Communications in Mathematics and Applications
Vol. 12, No. 3, pp. 749–754, 2021
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications

http://www.rgnpublications.com

DOI: 10.26713/cma.v12i3.1653

Research Article

Relationship Between a Central Limit Theorem and Hotelling's T^2-Statistic in the Context of the Stochastic EM Algorithm Used in Mixture Analysis

Athanase Polymenis*

Department of Economics, University of Patras, University Campus at Rio, Rio-Patras 26504, Greece

Received: July 13, 2021 **Accepted:** October 15, 2021

Abstract. In earlier research the asymptotic distribution of a test statistic, that uses the algebraic representation of Hotelling's T^2 and is pertaining to a process generated from the *Stochastic EM* (SEM) algorithm, was established in order to assess the performance of the EM algorithm in the estimation of the number of components in finite mixtures; theory concerning the distribution of T^2 was based on a regularity assumption stating that the vector random process generated from SEM is normally distributed. In the present paper a central limit theorem and some theory concerning second order moments are used in order to investigate corresponding results obtained in case the process is generated from the stationary state of SEM without making any assumption of normality. A comparison between our findings and usual asymptotic theory for independently distributed vector random variables is also provided.

Keywords. Stochastic EM; Stationary process; Hotelling's statistic; Asymptotic distribution

Mathematics Subject Classification (2020). 62E20; 62F05; 60E05

1. Introduction

Let x be a vector random variable with probability density function the mixture density $f(x,q) = \sum_{k=1}^{K} p_k f(x_i,\theta_k)$, where k stands for the number of the mixture components, p_k is

***Email:** athanase@upatras.gr

the weight of the k-th component ($0 < p_k < 1$), and θ_k denotes the parameter(s) associated with the k-th component. The parameters q are the weights and θ_k. A difficult problem that arises in such models, which has not been solved in its generality, is to estimate the true number of components; the problem occurs because standard asymptotic theory breaks down (Titterington, Smith and Makov [11, pp. 152 – 153]).

In order to circumvent this problem many algorithms have been proposed in existing literature. In the present work we consider the Stochastic EM (or SEM) algorithm as a method of estimation of the number of components(Celeux and Diebolt [4]). Results provided in the latter paper clearly show how useful can be this algorithm for dealing with the mixture problem. From a methodological point of view SEM is just a stochastic version of the well-known EM algorithm. For a comprehensive description of EM see Redner and Walker [10]. Advantages of SEM over EM are fully described in Celeux and Diebolt [6]. In view of its nice properties the SEM algorithm was also proposed to be used as a validity tool for assessing estimates of numbers of components provided by EM via a test statistic which has the same algebraic form as the usual Hotelling's $T^2$for independent vector random variables (Celeux [2]). In previous research the distribution of this statistic was obtained for a sufficiently large number of SEM iterations (Polymenis [7]). This statistic is revisited in this text and the aim is to investigate properties at the stationary state of the SEM algorithm as this will be described in detail in Section 2. For reasons of clarity we now present the SEM algorithm which is as follows.

Define an upper bound K for the (unknown) number of components. Then the SEM iteration $q_r \to q_{r+1}$ comprises of the following three steps.

E-Step: for $k = 1,\ldots,K$ and $i = 1,\ldots,N$, compute $t_k^r(x_i) = p_k^r f(x_i,\theta_k^r) / \sum_{j=1}^{K} p_j^r f(x_i,\theta_j^r)$.

S-step: for every observed x_i ($i = 1,\ldots,N$), draw a single multinomial observation $z^r(x_i) = (z_k^r(x_i), k = 1,\ldots,K)$ with probabilities $t_k^r(x_i)$, $k = 1,\ldots,K$. This procedure amounts to drawing the pseudo-completed sample $y_i = (x_i, z(x_i))$ by replacing each missing quantity $z(x_i)$ by a value drawn at random according to the probabilities $t_k(x_i)$.

M-step: compute the maximum likelihood estimates q_{r+1} using the pseudo-completed sample provided by the S-step. The procedure amounts to computing $p_k^{r+1} = (1/N)\sum_{i=1}^{N} z_k^r(x_i)$ and estimating θ_k^{r+1}.

The remainder of the present text is organized as follows. In Section 2 the models are presented and the problem we are dealing with is described. Section 3 provides mathematical evidence concerning our rationale. Concluding remarks are provided in Section 4.

2. Description of the Problem

As mentioned in the introduction the Stochastic EM algorithm has been found useful for estimating numbers of components in finite mixture models. From a mathematical point of view SEM generates a vector random process which is a Markov chain. In an earlier paper [7] it has been established that, as the sample size N tends to infinity, this chain is an autoregressive process of order 1, denoted as $AR(1)$, and can be written as

$$X_{r+1} = aX_r + \frac{s}{\sqrt{N}}\varepsilon_{r+1} \quad \text{(with initial condition } X_0 = 0), \tag{2.1}$$

where $r > 0$ denotes the number of SEM iterations, ε_r are independently and identically distributed p-dimensional normal random vectors and $\varepsilon_r \sim N(0, I)$, where 0 is a p-dimensional random vector will all entries equal to 0, I is the $p \times p$ identity matrix, s a $p \times p$ matrix with real entries, and a a $p \times p$ matrix with real entries that has all its eigenvalues less than 1 in absolute value (Polymenis [7]). The vector X_r appearing in eq. (2.1) is $q_r - q_N$, where q_r stands for the SEM estimate at iteration r, and q_N is the asymptotically (as $N \to \infty$) convergent solution of EM (Redner and Walker [10]). This will be the model of interest in the present paper and the process appearing in eq. (2.1) will be mentioned as process $\{X_r\}$ in this text. Note that in the one-dimensional case ($p = 1$), we have that $\varepsilon_r \sim N(0,1)$, where $0 \in \mathbb{R}$, s is a positive constant, and a is a constant such that $|a| < 1$. As mentioned in Polymenis [8], at its stationary state eq. (2.1) takes the form

$$X^{(m)} = (s/\sqrt{N}) \sum_{i=0}^{\infty} a^i \varepsilon_{m-i} \quad (m = \dots, -1, 0, 1, 2, \dots), \tag{2.2}$$

denoted by process $\{X^{(m)}\}$ in this text; remark that $X^{(m)}$ corresponds to the stationary solution of eq. (2.1) and process $\{X^{(m)}\}$ is strongly stationary. Note that for the one-dimensional case corresponding to $p = 1$, eq. (2.2) is the same as equation (3.5.20) of Priestley [9, pp. 121 – 122]. An important remark that plays a key role in the development of our theory is that there is an assumption of normality concerning the random errors involved in eq. (2.1) which is part of some regularity assumptions pertaining to theory underlying SEM (Celeux and Diebolt [5, 6]). This assumption stems from the fact that for the one-dimensional case $p = 1$ the random errors in eq. (2.1) have been shown to be normally distributed (Celeux [3, pp. 114 – 115]).

Let us now consider the validity test statistic $T_r^2 = r\overline{X}' M_r^{-1} \overline{X}$ (with $\overline{X} = \sum_{i=1}^{r} X_i / r$, and $M_r = \sum_{i=1}^{r} (X_i - \overline{X})(X_i - \overline{X})'/(r-1)$)which was proposed in the literature (Celeux [2]) in order to assess the validity of the number of estimated components obtained from the well-known EM algorithm. This statistic has the same algebraic representation as Hotelling's T^2, used for independent vector normal random variables, and process $\{X_i\}$ ($i = 1, \dots, r$) satisfies eq. (2.1). The distribution of T_r^2, when r is sufficiently large, under a null hypothesis H_0 corresponding to eq. (2.1), with normally distributed vector random errors, was established in previous research (Polymenis [7]). On the other hand we now mention a central limit theorem which will be helpful for supporting theory presented in the next section. This theorem results from Theorems 1 and 3 of Polymenis [8], and states that in case we do not make any assumption of normality concerning eq. (2.1), the distribution function of $\sqrt{n}\overline{X}^{(\cdot)}$, with $\overline{X}^{(\cdot)} = \sum_{i=1}^{n} X^{(i)}/n$ (where $X^{(i)}$ satisfies eq. (2.2)) will converge, as $n \to \infty$, to a normal distribution function with zero mean vector and covariance matrix equal to $\frac{(I-a)^{-1} s s' (I-a')^{-1}}{N}$. Using these results our main goal is to examine asymptotic distribution properties for the aforementioned algebraic representation considered at the stationary state of the SEM algorithm and compare these properties to corresponding ones from T_r^2 (Polymenis [7]).

3. The Main Result

Part A. In view of the central limit theorem reported in Section 2, the crucial assumption we make is that the random errors appearing in eq. (2.1) are no more normally distributed which in turn implies that process $\{X_r\}$ is not normally distributed as well. Using the idea of asymptotic stationarity introduced by Priestley [9] it has been shown in previous literature that, for r sufficiently large, $E[X^{(0)}(X^{(0)})']$ can be approximated by matrix C, where $C = \frac{\sum_{k=0}^{\infty}(a^k ss'(a')^k)}{N}$ (Polymenis [8, Proof of Theorem 3]), and symbol prime denotes the transpose matrix. Let us consider the algebraic representation $T_n^2 = n\overline{(X^{(\cdot)})}'\{E[X^{(0)}(X^{(0)})']\}^{-1}\overline{X^{(\cdot)}}$ which has the same form as T_r^2 but includes only variates from the stationary state of the SEM algorithm. Since the distribution of $T_n^2 = n\overline{(X^{(\cdot)})}'\{E[X^{(0)}(X^{(0)})']\}^{-1}\overline{X^{(\cdot)}}$is approximately equal to that of $n\overline{(X^{(\cdot)})}'C^{-1}\overline{X^{(\cdot)}}$ when r is sufficiently large, and using the aforementioned central limit from Polymenis [8], it results that T_n^2 has the same distribution as $X'C^{-1}X$, as $n \to \infty$, and X is a normally distributed vector random variable with zero mean vector and constant covariance matrix $\frac{(I-a)^{-1}ss'(I-a')^{-1}}{N}$. For reasons of clarity we also report that in the one-dimensional case, $p = 1$, it was shown that $E[(X^{(0)})^2]$ (the one-dimensional version of $E[X^{(0)}(X^{(0)})']$) can be approximated by $\frac{s^2}{N(1-a^2)}$, for r sufficiently large (Polymenis [8, Proof of Theorem 2(B)]), and $\frac{s^2}{N(1-a^2)}$ is the one-dimensional version of C. It results that in one-dimensional case, the distribution of $T_n^2 = t_n^2$ will be equal, as $n \to \infty$, to that of $\frac{X'X}{\left(\frac{s^2}{N(1-a^2)}\right)}$, and X is a normally distributed random variable with zero mean and constant variance $\frac{s^2}{N(1-a)^2}$ (i.e. the one-dimensional version of matrix $\frac{(I-a)^{-1}ss'(I-a')^{-1}}{N}$).

Part B. Let us now consider process $\{X_r\}$ as in eq. (2.1), and $\{X_r\}$ is normally distributed. Since, on one hand, C is the limit (as $r \to \infty$) in probability of M_r and thus C^{-1} is the limit in probability of M_r^{-1} (Polymenis [7, Proof of Theorem 4]), and on the other hand, by Theorem 3 of Polymenis [7], $\lim_{r\to\infty} \text{var}(\sqrt{r}\overline{X}) = \frac{(I-a)^{-1}ss'(I-a')^{-1}}{N}$, it results that, under H_0, $T_r^2 = r\overline{X}'M_r^{-1}\overline{X}$ has the same distribution as $X'C^{-1}X$ when $r \to \infty$, and X is a normally distributed vector random variable with zero mean vector and constant covariance matrix $\frac{(I-a)^{-1}ss'(I-a')^{-1}}{N}$. As before we now provide some explanations concerning case $p = 1$. In this case, as $r \to \infty$, the denominator of $T_n^2 = \frac{r\overline{X}^2}{\frac{1}{r-1}\sum_{i=1}^{r}(X_i-\overline{X})^2}$ has the same limit in probability as $\frac{1}{r-1}\sum_{i=1}^{r} X_i^2$ (Polymenis [7, Proof of Theorem 4]). Furthermore it has been shown on one hand that $E\left(\sum_{i=1}^{r} X_i^2\right) = \left(\frac{s^2}{N(1-a^2)}\right)r - \frac{a^2(1-a^{2r})s^2}{N(1-a^2)^2}$ (Polymenis [7, Appendix 4]), and thus $E\left[\frac{1}{r-1}\sum_{i=1}^{r} X_i^2\right]$ converges to $\frac{s^2}{N(1-a^2)}$ as $r \to \infty$, and on the other hand that $\text{var}\left(\sum_{i=1}^{r} X_i^2\right)$ is of order r (Polymenis [7, Appendix 5]) implying that $\text{var}\left(\frac{1}{r-1}\sum_{i=1}^{r} X_i^2\right)$ converges to 0 as $r \to \infty$. It results that, as $r \to \infty$, $\frac{1}{r-1}\sum_{i=1}^{r} X_i^2$ converges in probability to $\frac{s^2}{N(1-a^2)}$, which is equal to $E[(X^{(0)})^2]$ obtained in Part A. It results that in one-dimensional case, the distribution of $T_r^2 = t_r^2$ will be equal, as $r \to \infty$, to that of $\frac{X'X}{\left(\frac{s^2}{N(1-a^2)}\right)}$, and X is a normally distributed random variable with zero mean and constant variance $\frac{s^2}{N(1-a)^2}$.

Results reported in Parts A and B show that the asymptotic (as $n \to \infty$) distribution of T_n^2 is equal to that of T_r^2 under H_0 when r is sufficiently large; note that the latter was shown to be approximately equal to the distribution of $\delta_1 Z_1^2 + \cdots + \delta_p Z_p^2$, where δ_i $(i = 1, \ldots, p)$ are eigenvalues of the matrix $C^{-1/2} \frac{(I-a)^{-1} ss'(I-a')^{-1}}{N} C^{-1/2}$, and Z_i^2 are $\chi^2(1)$ independently and identically distributed variates (Polymenis [7, Theorem 4]). These results lead then to the following theorem.

Theorem 3.1. *Let us assume that the p-dimensional process $\{X_r\}$ satisfying eq.* (2.1) *is no more normally distributed and so neither is the corresponding strongly stationary process $\{X^{(m)}\}$ $(m = \ldots, -1, 0, 1, \ldots)$ which satisfies eq.* (2.2)*. The distribution of $T_n^2 = n(\overline{X^{(\cdot)}})' C^{-1} \overline{X^{(\cdot)}}$ is equal, as $n \to \infty$, to the distribution of $\delta_1 Z_1^2 + \cdots + \delta_p Z_p^2$ which is the approximate distribution of the statistic $T_r^2 = r\overline{X}' M_r^{-1} \overline{X}$, obtained for r sufficiently large under the null hypothesis that process $\{X_r\}$ satisfies eq.* (2.1) *and is assumed to be normally distributed.*

Theorem 3.1 can be viewed as the analogue, for random processes arising from the stationary state of the SEM algorithm, to Theorem 5.2.3 of Anderson [1, p. 163] for independent processes. In the special case where matrix a appearing in eq. (2.1) is equal to matrix [0], i.e. the matrix with all its entries equal to 0, process $\{X_r\}$ is independent, and thus Theorem 3.1 is equivalent to Theorem 5.2.3. Note that in this case the eigenvalues δ_i are all equal to 1 (as expected by the asymptotic $\chi^2(p)$ result of Theorem 5.2.3) because equality $\frac{(I-a)^{-1} ss'(I-a')^{-1}}{N} = C(I-a')^{-1} + (I-a)C - C$ (Polymenis [7, Appendix 2]) becomes $\frac{ss'}{N} = C + C - C = C$, and so $C^{-1/2} \frac{(I-a)^{-1} ss'(I-a')^{-1}}{N} C^{-1/2} = C^{-1/2} C C^{-1/2} = I$; thus the regularity assumption of normality pertaining to SEM is no more needed for obtaining the distribution of T_r^2 when r is sufficiently large.Finally remark that for the one-dimensional case ($p = 1$), C is equal to $\frac{s^2}{N(1-a^2)}$ on one hand and $\frac{(I-a)^{-1} ss'(I-a')^{-1}}{N}$ is equal to $\frac{s^2}{N(1-a)^2}$ on the other hand (as aforementioned), and thus $\frac{s^2}{N(1-a^2)} = \frac{s^2}{N}$ and $\frac{s^2}{N(1-a)^2} = \frac{s^2}{N}$ when $a = 0$; it results that $\delta_1 = \left(\frac{s^2}{N}\right) / \left(\frac{s^2}{N}\right) = 1$, which corresponds to the one-dimensional version of matrix I.

4. Conclusions

In the present paper asymptotic distribution properties of a Hotelling's T^2 type of statistic concerning variates generated from the stochastic EM algorithm at its stationary state, and appearing in eq. (2.2), were investigated and were found similar to corresponding properties concerning variates from the AR(1) process represented by eq. (2.1). These results rely on a central limit theorem and on the fact that second order moments obtained at the stationary state of SEM are equal to those obtained when process $\{X_r\}$ satisfies eq. (2.1) and the number of iterations r of the SEM algorithmis sufficiently large.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd edition, Wiley, New York (1984).

[2] G. Celeux, Validity tests in cluster analysis using a probabilistic teacher algorithm, in: *Compstat*, Proceedings in Computational Statistics, 7th Symposium held in Rome 1986, F. De Antoni, N. Lauro and A. Rizzi (Editors), **86**, *Physica-Verlag* (1986), 163 – 168, DOI: 10.1007/978-3-642-46890-2.

[3] G. Celeux, *Reconnaissance de Mélanges de Densités de Probabilité et Applications à la Validation des Résultats en Classification*, PhD Thesis, University Paris 9, Dauphine (1987).

[4] G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Computational Statistics Quarterly* **2** (1985), 73 – 82.

[5] G. Celeux and J. Diebolt, Etude du comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les mélanges de lois de probabilité, *Research Report* **RR-0563**, INRIA, inria-00075991 (1986), URL: https://hal.inria.fr/inria-00075991.

[6] G. Celeux and J. Diebolt, A random imputation principle: the stochastic EM algorithm, *Research Report* **RR-0901**, INRIA, inria-00075655 (1988), URL: https://hal.inria.fr/inria-00075655.

[7] A. Polymenis, A note on a validity test using the stochastic EM algorithm in order to assess the number of components in a finite mixture model, *Statistics* **42**(3) (2008), 261 – 274, DOI: 10.1080/02331880701829922.

[8] A. Polymenis, Asymptotic results for a stationary random process generated by the stochastic EM algorithm in mixture analyses, *Journal of Statistics and Management Systems* **21**(2) (2018), 251 – 260, DOI: 10.1080/09720510.2017.1418211.

[9] M. B. Priestley, *Spectral Analysis of Time Series, Vol. I – Univariate Series*, Academic Press (1981).

[10] R. A. Redner and H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review* **26** (1984), 195 – 239, URL: https://www.jstor.org/stable/2030064.

[11] D. M. Titterington, A. F. M. Smith and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York (1985).