only avoids the well-known problem that the ordinary likelihood of the normal mixture model is unbounded for any given sample size, but also removes redundant components.

Keywords: PMLE, consistency, normal mixture.

This research was supported by Kookmin University in 2009.

[1] Associate Professor, College of Business Administration, Kookmin University, JeongNeung-Dong, SeongBuk-Gu, Seoul 136-702, Korea. E-mail: sahn@kookmin.ac.kr