Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions

Jean Diebolt, Gilles Celeux

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

1992

25ème anniversaire

N° 1591

Programme 5

Traitement du Signal,

Automatique et Productique

ASYMPTOTIC PROPERTIES OF A STOCHASTIC EM ALGORITHM FOR ESTIMATING MIXING PROPORTIONS

Jean DIEBOLT

Gilles CELEUX

Février 1992

RR.1591

Asymptotic Properties of a Stochastic EM Algorithm for Estimating Mixing Proportions

Jean Diebolt*
LSTA, Université Paris 6, Pl. Jussieu 75252 Paris Cedex 05

Gilles Celeux
INRIA Rocquencourt 78153 Le Chesnay Cedex

Abstract

The purpose of this paper is to study the asymptotic behavior of the Stochastic EM algorithm (SEM) in a simple particular case within the mixture context. We consider the estimation of the mixing proportion p of a two-component mixture of densities assumed to be known. We establish that the stationary distribution of the ergodic Markov chain generated by SEM is asymptotic, as the sample size N tends to infinity, to a Gaussian distribution with mean the consistent maximum likelihood estimate of p and variance proportional to $N^{-1/2}$. *Similarly, we determine the limiting distributions of two sequential versions of SEM and study their asymptotic relative efficiency.*

Keywords: *Stochastic Algorithm; EM; Mixing Proportions; Asymptotic Behavior; Sequential Algorithms; Asymptotic Efficiency.*

*This paper was written while the first author was a Visiting Scholar at the University of Washington, Seattle; he was supported by a NATO grant, the CNRS and the ONR contract N-00014-91-J-1074.

Propriétés asymptotiques de l'algorithme SEM pour l'estimation des proportions d'un mélange de lois de probabilité

Résumé

Le but de cet article est d'étudier le comportement asymptotique de l'algorithme SEM dans un cas particulier simple d'estimation de la proportion d'un mélange de deux densités de probabilité connues. On montre que la distribution stationnaire de la chaine de Markov ergodique engendrée par l'algorithme SEM est asymptotiquement (lorsque la taille N de l'échantillon tend vers l'infini) une loi normale dont la moyenne est l'unique estimateur consistant du maximum de vraisemblance et dont la variance est proportionnelle à $N^{-1/2}$. De manière analogue, on détermine la loi limite de deux versions séquentielles de l'algorithme SEM et étudions leur efficacité asymptotique relative.

Mots-clés : *Algorithme stochastique, proportions d'un mélange, comportement asymptotique, algorithmes séquentiels, efficacité asymptotique.*

1 Introduction

The purpose of the present article is to study the asymptotic behavior of the random sequence of parameters generated by the Stochastic EM algorithm (SEM algorithm, see, e.g., Celeux and Diebolt (1985)) as the sample size $N \to \infty$, in a simple particular case within the mixture context.

The EM algorithm (Dempter, Laird and Rubin, 1977) is a widely applicable approach for computing maximum likehood (ML) estimates for incomplete data. Despite appealing features, the EM algorithm has several severe well-documented drawbacks.

In an attempt to overcome some of these drawbacks, we have defined and studied a stochastic version of the EM algorithm, that we have called the SEM algorithm, in Broniatowski, Celeux and Diebolt (1983) and Celeux and Diebolt (1985, 1986a, 1987). Instead of maximizing the expected complete-data loglikehood conditional on the observations $\mathbf{x}_{(N)} = \{x_1, \ldots, x_N\}$, the SEM algorithm first simulates the missing data $\mathbf{z}_{(N)}$ from the conditional density $\mathbf{k}(\mathbf{z}_{(N)}|\mathbf{x}_{(N)}, \theta^{(m)})$, where $\theta^{(m)}$ is the current guess of the parameter, and then computes the maximum of the pseudo-completed likelihood function, thus producing the updated estimator $\theta^{(m+1)}$. Note that the SEM algorithm can be seen as a particular case of the MCEM algorithm of Wei and Tanner (1990), with $q = 1$ in their notation, and that these authors overlooked Celeux and Diebolt's previous papers. (An answer to Wei and Tanner (1990) can be found in Biscarat, Celeux and Diebolt (1992).)

The random sequence $\{\theta^{(m)}\}$ generated by SEM is a homogeneous Markov chain which turns out to be ergodic in most of the cases of interest (see Section 2 for a proof of ergodicity in the particular mixture case under consideration). Let Ψ_N denote its stationary distribution, where the subscript N indicates dependence upon the observed sample $\mathbf{x}_{(N)} = \{x_1, \ldots, x_N\}$. The estimator of θ provided by SEM is the mean θ_N^{sem} of the distribution Ψ_N. It can be approximated by averaging over a sufficient number of $\theta^{(m)}$'s after $\theta^{(m)}$ has approximately reached its stationary regime (see Section 2).

Celeux and Diebolt (1985, 1986a) provide experimental evidence which shows that, for reasonable sample sizes, SEM is often preferable to EM. It avoids saddle-points as well as nonsignificant local maxima of the likelihood function and, in some cases, greatly accelerates the convergence. Moreover, for mixtures, it allows misspecification of the number of components since an upper bound of the number of components is sufficient to ensure convergence to the actual number of components if the sample size is large enough. Finally, in some particular cases, SEM may even provide a good alternative when the E-step evaluation of the EM algorithm is too intricate. For instance, when considering censored data it is much easier to simulate the censored data than to work with the expected complete likelihood conditional upon $\mathbf{x}_{(N)}$ (see Wei and Tanner (1990) and Chauveau (1991)).

On the other hand, Diebolt and Robert (1992) have highlighted the links between SEM and Bayesian sampling. SEM can be viewed as a simplified version of the Data Augmentation algorithm of Tanner and Wong (1987) with noninformative priors, where the step of simulation of the posterior distribution conditional upon the pseudo-completed data, $\pi(\theta|\mathbf{x}_{(N)}, \mathbf{z}^{(m)})$, is replaced by the computation of the mean of $\pi(\theta|\mathbf{x}_{(N)}, \mathbf{z}^{(m)})$. A similar parallel can be exhibited for Gibbs sampling. The interest of the SEM alternative in this perspective is that it allows for working out an estimate of θ even when the distributions under consideration are not conjugate. For instance, Chauveau (1991) makes use of SEM for this reason when dealing with mixtures of Weibull distributions.

The numerical simulation results of Celeux and Diebolt show that the stationary distributions Ψ_N of SEM is usually concentrated around a significant local maximum of the likelihood. In the present paper, we address the following basic problems :

1. Is $\theta_N^{sem} = Mean(\Psi_N)$ a consistent estimator of θ ?
2. What is the order of $\theta_N^{sem} - \theta_N$, where θ_N is the unique consistent solution of the likelihood equations (e.g., Redner and Walker, 1984) ?
3. Is the conditional distribution of $N^{1/2}(W_N - \theta_N)$ given the observed sample $\mathbf{x}_{(N)} = \{x_1, \ldots, x_N\}$ asymptotically distributed as a normal distribution with mean 0 and positive variance matrix, where W_N is a random variable drawn from Ψ_N ?

Since the theoretical results on the convergence of EM (Wu, 1983 and Redner and Walker, 1984) are essentially of a local nature and the standard asymptotic Bayesian theory cannot be used, Problems (1)-(3) appear rather formidable. Indeed, we did not find how to treat them in the general case with several local maxima as well as saddle-points. This is the reason why we focused on a particular case where the likelihood function (l.f.) is concave (see Section 2). Of course, in such a case, EM gives good results, and from a practical point of view, SEM is not useful. However, results obtained in this particular case have their own theoretical interest and permit us to gain insight into what is really happening and the mathematics beyond Problems (1)-(3). Moreover, these results suggest what answers can be expected in more general situations.

In Section 2, we present the simple mixture model that we will consider throughout the paper and derive preliminary results about the l.f., EM and SEM in this particular case.

Section 3 is devoted to our main result, stated as Theorem 1. This theorem gives affirmative answers to Problems (1) and (3) for the model under consideration. It also provides a (non-optimal) estimate of the rate of $\theta_N^{sem} - \theta_N$ (Problem (2)), as well as an estimate of the rate of the conditional variance of θ_N^{sem} given $\mathbf{x}_{(N)}$. Furthermore, the results in Theorem 1 imply that θ_N^{sem} is an asymptotically unbiased and optimal estimator of θ and the stationary rescaled SEM sequence $Y^{(m)} = N^{1/2}(\theta^{(m)} - \theta_N)$ converges in distribution, as $N \to \infty$, to the distribution of the stationary autoregressive sequence $\{Z_\star^{(m)}\}$ defined by

$$Z_\star^{(m+1)} = r^\star Z_\star^{(m)} + \sigma^\star \epsilon^{(m)}, \tag{1.1}$$

where the $\epsilon^{(m)}$'s are Gaussian i.i.d. random variables with mean 0 and variance 1, $\epsilon^{(m)}$ is dependent of $Z_\star^{(o)}, \ldots, Z_\star^{(m)}$, and $r^\star$, $0 < r^\star < 1$, and $\sigma^\star, \sigma^\star > 0$, are defined in (2.8) and (2.19) in terms of the complete, conditional and observed Fisher information values, respectively.

Section 4 examines two different sequential versions of SEM. The "*one-step*" version has been implicitly studied in Silverman (1980), but has its asymptotic efficiency can equal to zero. Our Theorem 3 states the a.s. convergence of the "*global*" version and its asymptotic normality. Since the asymptotic variance can be made explicit, we can examine in detail its asymptotic efficiency, which turns out to be of the same order as the optimal bound.

2 Preliminary results

2.1 The mixture problem

Throughout this paper, the observed data $\mathbf{x}_{(N)} = \{x_1, \cdots, x_N\}$ will be realizations of i.i.d. random variables from the mixture density $h(x, p^*)$, where

$$h(x,p) = pf_1(x) + (1-p)f_2(x), \tag{2.1}$$

where $f_1(x)$ and $f_2(x)$ are known densities with respect to a σ-finite measure $\mu(dx)$ on some separable measurable space $\mathbf{E}$, and the parameter p satisfies $0 < p < 1$. We will assume that $\mu\{x : f_1(x) \neq f_2(x)\} \neq 0$ and that $f_1(x)$ and $f_2(x)$ are positive on their respective supports. The statistical problem under consideration is to find a good estimate of p^* on the basis of $\mathbf{x}_{(N)}$.

Before proceeding, a formal point has to be made, since the study of the asymptotic behavior as the sample size $N \to \infty$ of a stochastic algorithm involves two different probability spaces: The sample space and the sample of pseudorandom drawings. We will interpret each sample $\mathbf{x}_{(N)}$ of size N as the projection on the N first coordinates of a sequence $\mathbf{x} = \{x_i; i \geq 1\}$ drawn from the product space $\mathbf{X} = \mathbf{E}^{\{i:i\geq 1\}}$ endowed with the probability distribution

$$\mathsf{P}_{\mathbf{X}} = \prod_{i=1}^{\infty} h(x_i, p^*)dx_i. \tag{2.2}$$

The formal description of the pseudorandom drawings is postponed to Subsection 2.3.

Next, let us describe the underlying complete data structure of the statistical problem under consideration. The complete data is $(\mathbf{x}_{(N)}, \mathbf{z}_{(N)}) = \{(x_i, z_i); i = 1, \ldots, N\}$, where $z_i = 1$ or 0 according as x_i has been drawn from $f_1(x)$ or $f_2(x)$, and the z_i's are independent. Thus, each z_i is a Bernoulli r.v. with parameter $t_i^* = t(x_i, p^*)$, where

$$t(x,p) = \frac{pf_1(x)}{pf_1(x) + (1-p)f_2(x)}. \tag{2.3}$$

2.2 The EM algorithm

The mth iteration $p^{(m+1)} = T_N(p^{(m)})$ of EM consists in the E-step: Compute $t_i^{(m)} = t(x_i, p^{(m)})$ for $i = 1, \ldots, N$, followed by the M-step: Compute $p^{(m+1)} =$

$T_N(p^{(m)})$, where

$$T_N(p) = \frac{1}{N}\sum_{i=1}^{N} t(x_i, p) \quad \text{for } p \in (0,1). \tag{2.4}$$

Thus, letting $T_N(0) = 0$ and $T_N(1) = 1$, the EM algorithm indeed consists in iterating the function $T_N : [0,1] \to [0,1]$, starting from an initial position $p^{(o)} \in (0,1)$. We have the following preliminary results, where

$$L_N(p) = \sum_{i=1}^{N} \log h(x_i, p) \tag{2.5}$$

denotes the loglikelihood function and the observed, complete and conditional Fisher information values J_{obs}, J_c and J_{cond}, respectively, are defined as in Titterington, Smith and Makov (1985).

Lemma 2.1 *(i) The function $T_N(p)$ is increasing over $[0,1]$.*

(ii) We have, for all p in $[0,1]$,

$$T_N(p) - p = p(1-p)\frac{L'_N(p)}{N}. \tag{2.6}$$

(iii) For $\mathbf{P_X}$-almost every $\mathbf{x} \in \mathbf{X}$, there exists an integer $N_0 = N_0(\mathbf{x})$ such that $L''_N(p) < 0$ for all p in $(0,1)$ whenever $N \geq N_0$, i.e. the loglikelihood is a concave function for $N \geq N_0$.

(iv) If p_N is the unique maximizer of $L_N(p)$ when $N \geq N_0$, p_N is the unique stable fixed point of $T_N(p)$ over $[0,1]$, with

$$r_N = T'_N(p_N) = 1 + p_N(1-p_N)\frac{L''_N(p_N)}{N} \in (0,1), \tag{2.7}$$

and $T_N(p) > p$ for $0 < p < p_N$ and $T_N(p) < p$ for $p_N < p < 1$.

(v) Each sequence $\{p^{(m)}\}$ generated by EM starting from $p^{(0)} \in (0,1)$ converges to p_N with a geometric rate.

(vi) The derivative r_N of T_N at p_N converges $\mathbf{P_X}$-a.s. to

$$r^* = 1 - \frac{J_{obs}}{J_c} = \frac{J_{cond}}{J_c} \in (0,1) \tag{2.8}$$

as $N \to \infty$.

A brief proof of Lemma 2.1 can be found in the Appendix.

Remark 2.1 - *Lemma 2.1 shows that, in the simple incomplete data statistical problem under consideration, EM does very well if $N \geq N_0$. Thus, from a practical point of view, there is no need for any improved algorithm in this particular case. However, we have explained in Section 1 why the asymptotic behavior of SEM as $N \to \infty$ in this context deserves a careful study.*

2.3 The SEM algorithm

In the present context, the Stochastic Imputation Principle (e.g., Celeux and Diebolt, 1987) produces the updated estimate $p^{(m+1)}$ as the ML estimate based on the pseudo-completed sample $(\mathbf{x}_{(N)}, \mathbf{z}^{(m)}) = \{(x_i, z_i^{(m)}); i = 1, \ldots, N\}$, where $\mathbf{z}^{(m)} \in \mathbf{Z}_{(N)} = \{0,1\}^N$, each $z_i^{(m)}$ is a Bernoulli r.v. with parameter $t_i^{(m)} = t(x_i, p^{(m)})$ given by (2.3) and the $z_i^{(m)}$'s, $i = 1, \ldots, N$, are drawn independently. This yields, in view of (2.4),

$$p^{(m+1)} = \frac{1}{N} \sum_{i=1}^{N} z_i^{(m)}, \tag{2.9}$$

so that the random sequence $\{p^{(m)}\}$ is a homogeneous Markov chain taking its values in $\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$. In order to remove the absorbing states 0 and 1, we first make choice of a sequence of thresholds $c(N)$, $\frac{1}{N} \leq c(N) < 1 - c(N) \leq \frac{N-1}{N}$, such that $c(N) \to 0$ as $N \to \infty$, and of a probability distribution Γ_N on the set $\{\frac{j}{N} : j = 0, 1, \cdots, N \;\text{ and }\; c(N) \leq \frac{j}{N} \leq 1 - c(N)\}$. The SEM algorithm then proceeds as follows. E-step: Compute $t(x_i, p^{(m)}) = t_i^{(m)}$ for $i = 1, \cdots, N$ using (2.3). S-step: For $i = 1, \cdots, N$, draw independently the Bernoulli r.v.'s $z_i^{(m)}$ with parameter $t_i^{(m)}$ and compute

$$p^{(m+\frac{1}{2})} = \frac{1}{N} \sum_{i=1}^{N} z_i^{(m)}. \tag{2.10}$$

If $p^{(m+\frac{1}{2})} \in J_N = [c(N), 1 - c(N)]$, then go to M-step. Otherwise, draw $p^{(m+1)}$ from the preassigned distribution Γ_N and go to E-step. M-step:

$$p^{(m+1)} = p^{(m+\frac{1}{2})}. \tag{2.11}$$

This procedure avoids $p^{(m)}$ being stuck at $p = 0$ or $p = 1$, whereas the sequence defined in this way is still a homogeneous Markov chain. Next, we turn to the ergodicity of this Markov chain.

Lemma 2.2 *The homogeneous Markov chain $\{p^{(m)}\}$ generated by SEM is geometrically ergodic and the support of its stationary distribution Ψ_N is contained in J_N.*

The proof of Lemma 2.2 can be found in the Appendix.

Remark 2.2 - *As a consequence, since $\{p^{(m)}\}$ is a finite-state ergodic Markov chain, it is uniformly strongly mixing with a geometric rate. Hence, the SLLN and a suitable version of the CLT (e.g., Davydov, 1973) apply.*

We conclude this section by showing that the sequence generated by SEM can be viewed as a random perturbation of the discrete-time dynamical system on $[0,1]$ generated by EM. First, we need to have a workable representation of the r.v.'s $z_i^{(m)}$. They can be written as

$$z_i^{(m)} = \mathbf{1}_{[0,t(x_i,p^{(m)})]}(\omega_i^{(m)}), i = 1,\ldots,N, \tag{2.12}$$

where $\mathbf{1}_{[a,b]}(s)$ is the indicator function of the interval $[a,b]$ and the ω_i's, $i = 1,\ldots,N$, are i.i.d. random variables uniformly distribued on $[0,1]$ such that the sample $\omega^{(m)} = (\omega_1^{(m)},\ldots,\omega_N^{(m)})$ is independent of $p^{(0)},\ldots,p^{(m)}$. We have

$$p^{(m+\frac{1}{2})} = T_N(p^{(m)}) + U_N(p^{(m)}), \tag{2.13}$$

where for each p, $0 < p < 1$,

$$U_N(p) = N^{-1/2}S_N(p)\eta_N(p,\omega), \tag{2.14}$$

where $S_N(p) > 0, S_N^2(p) = p(1-p)T_N'(p)$ converges $\mathbf{P_X}$-a.s. as $N \to \infty$ to

$$S^2(p) = p(1-p)\int f_1(x)f_2(x)\frac{h(x,p^*)}{h(x,p)}\mu(dx), \tag{2.15}$$

and the r.v. $\omega \to \eta_N(p,\omega)$ defined by

$$\eta_N(p,\omega) = N^{-1/2}S_N^{-1}(p)\sum_{i=1}^{N}\{\mathbf{1}_{[0,t(x_i,p)]}(\omega_i) - t(x_i,p)\}, \tag{2.16}$$

has mean 0 and variance 1 for each p in $(0,1)$. With the above representation of the $z_i^{(m)}$'s, the probabilistic setup of the successive random drawings involved in the S-step of SEM can be made precise: each $\omega^{(m)}$ can be viewed as a whole sequence of $\mathbf{\Omega} = [0,1]^{\{i:i\geq 1\}}$ endowed with the product σ-field and the probability $\mathbf{P}_{\mathbf{\Omega}}(d\omega) = \Pi_{i\geq 1}\lambda(d\omega_i)$, where λ denotes the Lebesgue measure on $[0,1]$, whereas $\eta_N(p,\omega^{(m)})$ only involves the first N coordinates $\omega_1^{(m)},\ldots,\omega_N^{(m)}$ of the sequence $\omega^{(m)}$. We will denote by $E_{\mathbf{\Omega}}(Y)$ the expectation of any r.v. $Y(\omega)$ with respect to this probability $\mathbf{P}_{\mathbf{\Omega}}$. For instance, $E_{\mathbf{\Omega}}(\eta_N(p,\omega)) = 0$ and $E_{\mathbf{\Omega}}\{\eta_N^2(p,\omega)\} = 1$ for all p in $(0,1)$.

Since the CLT implies that, for all p in $(0,1)$ and $\mathbf{P}_X$-a.e. $\mathbf{x}$, $\eta_N(p,\omega)$ converges in $\mathbf{P}_{\mathbf{\Omega}}$-distribution as $N \to \infty$ to a Gaussian r.v. $\epsilon(\omega)$ with mean 0 and variance 1, we can expect that, for large N, (2.10)-(2.14) can be approximated by

$$p^{(m+\frac{1}{2})} \approx T_N(p^{(m)}) + N^{-1/2}S_N(p^{(m)})\epsilon^{(m)}, \tag{2.17}$$

with $\epsilon^{(m)} = \epsilon(\omega^{(m)})$, so that, if we can show that the stationary measure Ψ_N of $\{p^{(m)}\}$ is well concentrated around p_N, then (2.17) turns out to be approximately

$$p^{(m+\frac{1}{2})} \approx p_N + r_N(p^{(m)} - p_N) + N^{-1/2}S_N(p^{(m)})\epsilon^{(m)}. \tag{2.18}$$

Furthermore, since $r_N \to r^*$ and

$$\sigma_N^2 = S_N^2(p_N) \to \sigma^{*2} = S^2(p^*) = p^*(1-p^*)r^* = J_{cond}/J_c^2 \tag{2.19}$$

as $N \to \infty$,

$$p^{(m+\frac{1}{2})} - p_N \approx r^*(p^{(m)} - p_N) + N^{-1/2}\sigma^*\epsilon^{(m)}, \tag{2.20}$$

so that, in general, $p^{(m+1)} = p^{(m+\frac{1}{2})}$ if $p^{(m)}$ remains near p_N most of the time. If we can make the approximations (2.17)-(2.20) precise and uniform with respect to $p^{(m)}$ in a suitable sense and show that $p^{(m)}$ remains near p_N, then we will have essentially proved Theorem 1, which we will now state and prove.

3 Main Results

3.1 Theorem 1

Before stating our Theorem 1, we need to introduce

$$R(N) = R(\mathbf{x}, N) = \sup_{p \in J_N, p \neq p_N} \left| \frac{T_N(p) - p_N}{p - p_N} \right|. \tag{3.1}$$

It follows from the results in Subsection 2.2 and the Appendix that $0 < R(N) < 1$ $\mathbf{P_X} - a.s.$ for N large enough, and $1 - R(N) \to 0$ as $c(N) \to 0$. Furthermore, the rate of convergence to 0 of $1 - R(N)$ can be arbitrarily slow provided that the rate of $c(N)$ as been chosen slow enough.

Theorem 1 *Suppose that the following assumptions (H1)-(H4) hold.*
(H1) The densities $f_1(x)$ and $f_2(x)$ satisfy $\mu\{x : f_1(x) \neq f_2(x)\} \neq 0$.
(H2) The probability distribution Γ_N (see Subsection 2.3) used to draw the updated SEM estimator $p^{(m+1)}$ when $p^{(m+(1/2))}$ is not in J_N is the Dirac measure at some $j(N)/N \in (c(N), 1 - c(N))$, $1 \leq j(N) \leq N - 1$.
(H3) $Nc(N) \to \infty$ as $N \to \infty$.
(H4) $N\{1 - R(N)\}^4 \to \infty$ as $N \to \infty$.
Then

(i) If W_N is a r.v. from the stationary distribution Ψ_N of SEM, then $N^{1/2}(W_N - p_N)$ converges in distribution as $N \to \infty$ to a Gaussian r.v. with mean 0 and variance $v^* = \sigma^{*2}/(1 - r^{*2})$, where $r^* = J_{cond}/J_c \in (0,1)$ and $\sigma^{*2} = p^*(1 - p^*)r^* = J_{cond}/J_c^2$.

(ii) For all N large enough,

$$|p_N^{sem} - p_N| \leq N^{-1/2}a(N) + O\left(\frac{\alpha^2(N)}{N}\right), \tag{3.2}$$

where $p_N^{sem} = Mean(\Psi_N)$ and

$$a(N) = O(\alpha(N) + \delta(N)), \tag{3.3}$$

$$\alpha(N) = O\left(N^{-1/2}\{1 - R(N)\}^{-2}\right) + O\left(\{Nc(N)\}^{-1/16} log^{1/4}\left\{\frac{1}{Nc(N)}\right\}\right) \tag{3.4}$$

and

$$\delta(N) = O\left(|r_N - r^*| + |\sigma_N - \sigma^*|\right). \tag{3.5}$$

Furthermore, if $a(N)\{1 - R(N)\}^{-1} \to 0$ *as* $N \to \infty$, *then*

$$|Var(\Psi_N) - \frac{v^*}{N}| = o(\frac{1}{N}). \tag{3.6}$$

(iii) If $P_N(t)$, $-\infty < t < \infty$, *denotes the d.f. of* Ψ_N *and* Φ_N *denotes the normal d.f. with mean* p_N *and variance* $v_N/N = \sigma_N^2/\{N(1-r_N^2)\}$, *then, for all* τ_0 *such that* $0 \le p^* - \tau_0 < p^* + \tau_0 \le 1$,

$$\sup_{t \notin [p^*-\tau_0, p^*+\tau_0]} |P_N(t) - \Phi_N(t)| = O\left(\frac{\alpha^2(N)}{N}\right) \tag{3.7}$$

$$\sup_{t \in [p^*-\tau_0, p^*+\tau_0]} |P_N(t) - \Phi_N(t)| = O\left(\alpha^{2/3}(N)\right). \tag{3.8}$$

Remark 3.1 *The assumption (H2) can be greatly relaxed to allow for more general* Γ_N *distributions. It suffices to make proper choices of* $\tilde{S}_N(p)$ *and* $\xi_N(p,\omega)$ *for* $p \notin J_N$, *but for these more general* Γ_N*'s the proof of Theorem 1 is more involved. Here, we have stated Theorem 1 under (H2) for clarity.*

Remark 3.2 *The assertion (i) tells us that the stationary distribution* Ψ_N *of SEM in the particular mixture context detailed in Section 2 is asymptotically normal with mean* p_N *and variance* v^*/N, *where* $v^* = J_{obs}^{-1}\{1+(J_c/J_{cond})\}^{-1} < 1/(2J_{obs})$. *Thus, the variance of a sample* $\{p^{(m)} : m = 1, \cdots, M\}$ *of the stationary SEM sequence is only a fraction of the variance* $1/J_{obs}$ *of the ML estimator* p_N. *This is natural since, as explained in Section 1, SEM can be roughly viewed as a particular version of the Gibbs sampler, where the step of simulation of* $\theta^{m+1} \sim \pi(\theta|\mathbf{x}_{(N)}, \mathbf{z}^{(m)})$ *is replaced by the updating* $\theta^{(m+1)} =$ *Mean of* $\pi(\theta|\mathbf{x}_{(N)}, \mathbf{z}^{(m)})$, *which reduces the variance of the generated sequence.*

Remark 3.3 *The assertion (ii) entails that the SEM estimator* $p_N^{sem} =$ *Mean* (Ψ_N) *is asymptotically unbiased and its sample variance is equal to the sample variance of the ML estimator* p_N *up to a term of the order of* $a(N)/N = o(1/N)$. *Thus,* p_N^{sem} *is asymptotically optimal.*

Remark 3.4 *From Lemma A.4 in the Appendix, it results that* $\delta(N) = O(\ell(N)/\sqrt{N})$ $P_{\mathbf{X}}-a.s.$*, where* $\ell(N) = \sqrt{2\ell_2(N)}$ *and* ℓ_2 *denotes the iterated logarithm.*

Remark 3.5 *Theorem 1 (i) can be directly generalized to the case where the mixture* $h(x, \mathbf{p}) = \Sigma_{1\leq\ k\leq K} p_k f_k(x)$ *has* $K \geq 3$ *components and the parameter to be estimated is* $\mathbf{p} = (p_1, \cdots, p_{K-1})$*. In contrast, the assertions(ii) and (iii) do not seem to be easily extendable via our method of proof.*

Remark 3.6 *Theorem 1 suggests that similar results hold in the context of general mixture problems, and even in the general missing or incomplete data context under reasonable assumptions. The only general result in this perspective is a theorem in Celeux and Diebolt (1986b) which states essentially that the assertion (i) holds under the restrictive assumption that the EM operator* $T_N(\theta)$ *has only one fixed point in the compact* G_N *corresponding to the interval* J_N*, and that this unique fixed point is stable. This theorem supports the conjecture that (i) holds in a rather general context, since, although* $T_N(\theta)$ *has many fixed points whenever* G_N *is reasonably large, the unique consistent estimator* θ_N *becomes prominent, whereas the other fixed points of* $T_N(\theta)$ *fluctuate and fade away as* $N \to \infty$*. In a very loose sense, this means that* $T_N(\theta)$ *has asymptotically a unique fixed point in* G_N*, which turns out to be stable since it is a maximum of the l.f.*

Remark 3.7 *Note that an alternative to the SEM algorithm is the SAEM algorithm (Celeux and Diebolt, 1992), which is somewhat in the spirit of simulated annealing. Celeux and Diebolt (1992) show that, for any given sample* $\mathbf{x}_{(N)}$ *with N large enough, SAEM converges a.s. to a local maximum of the l.f. in the context of general mixtures of densities from some exponential family, under reasonable assumptions concerning the fixed points of* $T_N(\theta)$ *in* G_N*. Furthermore, Biscarat (1992) establishes a more general result, which allows to take care of other important incomplete data settings. See also Biscarat, Celeux and Diebolt (1992) for a similar theoretical study of a simulated annealing type version of the MCEM algorithm introduced in Wei and Tanner (1990).*

3.2 Proof of Theorem 1

We begin with a brief outline of the proof of Theorem 1. In Part I, we establish results analogous to (i)-(iii) for the auxiliary sequence

$$V^{(m)} = \sqrt{N}\left(q^{(m)} - p_N\right), \tag{3.9}$$

where the homogeneous Markov chain $\left\{q^{(m)}\right\}$ is recursively defined by

$$q^{(m+1)} = \tilde{T}_N\left(q^{(m)}\right) + \frac{\tilde{S}_N\left(q^{(m)}\right)}{\sqrt{N}}\xi_N\left(q^{(m)}, \omega^{(m)}\right), \tag{3.10}$$

where $\tilde{T}_N(p) = T_N(p)$ for $p \in J_N$ and $\tilde{T}_N(p) =$ some $\dfrac{j(N)}{N}$ in J_N for $p \notin J_N$, $\tilde{S}_N(p) = S_N(p)$ for $p \in J_N$ and $S_N(p) = 0$ for $p \notin J_N$, and $\xi_N(p,\omega) = \eta_N(p,\omega)$ for $p \in J_N$, $\xi_N(p,\omega) = \eta_N(c(N),\omega)$ for $0 \leq p \leq c(N)$ and $\xi_N(p,\omega) = \eta_N(1-c(N),\omega)$ for $1 - c(N) \leq p \leq 1$. We first show that $\left\{q^{(m)}\right\}$ is ergodic (Step 1) with stationary distribution denoted by Λ_N. Then, we derive an upper bound for $E_\Omega\left(\left|V^{(m)}\right|^4\right)$, introducing

$$0 < R(N) = \sup_{p \in J_N, p \neq p_N} \left|\frac{T_N(p) - p_N}{p - p_N}\right| < 1, \tag{3.11}$$

in Step 2. In Step 3, we deduce from technical results about $\tilde{T}_N(p)$ and $\tilde{S}_N(p)$ (Lemmas 3.3 and 3.4) and from the Skorohod representation together with bounds related to the Berry-Esseen Inequality (see Lemma 3.5) an upper bound for $E\left(\left|V^{(m)} - Z^{(m)}\right|^2\right)$, where $Z^{(0)} = V^{(0)}$ and

$$Z^{(m+1)} = r_N Z^{(m)} + \sigma_N \varepsilon^{(m)}, \tag{3.12}$$

with $\varepsilon^{(m)}$ a Gaussian r.v. with mean 0 and variance 1, independent from $Z^{(0)}, \ldots, Z^{(m)}$. In Step 4, we deduce from Lemma 3.5 results for $V^{(m)}$ analogous to (i)-(iii) and an upper bound for $\Lambda_N(J_N^c)$, where J_N^c denotes the complement of J_N in $[0,1]$.

In Part II, we show how to obtain from these results corresponding upper bounds for $E_\Omega\left(\left|Y^{(m)} - Z^{(m)}\right|^2\right)$, where

$$Y^{(m)} = \sqrt{N}\left(p^{(m)} - p_N\right). \tag{3.13}$$

PART I

Step 1 First, we have to make sure that $\left\{q^{(m)}\right\}$ is ergodic. This is the purpose of Lemma 3.1 below.

Lemma 3.1 *The homogeneous Markov chain $\left\{q^{(m)}\right\}$ defined by (3.10) is ergodic. Moreover, its stationary distribution Λ_N has all its moments finite if $Nc(N) \to \infty$ as $N \to \infty$.*

The proof of Lemma 3.1 parallels that of Lemma 2.2 above, and can be found in the Appendix.

Step 2 We need the following upper bound for $E_{\mathbf{\Omega}}\left(\left|V^{(m)}\right|^4\right)$ involving $R(N)$ defined by (3.11).

Lemma 3.2 *Assume that $Nc(N) \to \infty$ as $N \to \infty$, and either $V^{(0)} = 0$ or $\left\{V^{(m)}\right\}$ is in its stationary regime. Then, for $\mathbf{P_X}$ - a.e. $\mathbf{x}$, there exists a finite integer $N_1(\mathbf{x})$ such that $N \geq N_1(\mathbf{x})$ implies*

$$\left\|V^{(m)}\right\|_4 \leq \frac{1}{1-R(N)} \text{ for all } m \geq 0, \tag{3.14}$$

where $\|V\|_p = (E_{\mathbf{\Omega}}(|V|)^p)^{1/p}$ for $p \geq 1$.

Proof. From (3.9)-(3.10) and the Minkowski Inequality, it follows that

$$\left\|V^{(m+1)}\right\|_4 \leq R(N)\left\|V^{(m)}\right\|_4 + \frac{1}{4}\left\|\xi_N\left(q^{(m)},\omega^{(m)}\right)\right\|_4, \tag{3.15}$$

since $0 < S_N(p) \leq 1/4$ for all p in $(0,1)$. Now, the same calculation as in the proof of Lemma 3.1 (Appendix) shows that, for all p in $[0,1]$,

$$\begin{aligned} E_{\mathbf{\Omega}}(|\xi_N(p,\omega)|^4) &\leq 1 + \frac{1}{4N\tilde{S}_N^2(p)} \\ &\leq 1 + \frac{1}{2Nc(N)r'_N}, \end{aligned} \tag{3.16}$$

where $r'_N = \inf_{[0,1]} |T'_N(p)| \to r' > 0$ for almost every $\mathbf{x}$ as $N \to \infty$ (Appendix). Since $Nc(N) \to \infty$ as $N \to \infty$, the result follows. □

Step 3 Before establishing Lemma 3.5, which is the core of the proof, we need two additional technical results, namely Lemmas 3.3 and 3.4 below, whose proofs are postponed to the Appendix for clarity.

Lemma 3.3 *(i) For* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$*, there exists a finite integer* $N_2(\mathbf{x}) \geq N_1(\mathbf{x})$ *such that* $N \geq N_2(\mathbf{x})$ *implies*

$$|T''_N(p)| \leq \frac{1}{p(1-p)} \text{ for all } p \text{ in } (0,1). \tag{3.17}$$

(ii) Let $0 < \varepsilon_0 < p^*$ *be given. Then, for* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$*, there exists a finite integer* $N_3(\mathbf{x}) \geq N_2(\mathbf{x})$ *such that* $N \geq N_3(\mathbf{x})$ *implies that* $|p_N - p^*| \leq \varepsilon_0$ *and* $[p^* - \varepsilon_0,\ p^* + \varepsilon_0]$ *is contained in* J_N *and, for all* h *such that* $p_N + h \in [0,1]$*, we have*

$$|\tilde{T}_N(p_N + h) - p_N - hr_N| \leq A_0|h|^2 \tag{3.18}$$

for some positive constant A_0.

Lemma 3.4 *Under the same assumptions as in Lemma 3.3, there exists a positive constant* B_0 *and, for* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$*, a finite integer* $N_4(\mathbf{x}) \geq N_3(\mathbf{x})$ *such that* $N \geq N_4(\mathbf{x})$ *implies*

$$|\tilde{S}_N(p_N + h) - \tilde{S}_N(p_N)| \leq B_0|h| \tag{3.19}$$

for all h *such that* $p_N + h \in [0,1]$.

The next lemma is the core of the proof of Theorem 1. Using Skorohod's (1956) representation argument and the Berry-Esseen Inequality, it provides a basic upper bound for $E_\lambda(|\xi_N(p,u) - \varepsilon(u)|^2)$ uniformly in $p \in [0,1]$, where $\xi_N(p,u)$ and $\varepsilon(u)$ denote the Skorohod representations of $\xi_N(p,\omega)$ and $\varepsilon(\omega)$, respectively, as defined below.

Lemma 3.5 *Assume that* $Nc(N) \to \infty$ *as* $N \to \infty$*. Then, there exists a probability space* $(\mathbf{U} = \{\mathbf{u} = (u_1, u_2, \cdots,)\}; \mathbf{P_U})$ *and r.v.'s* $\xi_N(p, u_i)$ *and* $\varepsilon(u_i)$*,* $i = 1, 2, \ldots$*, defined on this probability space, such that :*
(i) $\xi_N(p,u_i)$ *and* $\varepsilon(u_i)$ *have the same distributions as* $\xi_N(p,\omega)$ *and* $\varepsilon(\omega)$*, respectively, for each* $i = 1, 2, \cdots$ *and all* p *in* $[0,1]$.
(ii) For each fixed N *and* p*, the r.v.'s* $\xi_N(p,u_i)$*,* $i = 1, 2, \ldots$*, are i.i.d. and the r.v.'s* $\varepsilon(u_i)$*,* $i = 1, 2, \ldots$*, are i.i.d. and Gaussian with mean 0 and variance*

1.

(iii) For $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$*, there exists a finite integer* $N_5(\mathbf{x}) \geq N_4(\mathbf{x})$ *such that* $N \geq N_5(\mathbf{x})$ *implies, for all* $i = 1, 2, \ldots,$ *and* p *in* $[0,1]$,

$$\mathbf{E_U}(|\xi_N(p,u_i) - \varepsilon(u_i)|^2) \leq 10\gamma^{1/4}(N) \log^{1/2}\left(\frac{1}{\gamma(N)}\right) + 2\gamma^{1/2}(N), \quad (3.20)$$

where$\gamma(N) = O((Nc(N))^{-1/2})$.

Proof. We begin with some notation. For any distribution function (d.f.) $F(t)$, $-\infty < t < \infty$, let $F^{-1}(u) = \inf\{t : F(t) > u\}$, $u \in (0,1)$, denote the corresponding inverse function. Let $F_N(p,t)$ denote the d.f. of $\xi_N(p,\omega)$ and $\Phi(t)$ denote the standard normal d.f. The function $\xi_N(p,u) = \inf\{t : F_N(p,t) > u\}$ and $\varepsilon(u) = \Phi^{-1}(u)$, $u \in (0,1)$ are r.v.'s on the probability space $([0,1],\ B[0,1], \lambda(du))$, where $B[0,1]$ is the Borel σ- field of $[0,1]$ and $\lambda(du)$ is the Lebesgue probability measure on $[0,1]$, with the same distributions as $\xi_N(p,\omega)$ and $\varepsilon(\omega)$, respectively. For any fixed p, the CLT implies that $\xi_N(p,u) \to \varepsilon(u)\ \lambda(du)$ - a.s. as $N \to \infty$.

On the other hand, the Berry-Esseen Inequality (e.g., Shorack and Wellner (1986, p. 848)) implies that, for all p in $[0,1]$,

$$\begin{aligned} \sup_{-\infty<t<\infty} |F_N(p,t) - \Phi(t)| &\leq \frac{C_{BE}}{2}[N\tilde{S}_N^2(p)]^{-1/2} \\ &\leq \gamma(N), \end{aligned} \quad (3.21)$$

where $\gamma(N) = C_{BE}[r'Nc(N)]^{-1/2}$, with r' as in the Appendix, if N is large enough. Here, C_{BE} denotes the absolute positive constant involved in the Berry-Esseen Inequality.

Now let $B(N,p)$, $p \in [0,1]$, be the subset of $[0,1]$ defined by $B(N,p) = \{u \in (0,1) : |\xi_N(p,u) - \varepsilon(u)| > \sqrt{\gamma(N)}\}$. Owing to Shorack and Wellner (1986), Ex. 7 p. 65, we deduce from (3.21) that $\lambda\{B(N,p)\} < \sqrt{\gamma(N)}$, uniformly in p.

We are now in a position to derive (3.20). First note that, since $E_\lambda(\varepsilon^2) = E_\lambda\{\xi_N(p,.)^2\} = 1$, we have

$$E_\lambda(|\xi_N(p,.) - \varepsilon|^2) \leq 2E_\lambda\{|\varepsilon|\ |\varepsilon - \xi_N(p,.)|\}. \quad (3.22)$$

But, by the Cauchy-Schwarz Inequality, and the definition of $\varepsilon(u)$,

$$\int_{B(N,p)} |\varepsilon(u)|\ |\varepsilon(u)-\xi_N(p,u)|du \leq 2\left(\int_{B(N,p)}\{\Phi^{-1}(u)\}^2du\right)^{1/2} \tag{3.23}$$

whereas

$$\int_{B(N,p)^c} |\varepsilon(u)|\ |\varepsilon(u)-\xi_N(p,u)|du \leq \sqrt{\gamma(N)}. \tag{3.24}$$

In order to obtain a workable upper bound for the RHS of (3.23), we note that, since $\Phi^{-1}(u)$ is increasing on $(0,1)$ and symmetric about $1/2$, the integral $\int_{B(N,p)}\{\Phi^{-1}(u)\}^2du$ is lesser than the integral $\int_{C(N,p)}\{\Phi^{-1}(u)\}^2du$, where $C(N,p)$ has the same Lebesgue measure as $B(N,p)$ and is the union of $(0,a]$ and $[1-a,1)$ for some a, $0<a<1/2$. Since $\lambda\{B(N,p)\}<\sqrt{\gamma(N)}$, it follows that

$$\begin{aligned}\int_{B(N,p)}\{\Phi^{-1}(u)\}^2du &< 2\int_{1-\frac{\sqrt{\gamma(N)}}{2}}^{1}\{\Phi^{-1}(u)\}^2du \\ &= 2\int_{b(N)}^{\infty} t^2\varphi(t)dt,\end{aligned} \tag{3.25}$$

where $b(N)=\Phi^{-1}\left(1-\dfrac{\sqrt{\gamma(N)}}{2}\right)$ and $\varphi(t)=(2\pi)^{-1/2}\exp\left(-\dfrac{t^2}{2}\right)$. An integration by parts shows that the RHS of (3.25) is equal to

$$2\left\{(2\pi)^{-1/2}b(N)\exp\left(-\frac{1}{2}b^2(N)\right)+(1-\Phi)(b(N))\right\}.$$

Thus, for $b(N)\geq 1$,

$$\begin{aligned}\int_{B(N,p)}\{\Phi^{-1}(u)\}^2du &< 2b^2(N)\frac{\varphi(b(N))}{b(N)}+\sqrt{\gamma(N)} \\ &< 2b^2(N)\left(1+\frac{1}{b^2(N)}\right)(1-\Phi)(b(N))+\sqrt{\gamma(N)} \\ &\leq 4b^2(N)(1-\Phi)(b(N))+\sqrt{\gamma(N)} \\ &< (2b^2(N)+1)\sqrt{\gamma(N)}.\end{aligned} \tag{3.26}$$

Finally, for $b(N) \geq 1$, $\sqrt{\gamma(N)} = 2(1-\Phi)(b(N)) < 2\frac{\varphi(b(N))}{b(N)}$, implying that

$$\log\left(\frac{1}{\gamma(N)}\right) > b^2(N) + 2\log(b(N)) + \log\left(\frac{\pi}{2}\right) > b^2(N), \tag{3.27}$$

from which it follows that

$$\int_{B(N,p)} \{\Phi^{-1}(u)\}^2 du < \left\{2\log\left(\frac{1}{\gamma(N)}\right) + 1\right\} \sqrt{\gamma(N)}. \tag{3.28}$$

Putting (3.21)-(3.28) together, we obtain for all N large enough that

$$E_\lambda(|\xi_N(p,.) - \varepsilon|^2) \leq 10\gamma^{1/4}(N)\log^{1/2}\left(\frac{1}{\gamma(N)}\right) + 2\gamma^{1/2}(N). \tag{3.29}$$

□

Thus, to conclude the proof of Lemma 3.5, it suffices to take

$$\mathbf{U} = [0,1]^{\mathbf{N}^*}, \quad P_{\mathbf{U}}(d\mathbf{u}) = \prod_{i=1}^{\infty} \lambda(du_i),$$

$\xi_N(p, u_i) = \inf\{t : F_N(p,t) > u_i\}$ and $\epsilon(u_i) = \Phi^{-1}(u_i)$.

Step 4 We now compare the Skorohod version $V^{(m)}(\mathbf{u}) = \sqrt{N}(q^{(m)}(\mathbf{u}) - p_N)$ of (3.9) to the autoregressive linear Gaussian process

$$Z^{(m+1)}(\mathbf{u}) = r_N Z^{(m)}(\mathbf{u}) + \sigma_N \epsilon^{(m)}(\mathbf{u}), \tag{3.30}$$

where $\sigma_N = S_N(p_N)$ and $\epsilon^{(m)}(\mathbf{u}) = \epsilon(u_m)$ and to

$$Z_*^{(m+1)}(\mathbf{u}) = r^* Z_*^{(m)}(\mathbf{u}) + \sigma^* \epsilon^{(m)}(\mathbf{u}), \tag{3.31}$$

where r^* and σ^* have been defined in (2.8) and (2.19), respectively. In order to achieve these comparisons, we suppose that $V^{(m)}(\mathbf{u})$ is in its stationary regime, that $Nc(N) \to \infty$ as $N \to \infty$ and that $N \geq N_5(\mathbf{x})$ as defined in Lemma 3.5. For simplicity, we will suppress the argument $\mathbf{u}$ in the notation

and write $\xi_N^{(m)} = \xi_N(q^{(m)}, u_m)$.
From Lemmas 3.3 and 3.4 it follows that

$$|V^{(m+1)} - (r_N V^{(m)} + \sigma_N \xi_N^{(m)})| \leq \frac{A_0}{N}|V^{(m)}|^2 + \frac{B_0}{\sqrt{N}}|V^{(m)}||\xi_N^{(m)}| \tag{3.32}$$

which implies

$$\| V^{(m+1)} - Z^{(m+1)} \|_2 \leq r_N \| V^{(m)} - Z^{(m)} \|_2 + \alpha(N), \tag{3.33}$$

where

$$\alpha(N) = \frac{A_0}{\sqrt{N}\{1 - R(N)\}^2} + \frac{2B_0}{\sqrt{N}\{1 - R(N)\}} + \beta(N), \tag{3.34}$$

with

$$\beta_N = O\{\gamma^{1/8}(N) log^{1/4}\left(\frac{1}{\gamma(N)}\right)\}. \tag{3.35}$$

If we select $Z^{(0)} = V^{(0)}$, it follows from (3.33) that

$$\| V^{(m)} - Z^{(m)} \|_2 \leq \frac{\alpha(N)}{1 - r_N}, \tag{3.36}$$

with

$$Z^{(m)} = (r_N)^{(m)} V^{(0)} + \sigma_N \sum_{j=1}^{m} (r_N)^{m-j} \epsilon^{(j)}. \tag{3.37}$$

If, moreover, we select $Z_*^{(0)} = Z^{(0)} = V^{(0)}$, we obtain from (3.30)-(3.31) and (3.33)

$$\| V^{(m)} - Z_*^{(m)} \|_2 \leq \frac{\alpha(N) + \delta(m, N)}{1 - r_N}, \tag{3.38}$$

where

$$\delta(m, N) = |r_N - r^*| \left(\frac{(r^*)^m}{1 - R(N)} + \frac{\sigma^*}{1 - r^*} \right) + |\sigma_N - \sigma^*|. \tag{3.39}$$

Finally, if we select for each N large enough an integer $m(N)$ such that

$$\rho(N) = \frac{(r^*)^{m(N)}}{1 - R(N)} \to 0 \ as \ N \to \infty \tag{3.40}$$

and denote $\delta(N) = \delta(m(N), N)$ then we obtain a r.v. $V_N = V^{(m(N))}$ with distribution Λ_N and a Gaussian r.v. Z_N,

$$Z_N = \sigma^* \sum_{j=1}^{m(N)} (r^*)^{m(N)-j} \epsilon^{(j)}, \tag{3.41}$$

with mean 0 and variance $v^*\{1 - (r^*)^{2m(N)}\}$, such that

$$\| V_N - Z_N \|_2 \leq \frac{\alpha(N) + \delta(N)}{1 - r_N} + \rho(N). \tag{3.42}$$

Since $m(N)$ can be taken arbitrarily large, we can assume that

$$\| V_N - Z_N \|_2 \leq a(N), \tag{3.43}$$

where

$$a(N) = 2\frac{\alpha(N) + \delta(N)}{1 - r_N} + \rho(N). \tag{3.44}$$

and

$$\delta(N) \leq \frac{2\sigma^* |r_N - r^*|}{1 - r^*} + |\sigma_N - \sigma^*|. \tag{3.45}$$

Step 5 Before proceeding to Part II, we deduce from the results obtained above several properties of the asymptotic behavior of Λ_N from which the assertions (i)-(iii) of Theorem 1 will be derived in Part II.

1. In view of (3.42)-(3.45) and the convergence in distribution of Z_N to a Gaussian r.v. with mean 0 and variance v^*, it results from Lemma 2, p. 254, Feller (1971) that V_N converges in distribution to the same limit.

2. Since $\{V^{(m)}\}$ is assumed stationary,

$$\begin{aligned} \text{Mean } (\Lambda_N) &= E_{\mathrm{U}}(q^{(m)}) \\ &= E_{\mathrm{U}}\left(p_N + N^{-1/2} V^{(m)}\right) \\ &= p_N + N^{-1/2} E_{\mathrm{U}}(V_N) \end{aligned} \tag{3.46}$$

by taking $m = m(N)$. Thus

$$\begin{aligned} |\text{Mean } (\Lambda_N) - p_N| &\le N^{-1/2}\{|E_{\mathbf{U}}(Z_N)| + \| V_N - Z_N \|_1\} \\ &\le N^{-1/2} \| V_N - Z_N \|_2 \qquad (3.47) \\ &\le N^{-1/2} a(N), \end{aligned}$$

since Z_N has mean 0 and by making use of the Cauchy-Schwarz Inequality. It can be proved similarly that

$$|\text{Var } (\Lambda_N) - \frac{v^*}{N}| = o(\frac{1}{N}) \qquad (3.48)$$

provided that the rate of convergence of $c(N)$ to 0 is such that $a(N)/\{1-R(N)\}$ converges to 0.
Similar bounds can be derived for higher moments.

3. Let $Q_N(t) = P_{\mathbf{U}}\{q^{(m)} \le t\}$ be the distribution function of $q^{(m)}$ in its stationary regime and $\Phi_{m,N}(t) = P_{\mathbf{U}}\{p_N + \frac{Z^{(m)}}{\sqrt{N}} \le t\}$.
From (3.36) and Chebyschev Inequality it results that, for all real t and positive h,

$$|Q_N(t) - \Phi_{m,N}(t+h)| \le \frac{1}{h^2}\left[\frac{\alpha(N)}{1-r_N}\right]^2. \qquad (3.49)$$

Letting $m \to \infty$, we obtain that, for all real t,

$$|Q_N(t) - \Phi_N(t)| \le \frac{1}{h^2}\left[\frac{\alpha(N)}{1-r_N}\right]^2 + \frac{h}{\sqrt{N}}\left(\sup_{\left[t-\frac{h}{\sqrt{N}},t+\frac{h}{\sqrt{N}}\right]} \varphi_N\right), \qquad (3.50)$$

where $\Phi_N(t)$ is the normal distribution function with mean p_N and variance $\sigma_N^2/\{N(1-r_N^2)\}$ and $\varphi_N(t)$ is the corresponding density. A proper selection of $h = h_N$ in (3.50) yields for the sup-norm $\| Q_N - \Phi_N \|_\infty$:

$$\begin{aligned} \| Q_N - \Phi_N \|_\infty &\le \alpha^{2/3}(N)\left\{\left(\tfrac{1}{1-r_N}\right)^2 + \tfrac{1}{\sigma_N}\sqrt{\tfrac{1-r_N^2}{2\pi}}\right\} \\ &= O(\alpha^{2/3}(N)) \text{ as } N \to \infty. \end{aligned} \qquad (3.51)$$

4. Finally, we will use in Part II of the proof of Theorem 1 an upper bound for $\Lambda_N(J_N^c)$, where $J_N = [c(N), 1 - c(N)]$ and J_N^c is the complement of J_N in $[0,1]$. Since the measure Λ_N is concentrated on $[0,1]$, the distribution function of Λ_N satisfies $Q_N(0) = 0$ and $Q_N(1) = 1$, and we have $\Lambda_N(J_N^c) = Q_N(c(N)) + (1 - Q_N)(1 - c(N))$. Now, using inequality (3.50) with a proper choice of $h = h_N$ yields

$$\Lambda_N(J_N^c) = O\left(\frac{\alpha^2(N)}{N}\right) \text{ as } N \to \infty. \tag{3.52}$$

PART II

In order to derive (i)-(iii) of Theorem 1 from points (1)-(3) of Step 5 of Part I, we will first show that $\Lambda_N(B) \leq \Psi_N(B)$ for all Borel subsets B of J_N. To this end, we consider the Skorohod representation of $q^{(m)}$ and $p^{(m)}$, assuming that $q^{(0)} = p^{(0)} \in J_N$. Let $k_\ell(\ell \geq 1)$ denote the ℓ th exit time of $p^{(m+\frac{1}{2})}$ from J_N. For $0 \leq m \leq k_1, q^{(m)} = p^{(m)}$ and $q^{(k_1+1)} = p^{(k_1+\frac{1}{2})} \notin J_N$. Since $\tilde{T}_N(q) = j(N)/N \in J_N$ and $\tilde{S}_N(q) = 0$ for $q \notin J_N$, $q^{(k_1+2)} = j(N)/N$. Also, since $p^{(k_1+1)}$ is drawn from Γ_N and Γ_N is assumed to be the Dirac measure at $j(N)/N$, $p^{(k_1+1)} = q^{(k_1+2)} = j(N)/N$. An induction shows that, for $\ell \geq 0$,

$$q^{(m)} = p^{(m-\ell)} \; if \; k_\ell + \ell \leq m \leq k_{\ell+1} + \ell - 1, \tag{3.53}$$

with the convention $k_0 = 0$. Hence, if $I_B(q^{(m)}) = 1$, where $I_B(.)$ is the indicator function of $B \subset J_N$, then there exists $m' \leq m$ such that $I_B(p^{(m')}) = 1$, implying that

$$\frac{1}{M+1}\sum_{m=0}^{M} I_B(q^{(m)}) \leq \frac{1}{M+1}\sum_{m=0}^{M} I_B(p^{(m)}) \tag{3.54}$$

for all M. But, by ergodicity, letting $M \to \infty$ in (3.54) yields

$$\Lambda_N(B) \leq \Psi_N(B) \; for \; all B \subset J_N. \tag{3.55}$$

Next, since Ψ_N is concentrated on J_N, it follows from (3.55) and $\Lambda_N([0,1]) = \Psi_N([0,1]) = 1$ that the total variation $\| \Psi_N - \Lambda_N \|_{TV}$ satifies

$$\| \Psi_N - \Lambda_N \|_{TV} \leq 2\Lambda_N(J_N^c) \tag{3.56}$$

which implies (i) and (iii), in view of points (1) and (3) in Part I. Finally, since $p^{(m)}$ and $q^{(m)}$ are in $[0,1]$, (3.56) implies that

$$\begin{aligned} \mid Mean(\Psi_N) - Mean(\Lambda_N) \mid &\leq \int_0^1 t \mid \Psi_N - \Lambda_N \mid (dt) \\ &= O(\frac{\alpha^2(N)}{N}), \end{aligned} \tag{3.57}$$

and, similarly,

$$\mid Var(\Psi_N) - Var(\Lambda_N) \mid = O(\frac{\alpha^2(N)}{N}), \tag{3.58}$$

thus (ii) is obtained from point (2), which completes the proof of Theorem 1.□

4 Sequential Versions of SEM

In this section, we turn our attention to two different sequential versions of SEM. Sequential procedures are used when the observations $x_1, \ldots, x_n, \ldots$ are received one at a time and the estimation of the mixture parameters have to be updated before the next observation is received. The Chapter 6 in Titterington, Smith and Makov (1985) provides a thorough examination of sequantial methods in the mixture setting. Great attention is paid to the particular problem of estimating the mixing proportion p^* for two-component mixtures where the component densities are assumed known. This is the problem that we address in this section. Titterington, Smith and Makov review the main approaches to this problem, namely, the decision-directed method (Davisson and Schwartz, 1970), the Quasi-Bayes method (Makov and Smith, 1977), the probabilistic editor method (e.g., Owen, 1975), the method of moments (e.g., Odell and Basu, 1976), a Newton-Raphson-type gradient algorithm for finding the minimum of the Kullback-Leibler divergence (Kazakos, 1977) and the probabilistic teacher method (Agrawala, 1970, Silverman, 1980). This last approach is nothing but a one-step sequential version of SEM and is the object of Subsection 4.1 below. Moreover, Titterington (1984) introduces a general recursive method which, in the particular mixture problem under consideration, can be written as

$$p^{(m+1)} = p^{(m)} + (m+1)^{-1} J_c^{-1}(p^{(m)}) \mathrm{S}(x_{m+1}, p^{(m)}) \tag{4.1}$$

where $J_c(p) = [p(1-p)]^{-1}$ is the complete data Fisher information for a single observation from $h(x,p) = pf_1(x) + (1-p)f_2(x)$ and $\mathrm{S}(x,p) = [f_1(x) - f_2(x)]/h(x,p)$ is the score $(\partial/\partial p)\log h(x,p)$.

As noted in Titterington *et al.* (1985), p. 184: "A convenient way to approach the study of the asymptotic properties of the various proposed procedures is through the theory of stochastic approximation which exploits the martingale strucure implicit in the recursions involved in these methods" (see (4.1), for instance). Thus, the consistency of the estimators derived from the Quasi-Bayes method, the Kazakos algorithm, the probabilistic teacher method and the Titterington algorithm have been proved using results from martingale theory.

A good measure of the efficiency of a sequence of estimators $\{p^{(m)}\}$ is its asymptotic relative efficiency defined as

$$ARE = \lim_{n\to\infty} \frac{mV^*}{Var(p^{(m)})} \tag{4.2}$$

where $V^* = J_{obs}^{-1}$ is the Cramer-Rao lower bound. Kazakos (1977) has designed his algorithm to be fully efficient, i.e. to have $ARE = 1$. But his scheme requires numerical integration which are computationnally unattractive. Now, it is a striking fact that the ARE's of the Quasi-Bayes, the probabilistic teacher and the Titterington algorithms are positive iff the ratio $J_{obs}/J_c > 1/2$.

4.1 The one-step sequential SEM algorithm

This is the standard sequential version of SEM. Each time a new observation x_{m+1} is received, only the classification z_{m+1} is drawn at random from the current posterior probability $t(x_{m+1}, p^{(m)})$. There is no feedback as to the correctness of previous decisions: The other $z^{(j)}$'s, $1 \leq j \leq m$, are kept constant. More formally the one-step sequential SEM works as follows. Denote by $p^{(m)}(m \geq 1)$ the estimate of p^* computed on the basis of the observations $x_1, \ldots, x_m$ and by $p^{(0)}$ the starting point of the algorithm. After x_{m+1} has been received, the E-step consists of computing the posterior probability $t_{m+1} = t(x_{m+1}, p^{(m)})$, according to (2.3). The S step draws the r.v. z_{m+1} from a Bernoulli distribution with parameter t_{m+1}. The M-step updates $p^{(m)}$ as

$$p^{(m+1)} = p^{(m)} + \frac{z_{m+1} - p^{(m)}}{m+1}. \tag{4.3}$$

As noted above, the recursion (4.3) can be viewed as the probabilistic teacher algorithm of Agrawala (1970). Thus, the results obtained by Silverman

(1980) can be transferred. These results are summarized in the Theorem 2 below.

Theorem 2 (Silverman 1980) *(i)* $p^{(m)} \to p^*$ *a.s. as* $m \to \infty$.
(ii) The ARE of the one-step sequential SEM algorithm is equal to

$$ARE = \max(0, 2 - \frac{J_c}{J_{obs}}) \tag{4.4}$$

4.2 The global sequential SEM algorithm

In this subsection, we inherit the notation of Sections 2 and 3. The main object of this subsection is to explain and prove Theorem 3. This theorem concerns convergence properties of the particular sequential version of SEM that we call the global sequential SEM algorithm. This version works as follows. Denote by $p^{(m)}(m \geq 1)$ the estimate of p^* computed on the basis of the observations $x_1, \ldots, x_m$ and by $p^{(0)}$ the starting point of the algorithm. After the $(m+1)$th observation, x_{m+1}, has been recorded, the E-step updates the posterior probabilities as $t_i^{(m+1)} = t(x_i, p^{(m)}), i = 1, \ldots, m$, and computes the new posterior probability $t_{m+1}^{(m+1)} = t(x_{m+1}, p^{(m)})$, according to (2.3). The S-step draws independently each r.v. $z_i^{(m+1)}, i = 1, \ldots, m+1$, from a Bernoulli distribution with parameter $t_i^{(m+1)}$. The M-step updates p^m as

$$p^{(m+1)} = \tilde{T}_{m+1}(p^{(m)}) + (m+1)^{-1}\tilde{S}_{m+1}(p^{(m)})\xi^{(m+1)}, \tag{4.5}$$

where $\xi^{(m+1)} = \xi_{m+1}(p^{(m)}, \omega^{(m+1)})$.

The important difference with the one-step sequential SEM algorithm is that all the observations are again randomly attributed to one of the components of the mixture after each new observation has been recorded. Accordingly, there are more and more computations involved in the m th iteration as m increases. On the other hand, one can expect that the convergence to p^* will be much quicker. This is suggested by the assertion (iv) of Theorem 3 below, which states that the ARE of the global sequential version of SEM is positive (at least under the assumption that $c(N)$ and $R(N)$ be constant). On the contrary, the ARE of the one-step sequential version of SEM is zero whenever $J_c/J_{obs} \geq 2$ (see (4.4)).

Note that, since Theorem 3 (iii) establishes the a.s. convergence of $p^{(m)}$ to $p^* \in (0,1)$ as $m \to \infty$, there is a.s. at most a finite number of $p^{(m)}$'s outside the intervals J_m. This means that the choice of the procedure when $p^{(m)}$ exits from J_m is not important. This is the reason why we have made choice of the procedure implicitly contained in (4.5): It is the most convenient for proving Theorem 3, following our approach.

Theorem 3 (i) asserts that, under assumptions which parallel (H1)-(H4) above, the $\mathbf{P}_\Omega$ - distribution of $p^{(m)} = p^{(m)}(\mathbf{x}, \omega)$ is $\mathbf{P}_\mathbf{X}$ - a.s. asymptotically normal with mean p_m and variance $v^* m^{-1}$, where $p_m = p_m(\mathbf{x})$ is the ML estimate of p^* based on the sample $\{x_1, \ldots, x_m\}$ and $v^* = (\sigma^*)^2\{1-(r^*)^2\}^{-1}$, with r^* and σ^* defined in (2.8) and (2.19), respectively.

Theorem 3 (ii) provides $\mathbf{P}_\mathbf{X}-$ a.s. asymptotic upper bounds for $|E_\Omega(p^{(m)}) - p_m|$ and $|Var_\Omega(p^{(m)}) - v^* m^{-1}|$. These bounds make (i) more precise.

Before stating Theorem 3, we need to state the assumptions (H5) and (H6), which are as follows.

(H5) For $\mathbf{P}_\mathbf{X}-$ a.e.$\mathbf{x}$, $\Sigma_{m\geq 1} m^{-2}\{1 - R(m)\}^{-4} < \infty$ and $\Sigma_{m\geq 1} m^{-1}\beta^2([\theta m]) < \infty$ for any $\theta, 0 < \theta < 1$, where $[t]$ denotes the largest integer $\leq t$.

(H6) There exists an exponent $\mu, 0 \leq \mu < 1$, such that $m^{-\mu} = O(c(m))$. (For $\mu = 0$, this means that $c(m)$ is constant. Note that (H6) implies (H2).)

Finally, recall that $\ell(m) = \{2\ell_2(m)\}^{1/2}$, where $\ell_2(m)$ denotes the iterated logarithm. We can now state Theorem 3.

Theorem 3 *(i) Suppose that the assumptions (H1), (H3)-(H4) and (H6) hold. Then, $m^{1/2}$ $(p^{(m)} - p_m)$ converges $\mathbf{P}_\mathbf{X}$ - a.s. in $\mathbf{P}_\Omega$ - distribution to a Gaussian r.v. with mean 0 and variance v^*.*

(ii) Under the assumptions (H1), (H3)-(H4) and (H6), for $\mathbf{P}_\mathbf{X}$ - a.e. $\mathbf{x}$ and all m large enough,

$$|E_\Omega(p^{(m)}) - p_m| \leq m^{-1/2} a_{seq}(m) = o(m^{-1/2}) \quad (m \to \infty) \tag{4.6}$$

and

$$|\,Var_\Omega^{1/2}(p^{(m)}) - m^{-1/2}(v^*)^{1/2}| \leq 2m^{-1/2} a_{seq}(m) + O(m^{-1}\ell(m)), \tag{4.7}$$

where, for all $\theta, 0 < \theta < 1$,

$$a_{seq}(m) = O(\beta([\theta m])) + O(m^{-1/2}\{1 - R(m)\}^{-2}) \quad (m \to \infty). \tag{4.8}$$

(iii) Under the assumptions (H1), (H3) and (H5)-(H6),

$$\lim_{m\to\infty} (p^{(m)} - p_m) = 0 \;\; \mathbf{P}_\mathbf{X} \otimes \mathbf{P}_\Omega - a.s. \tag{4.9}$$

Futhermore, under (H1), (H3) and (H6) with $0 \leq \mu < 1/4$,

$$\limsup_{m \to \infty} m^{\nu}|p^{(m)} - p_m| = 0 \quad \mathsf{P}_{\mathbf{X}} \otimes \mathsf{P}_{\Omega} - a.s. \tag{4.10}$$

for $0 \leq \nu < \min\{(1-\mu)/8, (1-4\mu)/2, (9-33\mu)/32\}$.

(iv) *Under the only assumption that* $c(m)$ *and* $R(m)$ *are constant with* $0 < c < 1/2$ *and* $0 < R < 1$*, the ARE of the global sequential SEM algorithm is positive.*

Proof of Theorem 3. The proof of Theorem 3 parallels that of Theorem 1, except for the proof of (iv). However, since it is delicate, we detail each of its steps. Step 1 contains the proof of some technical results about the rates of $c(m)$, $R(m)$ and related quantities ($m \to \infty$). Step 2 establishes a result analogous to (3.14). Step 3 constructs the Skorohod representation we need and state the estimate (4.22), analogous to (3.20). Step 4 obtains an estimate of the rate of convergence to 0 of $E_{\mathbf{U}}(|Y^{(m)} - Z^{(m)}|^2)$ as $m \to \infty$, where $E_{\mathbf{U}}$ denotes the expectation with respect to the Skorohod representation probability space, $Y^{(m)} = Y^{(m)}(\mathbf{u}) = m^{1/2}\{p^{(m)}(\mathbf{x}, \mathbf{u}) - p_m(\mathbf{x})\}$ and $\{Z^{(m)} = Z^{(m)}(\mathbf{u})\}$ represents a suitable Gaussian process, such that the r.v. $Z^{(m)}$ has mean 0 and variance $v^{(m)} \to v^*$ as $m \to \infty$. Step 5 deduces the assertions (i)-(iii) of Theorem 3 from the preceding steps. Finally, Step 6 establishes the assertion (iv).

We now proceed to explain the successive steps of the proof of Theorem 2.

Step 1 We begin with some technical results concerning the assumptions (H5) and (H6) in Theorem 3. These results are stated in Lemma 4.1, which is as follows.

Lemma 4.1 *Suppose that the assumption (H6) in Theorem 2 holds. Then,*

(i) *For any* $\theta, 0 < \theta < 1$,

$$\sum_{m=1}^{\infty} m^{-1}\beta^2([\theta m]) < \infty, \tag{4.11}$$

where $\beta(m)$ *is defined as in (3.35)-(3.36) and* $[t]$ *denotes the largest integer* $\leq t$.

(ii) A sufficient, but not necessary, condition for (H5) to hold is $\mu < 1/4$.

(iii) We have, for any $\theta, 0 < \theta < 1$,

$$m^{3/2} \prod_{k=1}^{m} R(k) \to 0 \quad \text{as } m \to \infty \tag{4.12}$$

and

$$m^{3/2} \prod_{k=[m\theta]}^{m} R(k) \to 0 \quad \text{as } m \to \infty. \tag{4.13}$$

The proof of Lemma 4.1 can be found in the Appendix.

Step 2 In Step 2, we establish the following Lemma 4.2, which concerns an estimate of

$$||Y^{(m)}||_{\Omega,4} = E_{\Omega}^{1/4}(|Y^{(m)}|^4) \quad (m \to \infty), \tag{4.14}$$

where $Y^{(m)} = m^{1/2}\{p^{(m)}(\mathbf{x},\omega) - p_m(\mathbf{x})\}$ and the notation E_Ω refers to expectation with respect to the probability measure P_Ω. Notice that the result that we obtain in Lemma 4.2, namely (4.15), is true for $\mathsf{P}_\mathbf{X}$ - a.e. sample sequence $\mathbf{x}$ and all m large enough, but cannot be integrated with respect to $\mathsf{P}_\mathbf{X}$.

Lemma 4.2 *Suppose that the assumptions (H1), (H3) and (H6) hold. Assume in addition that $m\{1 - R(m)\}^2 \to \infty$ as $m \to \infty$. Then, there exists for P_X - a.e. $\mathbf{x}$ a finite integer $m_0 = m_0(\mathbf{x})$ such that $m \geq m_0$ implies*

$$||Y^{(m)}||_{\Omega,4} \leq \{1 - R(m)\}^{-1}. \tag{4.15}$$

Proof of Lemma 4.2 We have for all $m \geq 1$

$$\begin{aligned} &||Y^{(m+1)}||_{\Omega,4} \leq (m+1)^{1/2} m^{-1/2} R(m+1) ||Y^{(m)}||_{\Omega,4} \\ &+(m+1)^{1/2} R(m+1)\Delta_{m+1} + w(m+1), \end{aligned} \tag{4.16}$$

where $\Delta_{m+1} = |p_{m+1} - p_m|$ and $w(m) = (1/4)||\xi^{(m)}||_{\Omega,4} \to 1/4$ $\mathsf{P}_\mathbf{X}$ - a.s. as $m \to \infty$ (see (2.8), (2.19) and (3.15)). This in turn leads to

$$||Y^{(m)}||_{\Omega,4} \leq I_m + II_m + III_m \quad \text{for all } m \geq 1, \tag{4.17}$$

where $I_m = m^{1/2}\Pi_m(2)||Y^{(1)}||_{\Omega,4} = o(1)$ in view of (H6) and Lemma 4.1,

$$\left\{\begin{array}{rcl} II_m & = & m^{1/2}\sum_{j=2}^{m}\Pi_m(j)\Delta_j \quad \text{and} \\ III_m & = & \sum_{j=2}^{m}(m/j)^{1/2}\Pi_m(j+1)w(j), \end{array}\right. \tag{4.18}$$

and $\Pi_m(j) = R(j)\ldots R(m)$ for $j = 1,\ldots,m$, with the convention that $\Pi_m(m+1) = 1$.

Since $R(j) \leq R(j+1)$ for $j \geq 1, \Pi_m(j) \leq \Pi_m(j+1)$ for $j = 1,\ldots,m$ and, by Lemma A.4 in the Appendix, there exists a.s. a finite integer $m_1 = m_1(\mathbf{x})$ such that $m \geq m_1$ implies $\Delta_m \leq (2J_c/J_{obs})m^{-1}$, splitting II_m into $A_m + B_m$, with $A_m = m^{1/2}\Sigma_{2\leq j\leq[\theta m]}\Pi_m(j)\Delta_j$ and $B_m = m^{1/2}\Sigma_{[\theta m]+1\leq j\leq m}\Pi_m(j)\Delta_j$, where $0 < \theta < 1$ and $[t]$ denotes the largest integer $\leq t$, yields for $m \geq m_2$:

$$\begin{array}{rcl} II_m & \leq & m^{1/2}\sum_{j=2}^{[\theta m]}\Pi_m([\theta m]) + m^{1/2}(2J_c/J_{obs})[\theta m]^{-1} \\ & & (1 + R(m) + R^2(m) + \ldots) \\ & \leq & m^{3/2}\Pi_m([\theta m]) + m^{-1/2}\theta^{-1}(2J_c/J_{obs})\{1 - R(m)\}^{-1} \\ & = & o(1) \end{array} \tag{4.19}$$

in view of (H6) and Lemma 4.1. Splitting III_m similarly provides

$$\begin{array}{rcl} III_m & \leq & (m/2)^{1/2}\Pi_m([\theta m]+1)[\theta m][\theta m]^{-1}\sum_{j=2}^{[\theta m]}w(j) \\ & & +(m/[\theta m])^{1/2}(\max_{[\theta m]\leq j\leq m} w(j))(1 + R(m) + \ldots) \\ & \leq & m^{3/2}\Pi_m([\theta m]+1)[\theta m]^{-1}\sum_{j=2}^{[\theta m]}w(j) + \theta^{-1/2}\{1 - R(m)\}^{-1} \\ & & (\max_{[\theta m]\leq j\leq m} w(j)). \end{array} \tag{4.20}$$

Since the Cesaro mean and $\max_{[\theta m]\leq j\leq m} w(j) \to 1/4$ as $m \to \infty$ and $m^{3/2}\Pi_m([\theta m]+1) = o(1)$, (4.16)-(4.20) together imply

$$||Y^{(m)}||_{\Omega,4} \leq o(1) + (1/3)\theta^{-1/2}\{1 - R(m)\}^{-1} \tag{4.21}$$

whenever $m \geq m_2 = m_2(\mathbf{x}) \geq m_1$. Choosing θ such that $(1/3)\theta^{-1/2} < 1$ leads to (4.15) for $m \geq m_0(\mathbf{x}) \geq m_2$, as required. $\square$

Step 3 In Step 3, we examine the construction of the Skorohod representation of the r.v.'s $\xi_m(p,\omega)$ and $\epsilon^{(m)}(\omega)$ for $p \in [0,1]$ and $m \geq 1$, as well as the estimate (4.22), uniform in $p \in [0,1]$, for the mean square distance between these respective representation r.v's, as $m \to \infty$.
Again, (4.22) is true for $\mathsf{P}_\mathbf{X}$ - a.e. sample sequence $\mathbf{x}$ and for m large enough. It cannot be integrated with respect to $\mathsf{P}_\mathbf{X}$.

Lemma 4.3 *Under the assumptions (H1) and (H3) of Theorem 3, there exists a probability space* $(\mathbf{U} = \{\mathbf{u} = (u_1, u_2, \ldots)\}; \mathsf{P}_\mathbf{U})$ *and r.v.'s* $\xi_m(p, u_m)$ *and* $\epsilon(u_m), m = 1, 2, \ldots,$ *such that*

(i) $\xi_m(p, u_m)$ *and* $\epsilon(u_m)$ *have the same distributions as* $\xi_m(p,\omega)$ *and* $\epsilon^{(m)}(\omega)$, *respectively, for each* $m = 1, 2, \ldots$ *and all* p *in* $[0,1]$.

(ii) For each fixed p, *the r.v.'s* $\xi_m(p, u_m), m = 1, 2, \ldots,$ *are mutually independent and the r.v.'s* $\epsilon(u_m), m = 1, 2, \ldots,$ *are i.i.d. and Gaussian with mean 0 and variance 1.*

(iii) For $\mathsf{P}_\mathbf{X}$ *- a.e.* $\mathbf{x}$, *there exists a finite integer* $m_3 = m_3(\mathbf{x}) \geq m_0(\mathbf{x})$ *such that, for all* $m \geq m_3$ *and* p *in* $[0,1]$,

$$E_\mathbf{U}(|\xi_m(p, u_m) - \epsilon(u_m)|^2) \leq 10\gamma^{1/4}(m)|\log\{\gamma(m)\}|^{1/2} + 2\gamma^{1/2}(m), \tag{4.22}$$

where $\gamma(m) = C_{BE}[r'mc(m)]^{-1/2}$ *is as in Section 3 (see (3.20)).*

The proof of Lemma 4.3 completely parallels that of Lemma 3.5, and is thus omitted.

Step 4 In Step 4, we show that there exists a Gaussian process $\{Z^{(m)} = Z^{(m)}(\mathbf{u})\}$, defined on the Skorohod representation probability space $(\mathbf{U}, \mathsf{P}_\mathbf{U})$ introduced in Step 3, such that the versions $Y^{(m)} = Y^{(m)}(\mathbf{x}, \mathbf{u})$ of $m^{1/2}\{p^{(m)}(\mathbf{x}, \omega) - p_m(\mathbf{x})\}$ defined by making use of the r.v.'s $\xi_m(p, u_m)$ introduced in Lemma 4.3 are close to $Z^{(m)} = Z^{(m)}(\mathbf{u})$ in the mean square sense, for $\mathsf{P}_\mathbf{X}$ - a.e. $\mathbf{x}$.
More precisely, let $\{Z^{(m)}(\mathbf{u})\}$ be the linear autoregressive Gaussian process,

defined on $(\mathbf{U}, \mathsf{P}_{\mathbf{U}})$ by the recursion formula

$$Z^{(m+1)}(\mathbf{u}) = (m+1)^{-1/2} r_{m+1} Z^{(m)}(\mathbf{u}) + \sigma_{m+1}\epsilon^{(m+1)}(\mathbf{u}) \quad \text{for } m \geq 1, \tag{4.23}$$

with $Z^{(1)}(\mathbf{u}) = \sigma_1 \epsilon^{(1)}(\mathbf{u})$. Then,

$$Z^{(m)}(\mathbf{u}) = \sum_{j=1}^{m} (m/j)^{1/2} \pi_m(j+1)\sigma_j \epsilon^{(j)}(\mathbf{u}) \quad \text{for } m \geq 1, \tag{4.24}$$

where $\pi_m(j) = r_j \ldots r_m$ for $j = 1, \ldots, m$ and $\pi_m(m+1) = 1$, is a Gaussian r.v. with mean 0 and variance

$$v^{(m)} = \sum_{j=1}^{m} (m/j)\pi_m^2(j+1)\sigma_j^2. \tag{4.25}$$

Moreover, define $Y^{(m)}(\mathbf{x}, \mathbf{u}) = m^{1/2}\{p^{(m)}(\mathbf{x}, \mathbf{u}) - p_m(\mathbf{x})\}$, where $p^{(0)}(\mathbf{x}, \mathbf{u}) = p^{(0)}$ and, for $m \geq 0$,

$$\begin{aligned} p^{(m+1)}(\mathbf{x}, \mathbf{u}) \quad = \quad & \tilde{T}_{m+1}\{p^{(m)}(\mathbf{x}, \mathbf{u})\} + (m+1)^{-1/2} \\ & \tilde{S}_{m+1}\{p^{(m)}(\mathbf{x}, \mathbf{u})\}\xi^{(m+1)}(\mathbf{x}, \mathbf{u}), \end{aligned} \tag{4.26}$$

where $\xi^{(m+1)}(\mathbf{x}, \mathbf{u}) = \xi_{m+1}\{p^{(m)}(\mathbf{x}, \mathbf{u}), u_{m+1}\}$.

In the sequel, we will suppress the notation indicating the dependence on $\mathbf{u}$ or $(\mathbf{x}, \mathbf{u})$, unless necessary. We will let $||Y||_{\mathbf{U},\alpha} = E_{\mathbf{U}}^{1/\alpha}(|Y|^\alpha)$ for all finite $\alpha \geq 1$.

Lemma 4.4 below provides an estimate for $E_{\mathbf{U}}(|Y^{(m)} - Z^{(m)}|^2)$ as $m \to \infty$. This estimate is uniform in $p \in [0, 1]$ and is true for $\mathsf{P}_{\mathbf{X}}$ - a.e. $\mathbf{x}$. Unfortunately, the integration of inequality (4.27) with respect to P_X does not lead to a similar estimate.

Lemma 4.5 below shows that the variance $v^{(m)}$ of $Z^{(m)}$ converges to v^* as $m \to \infty$ and provides an estimate for $|v^{(m)} - v^*|$.

Lemma 4.4 is as follows.

Lemma 4.4 *Under the assumptions (H1), (H3) and (H6) and the additional assumption $m\{1 - R(m)\}^4 \to \infty$ as $m \to \infty$, the uniform (in p) Skorohod representations $Y^{(m)}$ and $Z^{(m)}$ satisfy $\mathsf{P}_{\mathbf{X}}$ - a.s.*

$$||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} \leq a_{seq}(m) \tag{4.27}$$

for all m large enough, where, for all $\theta, 0 < \theta < 1$,

$$a_{seq}(m) = O(\beta([\theta m])) + O(m^{-1/2}\{1 - R(m)\}^{-2}) \quad (m \to \infty). \tag{4.28}$$

Proof of Lemma 4.4. It follows from Lemmas 3.5 and 3.4 that, for all $m \geq 1$ and $\mathbf{P_X}$ - a.e. $\mathbf{x}$,

$$\begin{aligned} &||Y^{(m+1)} - Z^{(m+1)}||_{\mathbf{U},2} \leq (m+1)^{1/2} m^{-1/2} r_{m+1} \\ &||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} + g(m+1), \end{aligned} \tag{4.29}$$

where $\{g(m)\}$ is a sequence of positive numbers decreasing to zero such that, for all $m \geq 2$ and $\mathbf{P_X}$ - a.e. $\mathbf{x}$.

$$\begin{aligned} g(m) \;\leq\; & \beta(m) + m^{1/2}\Delta_m + A_0 m^{1/2}\Delta_m^2 \\ & +2A_0(m+1)^{1/2}m^{-1/2}\Delta_m||Y^{(m-1)}||_{\mathbf{U},4} \\ & +B_0\Delta_m + A_0(m+1)^{1/2}m^{-1}||Y^{(m)}||^2_{\mathbf{U},4} + B_0 m^{-1/2}||Y^{(m)}||^2_{\mathbf{U},4} \\ & +B_0 m^{-1/2}||Y^{(m-1)}||_{\mathbf{U},4}||\xi^{(m)}||_{\mathbf{U},4} \\ \leq\; & \beta(m) + 2A_0 m^{-1/2}\{1 - R(m)\}^{-2} + o(m^{-1/2}\{1 - R(m)\}^{-2}), \end{aligned} \tag{4.30}$$

in view of Lemma A.4 and 4.2.

Now, (4.29) entails that, for all $m \geq 2$,

$$||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} \leq m^{1/2}\pi_m(2)K_1 + \sum_{j=2}^{m}(m/j)^{1/2}\pi_m(j+1)g(j), \tag{4.31}$$

where $K_1 = ||Y^{(1)} - Z^{(1)}||_{\mathbf{U},2} < \infty$ $\mathbf{P_X}$ - a.s.

Since, for all r such that $r^* < r < 1$, there exists $\mathbf{P_X}$ - a.s. a finite integer $m_4 = m_4(\mathbf{x}) \geq m_3$ such that $0 < r_m < r$ for all $m \geq m_4$, the same splitting technique as in the proof of Lemma 4.2 yields, for all $[\theta m] \geq m_4$,

$$\begin{aligned} &||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} \leq m^{1/2}\mathrm{MAX} r^{m-m_4} K_1 + \sum_{j=2}^{[\theta m]}(m/2)^{1/2} \\ &\mathrm{MAX} r^{m-[\theta m]} g(j) + (m/[\theta m])^{1/2} g([\theta m])(1 + r + r^2 + \ldots), \end{aligned} \tag{4.32}$$

where MAX=$\max(1, r_2 \ldots r_{m_4})$. Thus, $\mathbf{P_X}$ - a.s.,

$$\begin{aligned} ||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} &= O(m^{1/2}r^m) + O(m^{3/2}r^{(1-\theta)m}[\theta m]^{-1}\sum_{j=1}^{[\theta m]}g(j)) \\ &\quad +O(g([\theta m])) \quad (m \to \infty) \\ &= o(1) + O(g([\theta m])) \quad (m \to \infty), \end{aligned} \tag{4.33}$$

which gives (4.27)-(4.28), in view of (4.30), as required. $\square$

We now turn to Lemma 4.5.

Lemma 4.5 *For* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$*, we have*

$$|v^{(m)} - v^*| = O(m^{-1/2}\ell(m)) \quad (m \to \infty), \tag{4.34}$$

where $\ell(m) = \{2\ell_2(m)\}^{1/2}$ *and* $\ell_2(m)$ *is the iterated logarithm.*

Proof of Lemma 4.5. We again use the same splitting technique as above, but with $f(m) = m - s(m), s(m) \to \infty$ and $s(m) = o(m)$ as $m \to \infty$, instead of $f(m) = [\theta m]$. We have

$$\begin{aligned}
&|v^{(m)} - v^*| \leq m \sum_{j=1}^{f(m)-1} (r_{f(m)} \cdots r_m)^2 \sigma_j^2 + \ |\sigma_m^2 - (\sigma^*)^2| \\
&+|\sigma_m^2 - (\sigma^*)^2| m(m-1)^{-1} r_m^2 + \ldots + |\sigma_{f(m)}^2 - (\sigma^*)^2| \\
&mf(m)^{-1}(r_{f(m)+1} \cdots r_m)^2 + (\sigma^*)^2 |1 + (m-1)^{-1} r_m^2 + \ldots \\
&+ s(m)\{m - s(m)\}^{-1}(r_{f(m)+1} \cdots r_m)^2 - \{1 - (r^*)^2\}^{-1}|.
\end{aligned} \tag{4.35}$$

But, for $j \geq m_4(\mathbf{x}), r_j < r < 1$, whereas, by Lemma A.1, $|\sigma_j^2 - (\sigma^*)^2| = O(j^{-1/2}\ell(j))(j \to \infty)$ $\mathbf{P_X}$ - a.s. Thus, for m large enough,

$$\begin{aligned}
&|v^{(m)} - v^*| \leq mr^{2s(m)}\{f(m) - 1\}^{-1} \sum_{j=1}^{f(m)-1} \sigma_j^2 + O(f(m)^{-1/2}\ell\{f(m)\}) \\
&+(\sigma^*)^2[(m-1)^{-1}|r_m^2 - (r^*)^2| + \ldots + s(m)\{m - s(m)\}^{-1} \\
&|(r_{f(m)+1} \cdots r_m)^2 - (r^*)^{2s(m)}|] \\
&+(\sigma^*)^2|1 + (m-1)^{-1}(r^*)^2 + \ldots + s(m)\{m - s(m)\}^{-1} \\
&(r^*)^{2s(m)} - \{1 - (r^*)^2\}^{-1}.
\end{aligned} \tag{4.36}$$

Now, the first term in the RHS of (4.36) is $O(mr^{2s(m)}$ $(m \to \infty)$; the second term is $O(m^{-1/2}\ell(m))$ $(m \to \infty)$ because of properties of $s(m)$ as $m \to \infty$; the third term is $O(m^{-1/2}\ell(m))$ $(m \to \infty)$: In view of Lemma A.1, $|r_j^2 - (r^*)^2| = O(j^{-1/2}\ell(j))$ $(j \to \infty)$ $\mathbf{P_X}$ - a.s., thus $|(r_{m-j} \cdots r_m)^2 - (r^*)^{2j}| \leq jr^{2j}O(m^{-1/2}\ell(m))$ $(m \to \infty)$ if $1 \leq j \leq s(m)$, with $jr^{2j} =$

$j\exp\{-2|\log r|j\} \le \sup_{x>0} x\exp\{-2|\log r|x\} = (2e|\log r|)^{-1}$ and $(m-1)^{-1}+ \ldots + s(m)\{m-s(m)\}^{-1} \le s^2(m)\{m-s(m)\}^{-1} \sim m^{-1}s^2(m) = o(1)$ $(m \to \infty)$; the fourth term is bounded by $o(1) + O((r^*)^{2s(m)})$; finally, choosing $\lambda \log m \le s(m) = o(m)$ $(m \to \infty)$ with $\lambda > (2|\log r|)^{-1}$ above completes the proof. □

Step 5 We are now in a position to deduce the assertions (i)-(iii) of Theorem 2 from Lemmas 4.1-4.5.

Proof of (i): Since, by Lemma 4.5, $Z^{(m)}$ converges in $\mathbf{P_U}$ - distribution as $m \to \infty$ to a Gaussian r.v. with mean 0 and variance v^*, Lemma 4.4 implies that $Y^{(m)}$ converges in $\mathbf{P_U}$ - distribution as $m \to \infty$ to a Gaussian r.v. with mean 0 and variance v^*, for $\mathbf{P_X}$ - a.e. $\mathbf{x}$. Thus, the same is true for $Y^{(m)}(\mathbf{x}, \omega)$ in $\mathbf{P_\Omega}$ distribution. Hence (i).

Proof of (ii): From Lemma 4.4 and the Cauchy-Schwarz inequality, for $\mathbf{P_X}$ - a.e. $\mathbf{x}$ and all m large enough,

$$\begin{aligned} |E_{\mathbf{U}}(p^{(m)} - p_m - m^{-1/2}Z^{(m)})| &= |E_{\mathbf{U}}(p^{(m)}) - p_m| \\ &= |E_{\mathbf{\Omega}}(p^{(m)}) - p_m| \\ &\le m^{-1/2}||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2} \\ &\le m^{-1/2}a_{seq}(m). \end{aligned} \tag{4.37}$$

Hence,

$$|E_{\mathbf{\Omega}}(p^{(m)}) - p_m| \le m^{-1/2}a_{seq}(m) = o(m^{-1/2}) \quad (m \to \infty). \tag{4.38}$$

We now turn to the variance of $p^{(m)}$ with respect to $\mathbf{P_\Omega}$, $Var_{\mathbf{\Omega}}(p^{(m)})$. We have for $\mathbf{P_X}$ - a.e. $\mathbf{x}$ and all m large enough,

$$\Big| ||p^{(m)} - p_m||_{\mathbf{U},2} - m^{-1/2}||Z^{(m)}||_{\mathbf{U},2} \Big| \le m^{-1/2}||Y^{(m)} - Z^{(m)}||_{\mathbf{U},2}. \tag{4.39}$$

Thus,

$$\Big| ||p^{(m)} - p_m||_{\mathbf{\Omega},2} - m^{-1/2}(v^{(m)})^{1/2} \Big| \le m^{-1/2}a_{seq}(m). \tag{4.40}$$

Finally, from Lemmas 4.4 and 4.5 and (4.37)-(4.40) it follows that

$$\Big| \mathrm{Var}_{\mathbf{\Omega}}^{1/2}(p^{(m)}) - ||p^{(m)} - p_m||_{\mathbf{\Omega},2} \Big| \le ||E_{\mathbf{\Omega}}(p^{(m)}) - p_m||_{\mathbf{\Omega},2}. \tag{4.41}$$

Thus,

$$\left|\mathrm{Var}_{\mathbf{\Omega}}^{1/2}(p^{(m)}) - m^{-1/2}(v^{(m)})^{1/2}\right| \le 2m^{-1/2}a_{seq}(m) \tag{4.42}$$

and

$$\left|\mathrm{Var}_{\mathbf{\Omega}}^{1/2}(p^{(m)}) - m^{-1/2}(v^{*})^{1/2}\right| \le 2m^{-1/2}a_{seq}(m) + O(m^{-1}\ell(m)). \tag{4.43}$$

This completes the proof of (ii).

Proof of (iii): Let $P_m(t)$ and $\Phi_m(t), -\infty < t < \infty$, denote the d.f. of the r.v.'s $p^{(m)}(\mathbf{u}) - p_m$ and $m^{-1/2}Z^{(m)}(\mathbf{u})$, respectively, i.e., for $\mathbf{P_X}$ - a.e. $\mathbf{x}$ and $m \ge 1$,

$$\begin{cases} P_m(t) &= \mathbf{P_U}\{p^{(m)} - p_m \le t\} \\ \Phi_m(t) &= \mathbf{P_U}\{m^{-1/2}Z^{(m)} \le t\}. \end{cases} \tag{4.44}$$

Lemma 4.4 implies that, for $\mathbf{P_X}$ - a.e. $\mathbf{x}$, all $h > 0$ and all m large enough,

$$\sup_{-\infty<t<\infty} |P_m(t) - \Phi_m(t)| \le h^{-2}a_{seq}^2(m) + hm^{-1/2} \sup_{t-hm^{-1/2}\le s\le t+hm^{-1/2}} \phi_m(s), \tag{4.45}$$

where $\phi_m(s) = (d/ds)\Phi_m(s)$ is the normal density function with mean 0 and variance $m^{-1}v^{(m)}$. Let $\tau = \tau(m) > 0$ be given. Then, letting $h = \tau m^{1/2}$ in (4.45), we obtain that for $\mathbf{P_X}$ - a.e. $\mathbf{x}$, all m large enough and all $t \ge \tau$,

$$\begin{aligned} |P_m(t) - \Phi_m(t)| &\le \tau^{-2}m^{-1}a_{seq}^2(m) \\ &+ \tau m^{1/2}\left(2\pi v^{(m)}\right)^{-1/2}\exp\{-(1/2)m(v^{(m)})^{-1}(t-\tau)^2\} \end{aligned} \tag{4.46}$$

and, for all $t \le -\tau$,

$$\begin{aligned} &|P_m(t) - \Phi_m(t)| \le \tau^{-2}m^{-1}a_{seq}^2(m) + \tau m^{1/2}(2\pi v^{(m)})^{-1/2} \\ &\exp\{-(1/2)m(v^{(m)})^{-1}(t+\tau)^2\}. \end{aligned} \tag{4.47}$$

Since $\mathbf{P_U}\{|p^{(m)} - p_m| > t\} = (1 - P_m)(t) + P_m(-t)$ for all $t > 0$, picking a sequence $\{\tau(m)\}$ of positive numbers such that

$$\sum_{m=1}^{\infty} \tau^{-2}(m)m^{-1}a_{seq}^2(m) < \infty \tag{4.48}$$

and a sequence $\{t(m)\}$ such that $t(m) > \tau(m)$ for all m,

$$\sum_{m=1}^{\infty}(1 - \Phi_m)\{t(m)\} < \infty \tag{4.49}$$

and

$$\sum_{m=1}^{\infty} \tau(m)m^{1/2}\exp[-(1/2)m(v^{(m)})^{-1}\{t(m)-\tau(m)\}^2] < \infty \tag{4.50}$$

yields, in view of (4.46)-(4.47),

$$\mathbf{P_U}\{\limsup_{m\to\infty} t^{-1}(m)|p^{(m)} - p_m| \leq 1\} = 1 \quad \mathbf{P_X} - \text{a.s.} \tag{4.51}$$

It is possible to choose

$$t(m) = cst.m^{-\nu} \tag{4.52}$$

for any positive constant cst. and

$$0 \leq \nu < \min\{(1-\mu)/8, (1-4\mu)/2, (9-33\mu)/32\}. \tag{4.53}$$

For (H6) with $\mu = 0$ (i.e., $c(m)$ = constant), it is possible to choose

$$t(m) = cst.m^{-1/8} \tag{4.54}$$

for any positive constant cst. This concludes the proof of (iii).

Step 6 In this last step, we consider the ARE of the global sequential SEM algorithm and prove assertion (iv) of Theorem 3. This is the subject of the following lemma.

Lemma 4.6 *Under the assumption that $c(m) = c =$ constant and $R(m) = R =$ constant, with $0 < c < 1/2$ and $0 < R < 1$, the ARE of the global sequential SEM algorithm is positive.*

Proof of Lemma 4.6 It suffices to show that

$$E(|p^{(m)} - p_m|^2) = O(m^{-1}) \quad (m \to \infty), \tag{4.55}$$

where the expectation symbol E stands for $E_{\mathbf{X}\times\mathbf{\Omega}}$. To this end, we introduce the sigma-fields $\mathcal{F}_m$ generated by $x_1, \ldots, x_m$ and $\omega^{(1)}, \ldots, \omega^{(m)}$ $(m \geq 1)$.

We have

$$\begin{aligned}
E(|p^{(m+1)} - p_{m+1}|^2|\mathcal{F}_m) &= E\{|\tilde{T}_{m+1}(p^{(m)}) - p_{m+1}|^2|\mathcal{F}_m\} \\
&\quad +(m+1)^{-1}E\{\tilde{S}^2_{m+1}(p^{(m)})|\mathcal{F}_m\} \text{ a.s.} \\
&\leq R^2E(|p^{(m)} - p_{m+1}|^2|\mathcal{F}_m) \\
&\quad +(1/2)(m+1)^{-1} \text{ a.s.} \\
&\leq R^2|p^{(m)} - p_m|^2 + E(|p_{m+1} - p_m|^2|\mathcal{F}_m) \\
&\quad +2E(|p^{(m)} - p_m||p_{m+1} - p_m||\mathcal{F}_m) \\
&\quad +(1/2)(m+1)^{-1} \text{ a.s.} \\
&\leq R^2|p^{(m)} - p_m|^2 + E(|p_{m+1} - p_m|^2|\mathcal{F}_m) \\
&\quad +2|p^{(m)} - p_m|E^{1/2}(|p_{m+1} - p_m|^2|\mathcal{F}_m) \\
&\quad +(1/2)(m+1)^{-1} \text{ a.s.}
\end{aligned} \tag{4.56}$$

If we take the expectation of both members of (4.56), we obtain, by making use of the Cauchy-Schwarz inequality,

$$\begin{aligned}
&||p^{(m+1)} - p_{m+1}||_2^2 \leq R^2||p^{(m)} - p_m||_2^2 + ||p_{m+1} - p_m||_2^2 \\
&+2||p^m - p_m||_2||p_{m+1} - p_m||_2 + (1/2)(m+1)^{-1}.
\end{aligned} \tag{4.57}$$

Now, since p_m is the ML estimate of p^* based on $\{x_1, \ldots, x_m\}$, we have $||p_m - p^*||_2^2 = E(|p_m - p^*|)^2 \sim (mJ_{obs})^{-1}$ as $m \to \infty$. Thus, for any $\nu_0 > 0$, there exists a finite integer $M(\nu_0)$ such that $m \geq M(\nu_0)$ implies

$$\begin{aligned}
&||p^{(m+1)} - p_{m+1}||_2^2 \leq R^2||p^{(m)} - p_m||_2^2 + 4(1+\nu_0)^{1/2}(mJ_{obs})^{-1/2}||p^{(m)} - p_m||_2 \\
&+2(1+\nu_0)(mJ_{obs})^{-1} + (1/2)m^{-1}.
\end{aligned} \tag{4.58}$$

If we define $A(\nu_0) = A = 2(1+\nu_0)^{-1/2}J_{obs}^{-1/2}$, $B(\nu_0) = B = 2(1+\nu_0)J_{obs}^{-1} + (1/2)$ and $y_m = ||p^{(m)} - p_m||_2$, then (4.58) becomes

$$y_{m+1}^2 \leq R^2y_m^2 + 2Am^{-1/2}y_m + Bm^{-1} \tag{4.59}$$

for all $m \geq M(\nu_0)$. We now prove by induction that there exists an integer $M_1 \geq M(\nu_0)$ and a positive constant K such that $m \geq M_1$ implies $y_m \leq Km^{-1/2}$. To this end, assume that $y_m \leq Km^{-1/2}$ for some positive K and m large enough. Then, in view of (4.59),

$$y_{m+1}^2 \leq (R^2K^2 + 2AK + B)m^{-1}. \tag{4.60}$$

The RHS of (4.60) is lesser than $K^2(m+1)^{-1}$ if m is large enough and

$$(1-R^2)K^2 - 2AK - B > 0. \tag{4.61}$$

But (4.61) holds whenever $K > (1-R^2)^{-1}[A + \{A^2 + B(1-R^2)\}^{1/2}]$. □

This conclues the proof of Theorem 3. □

Remark 4.1 *Theorem 3 (ii) implies under the additional assumption (H5) that for* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$

$$\begin{aligned} m^{-1}\sum_{j=1}^{m}|E_{\mathbf{\Omega}}(p^{(j)}) - p_j| &\leq m^{-1}\left\{\sum_{j=1}^{m}|E_{\mathbf{\Omega}}(p^{(j)}) - p_j|^2\right\}^{1/2}\left(\sum_{j=1}^{m}1\right)^{1/2} \\ &\leq m^{-1/2}\left\{\sum_{j=1}^{\infty}j^{-1}a_{seq}^2(j)\right\} \\ &= O(m^{-1/2}) \quad (m \to \infty), \end{aligned} \tag{4.62}$$

and, similarly,

$$m^{-1}\sum_{j=1}^{m}\left|Var_{\mathbf{\Omega}}^{1/2}(p^{(j)}) - j^{-1/2}(v^*)^{1/2}\right| = O(m^{-1/2}) \quad (m \to \infty). \tag{4.63}$$

Remark 4.2 *The assertion (4.55) in the proof of Lemma 4.6 entails that* $p^{(m)}$ *is asymptotically unbiased, since*

$$\begin{aligned} |E_{\mathbf{X}\times\mathbf{\Omega}}(p^{(m)}) - E_{\mathbf{X}\times\mathbf{\Omega}}(p_m)| &\leq E_{\mathbf{X}\times\mathbf{\Omega}}^{1/2}(|p^{(m)} - p_m|^2) \\ &= O(m^{-1/2})\ (m \to \infty) \end{aligned} \tag{4.64}$$

and $E_{\mathbf{X}\times\mathbf{\Omega}}(p_m) = E_{\mathbf{X}}(p_m) = p^*$.

Remark 4.3 *The estimates in Theorem 3 are nonoptimal, since they essentially rely on an* $L^2(\mathbf{P_U})$ *estimate, namely (4.27). It can be conjectured that*

$$\limsup_{m\to\infty} m\ Var_{\mathbf{X}\times\mathbf{\Omega}}(p^{(m)}) \geq v^* + J_{obs}^{-1}. \tag{4.65}$$

We now try to support this conjecture. First of all, we know from (4.38) and (4.43) that $E_{\mathbf{\Omega}}(|p^{(m)} - p_m|^2) = m^{-1}v^* + o(m^{-1})$ $(m \to \infty)$ $\mathsf{P}_{\mathbf{X}}$ - a.s. Thus, $\limsup_m mE_{\mathbf{X}\times\mathbf{\Omega}}(|p^{(m)} - p_m|^2) \geq v^*$. Next, the ML estimation theory implies that $\limsup_m E_{\mathbf{X}}(|p_m - p^*|^2) = J_{obs}^{-1}$. Finally, the sample fluctuations of $m^{1/2}(p_m - p^*)$ and $m^{1/2}\{E_{\mathbf{\Omega}}(p^{(m)}) - p_m\}$ can be guessed to be asymptotically uncorrelated : indeed,

$$\begin{aligned} |E_{\mathbf{X}\times\mathbf{\Omega}}\{m^{-1/2}(p_m - p^*)D^{(m)}\}| &\leq E_{\mathbf{X}}(m^{1/2}|p_m - p^*|\ ||D^{(m)}||_{\mathbf{\Omega},2}) \\ &\leq E_{\mathbf{X}}^{1/2}(m|p_m - p^*|^2)\ ||D^{(m)}||_{\mathbf{X}\times\mathbf{\Omega},2} \\ &= O(||D^{(m)}||_{\mathbf{X}\times\mathbf{\Omega},2}) \quad (m \to \infty). \end{aligned} \tag{4.66}$$

where $D^{(m)} = Y^{(m)} - Z^{(m)}$.

Now it can be expected that $||Y^{(m)} - Z^{(m)}||_{\mathbf{X}\times\mathbf{\Omega},2} \to 0$ as $m \to \infty$. But, since $Z^{(m)}$ has $\mathsf{P}_{\mathbf{\Omega}}$ - mean $0, E_{\mathbf{X}\times\mathbf{\Omega}}\{m^{1/2}(p_m - p^*)(Y^{(m)} - Z^{(m)})\} = E_{\mathbf{X}\times\mathbf{\Omega}}\{m^{1/2}(p_m - p^*)Y^{(m)}\}$, with $Y^{(m)} = m^{1/2}(p^{(m)} - p_m)$. From a heuristic point of view, (4.65) tells us that the variance of $p^{(m)}(\mathbf{x}, \omega)$ can be split into the variance of p_m and the variance of the fluctuations related to the simulation S-step, the latter being of a magnitude $\geq v^* m^{-1}$ as $m \to \infty$. Recalling that $v^* = J_{obs}^{-1}\{1 + (J_c/J_{cond})\}^{-1} < (2J_{obs})^{-1}$, the conjecture (4.65) would imply, if it were true, that the ARE of the global sequential SEM algorithm is $\leq [1 + \{1 + (J_c/J_{cond})\}^{-1}]^{-1}$. If, furthermore, the inequality in (4.65) could be replaced by equality, then the ARE would be $> 2/3$ and would converge to 1 as J_{obs}/J_c converges to 1, i.e. as the mixture becomes more and more separable.

Remark 4.4 *As for the one-step sequential SEM algorithm, the global sequential SEM algorithm can be considered as a sequential Bayesian algorithm. Here, the underlying Bayesian algorithm is Tanner and Wong's (1987) one.*

Remark 4.5 *As for Theorem 1, extension of Theorem 3 (i) to the case where the mixture has $K \geq 3$ components is straightforward.*

Remark 4.6 *As for Theorem 1, extension of Theorem 3 (i) to a general mixture setup has been proved in Celeux and Diebolt (1986b) under the stringent assumption that $T_N(p)$ has only one fixed point in the compact G_N corresponding to J_N. This result suggests that a similar result holds in very general incomplete data settings.*

APPENDIX

Proof of Lemma 2.1

Proof of (i): From (2.3) and (2.4),

$$T_N'(p) = N^{-1} \sum_{i=1}^{N} f_1(x_i) f_2(x_i) h^{-2}(x_i, p) > 0 \quad \text{for all } p \text{ in } (0,1), \qquad \text{(A.1)}$$

where $h(x,p) = pf_1(x) + (1-p)f_2(x)$ (see (2.1)).

Proof of (ii): From (2.3) and (2.5),

$$\begin{aligned} L_N'(p) &= \sum_{i=1}^{N} \{f_1(x_i) - f_2(x_i)\} h^{-1}(x_i, p) \\ &= \sum_{i=1}^{N} [p^{-1} t(x_i, p) - (1-p)^{-1}\{1 - t(x_i, p)\}] \\ &= p^{-1}(1-p)^{-1} \sum_{i=1}^{N} [(1-p) t(x_i, p) - p\{1 - t(x_i, p)\}] \\ &= p^{-1}(1-p)^{-1} \sum_{i=1}^{N} \{t(x_i, p) - p\}, \end{aligned} \qquad \text{(A.2)}$$

hence (2.6).

Proof of (iii): From (A2),

$$L_N''(p) = -\sum_{i=1}^{N} \{f_1(x_i) - f_2(x_i)\}^2 h^{-2}(x_i, p) \leq 0 \quad \text{for all } p \text{ in } [0,1]. \qquad \text{(A.3)}$$

Now, either $f_1(x_i) = f_2(x_i)$ for $i = 1, \ldots, N$ or these exists $i, 1 \leq i \leq N$, such that $f_1(x_i) \neq f_2(x_i)$. In the first case, we have $L_N''(p) \equiv 0$, $L_N(p)$ is constant and $T_N(p) \equiv p$ on $[0,1]$. In the second case, $L_N''(p) < 0$ for all p in $[0,1]$; thus, $L_N(p)$ is concave on $[0,1]$ and has a unique maximizer on $[0,1]$.

Furthermore, for all x in X and p in $(0,1)$,

$$\begin{aligned}\{f_1(x)-f_2(x)\}^2h^{-2}(x,p) &\leq 2p^{-2}t^2(x,p)+2(1-p)^2\{1-t(x_i,p)\}^2 \\ &\leq 2\{p^{-2}+(1-p)^{-2}\} \quad (\text{since } 0 \leq t(x,p) \leq 1).\end{aligned} \tag{A.4}$$

Thus, the SLLN implies that for $\mathsf{P}_{\mathbf{X}}$ - a.e. $\mathbf{x}$ in $\mathbf{X}$ and all p in $(0,1)$

$$N^{-1}L''_N(p) \rightarrow L''(p) = -\int\{f_1(x)-f_2(x)\}^2h^{-2}(x,p)h(x,p^*)\mu(dx) \tag{A.5}$$
as $N \rightarrow \infty$.

By the assumptions of Theorem 1, $L''(p) < 0$. Hence (iii).

Proof of (iv) : From (2.6),

$$T'_N(p)-1 = (1-2p)N^{-1}L'_N(p)+p(1-p)N^{-1}L''_N(p) \quad \text{for all } p \text{ in } (0,1). \tag{A.6}$$

Thus, for $p = p_N$, where $L'_N(p_N) = 0$ and $L''_N(p_N) < 0$, we have

$$T'_N(p_N) = 1 + p_N(1-p_N)N^{-1}L''_N(p) < 1. \tag{A.7}$$

As $T'_N(p_N) > 0$ and $T_N(p_N) = p_N$ again by (2.6), (2.7) is proved.

Furthermore, since $L'_N(p) < 0$ for all $0 < p < p_N$ and $L'_N(p) > 0$ for all $p_N < p < 1$, the remainder of (iv) obtains again from (2.6). (Compare the proof in Silverman (1980).)

Proof of (v) : Assertion (v) is a direct consequence of (iv) and its proof will be omitted here.

Proof of (vi) : Remark that the empirical complete-data information value $J_{N,c} = p_N^{-1}(1-p_N)^{-1}$ and the empirical observed-data information value $J_{N,obs} = -N^{-1}L''_N(p_N)$ (e.g., Titterington *et al* (1985)). Thus, from (A.7),

$$r_N = T'_N(p_N) = 1 - J_{N,c}^{-1}J_{N,obs}. \tag{A.8}$$

Since $J_{N,c}^{-1}J_{N,obs} \rightarrow J_c^{-1}J_{obs}$ as $N \rightarrow \infty$ for $\mathsf{P}_{\mathbf{X}}$ - a.e. $\mathbf{x}$ and $J_c = J_{obs} + J_{cond}$, (vi) obtains.

It is worth noting that (A.8) also results from a general relation in Dempster *et al.* (1977) and that the information ratio $J_c^{-1}J_{obs}$ measures "the proportion of information about p without knowing the subpopulation membership [...] (and) might be interpreted as the ability of the data to distinguish the component densities" (Windham and Cutler (1991)). Indeed, it is well-known (e.g., Louis (1982) and Sundberg (1976)) that the convergence rate of the EM algorithm is the largest eingenvalue of the matrix $I - J_c^{-1}J_{obs}$, which is coherent with (vi) above. □

Lemma A.1 *(i) For* P_X *- a.e.* $\mathbf{x}$, *all* N *and all* p *in* $(0,1)$,

$$0 < T_N' \leq (1/2)p^{-1}(1-p)^{-1} \tag{A.9}$$

and

$$|T_N''(p)| \leq p^{-2}(1-p)^{-2} \tag{A.10}$$

Proof of Lemma A.1. Inequality (A.9) results from (A.1) and the elementary inequality

$$2p(1-p)f_1(x)f_2(x) \leq \{pf_1(x) + (1-p)f_2(x)\}^2. \tag{A.11}$$

Similarly,
$T_N''(p) = -2N^{-1}\Sigma_{1\leq i\leq N}f_1(x_i)f_2(x_i)\{f_1(x_i) - f_2(x_i)\}h^{-3}(x_i,p)$
and
$2p(1-p)f_1(x)f_2(x)h^{-2}(x,p) \leq 1$,
whereas
$|f_1(x) - f_2(x)|h^{-1}(x,p) \leq f_1(x)h^{-1}(x,p) + f_2(x)h^{-1}(x,p) = p^{-1}t(x,p) + (1-p)^{-1}\{1 - t(x,p)\} \leq p^{-1} + (1-p)^{-1} = p^{-1}(1-p)^{-1}$. □

Lemma A.2 *For* P_X *- a.e.* $\mathbf{x}$,

$$r_N' = \inf_{p\in(0,1)} T_N'(p) \to r' = \inf_{p\in(0,1)} T'(p) > 0 \quad as\ m \to \infty \tag{A.12}$$

where $T'(p) = \int f_1(x)f_2(x)h^{-2}(x,p)h(x,p^*)\mu(dx) = \lim_{N\to\infty} T_N'(p), 0 < p < 1$.

Proof of Lemma A.2. It is an easy convexity fact that $r_N' = T_N'(p_{inf,N})$ and $r' = T'(p_{inf})$, where $p_{inf,N}$ and p_{inf} denote the inflexion points of $T_N'(p)$ and $T'(p)$, respectively, i.e. $T_N''(p_{inf,N}) = T''(p_{inf}) = 0$. Furthermore, it can be shown that r_N' and r' are in $(0,1)$. Let $b, 0 < b < 1/2$, be so small that $T'(b)$ and $T'(1-b)$ are > 1. Since $T_N''(p)$ and $T''(p), 0 < p < 1$, are increasing functions, Dini lemma implies that $||T_N'' - T''||_I = \sup_{p\in I}|T_N''(p) - T''(p)| \to 0$ as $N \to \infty$, where $I = [b, 1-b]$. Also, $T_N'(b)$ and $T_N'(1-b)$ are > 1 for all N large enough, and $p_{inf,N}$ and p_{inf} are in I.

Finally, $|T_N''(p_{inf,N}) - T''(p_{inf,N})| = |T''(p_{inf,N})| \le ||T_N'' - T''||_I \to 0$ as $m \to \infty$. Thus, $T''(p_{inf,N}) \to 0$ as $N \to \infty$, which implies that $p_{inf,N} \to p_{inf}$ as $N \to \infty$ and $r_N' = T_N'(p_{inf,N}) \to r' = T'(p_{inf})$ as $N \to \infty$. □

Lemma A.3 *For* $\mathbf{P_X}$ *- a.e.* $\mathbf{x}$ *and all* N *large enough,*

$$|S_N'(p)| \le (1/2)p^{-1}(1-p)^{-1} + cst.p^{-3/2}(1-p)^{-3/2}, 0 < p < 1, \tag{A.13}$$

where cst. denotes some positive constant.

Proof of Lemma A.3 Since

$$S_N' = p^{-1/2}(1-p)^{-1/2}(1-2p)\{T_N'(p)\}^{1/2} + (1/2)p^{1/2}(1-p)^{1/2} \{T_N'(p)\}^{-1/2}T_N''(p), \tag{A.14}$$

we have for all p in $(0,1)$ that

$$|S_N'(p)| \le (1/2)p^{-1}(1-p)^{-1} + (1/2){r_N'}^{-1/2}p^{-3/2}(1-p)^{-3/2}, \tag{A.15}$$

in view of (A.9) and (A.10). □

Lemma A.4 *(i) We have*

$$|r_N - r^*| = O(\ell(N)N^{-1/2}) \quad (N \to \infty). \tag{A.16}$$

(ii) We have

$$|\sigma_N - \sigma^*| = O(\ell(N)N^{-1/2}) \quad (N \to \infty). \tag{A.17}$$

Proof of Lemma A.4 Proof of (i): From the general theory of ML estimation, $|p_N - p^*| = O(\ell(N)N^{-1/2}) \quad (N \to \infty)$ $\mathbf{P_X}$ - a.s. Thus, for N large enough, we have in view of Lemma A.1 that

$$\begin{aligned} |r_N - r^*| &\le |T_N'(p_N) - T_N'(p^*)| + |T_N'(p^*) - T'(p^*)| \\ &\le O(\ell(N)N^{-1/2}) + |T_N'(p^*) - T'(p^*)|. \end{aligned} \tag{A.18}$$

But $T_N'(p^*) - T'(p^*)$ has the form $N^{-1}\Sigma_{1\le i\le N}U_i$, where the r.v.'s U_i are i.i.d., bounded and have mean 0. Thus, by the LIL, $|T_N'(p^*) - T'(p^*)| = O(\ell(N)N^{-1/2}) \quad (N \to \infty)$.

The proof of (ii) proceeds similarly. □

Proof of Lemma 2.2 By the duality principle of Diebolt and Robert (1992), it is enough to prove that the sequence $\{\mathbf{z}^{(m)}\}$ is ergodic. Since it is a finite-state homogeneous Markov chain, it is enough to prove that all the transition probabilities are positive. Now, define

$$\mathbf{A}_N = \left\{ \mathbf{z} \in \mathbf{Z} : N^{-1} \sum_{i=1}^{N} z_i \in J_N \right\}, \tag{A.19}$$

where $\mathbf{Z} = \{0,1\}^N$ has 2^N elements. If $\mathbf{a}$ and $\mathbf{b}$ are any elements of $\mathbf{A}_N$,

$$\begin{aligned}
&\mathbf{P}_\Omega\{\mathbf{z}^{(m+1)}\mathbf{b}|\mathbf{z}^{(m)} = \mathbf{a}\} = \\
&\mathbf{P}_\Omega\{\mathbf{z}^{(m+1)} = \mathbf{b}|\mathbf{z}^{(m+(1/2))} \in \mathbf{A}_N, \mathbf{z}^{(m)} = \mathbf{a}\}\ \mathbf{P}_\Omega\{\mathbf{z}^{(m+(1/2))} \in \mathbf{A}_N|\mathbf{z}^{(m)} = \mathbf{a}\} \\
&+\mathbf{P}_\Omega\{\mathbf{z}^{(m+1)} = \mathbf{b}|\mathbf{z}^{(m+(1/2))} \notin \mathbf{A}_N, \mathbf{z}^{(m)} = \mathbf{a}\}\ \mathbf{P}_\Omega\{z^{(m+(1/2))} \notin \mathbf{A}_N|\mathbf{z}^{(m)} = \mathbf{a}\},
\end{aligned} \tag{A.20}$$

with

$$\mathbf{P}_\Omega\{\mathbf{z}^{(m+(1/2))} \in \mathbf{A}_N|\mathbf{z}^{(m)} = \mathbf{a}) > 0 \tag{A.21}$$

and

$$\mathbf{P}_\Omega\{\mathbf{z}^{(m+(1/2))} \notin \mathbf{A}_N|\mathbf{z}^{(m)} = \mathbf{a}) > 0, \tag{A.22}$$

since all the states $\mathbf{z} \in \mathbf{Z}$ can be reached from $\mathbf{z}^{(m)}$ with positive probability (because $t(x,p)$ is in $(0,1)$ for all p in $(0,1)$). Moreover, $\mathbf{P}_\Omega\{\mathbf{z}^{(m+1)} = .|\mathbf{z}^{(m+(1/2))} \notin \mathbf{A}_N\}$ is a given probability distribution related to Γ_N, whereas

$$\begin{aligned}
\mathbf{P}_\Omega\{\mathbf{z}^{(m+1)} &= \mathbf{b}|\mathbf{z}^{(m+(1/2))} \in \mathbf{A}_N, \mathbf{z}^{(m)} = \mathbf{a}\} = \mathbf{P}_\Omega\{\mathbf{z}^{(m+(1/2))} = \mathbf{b}|\mathbf{z}^{(m)} = \mathbf{a}\} \\
&= \prod_{i=1}^{N} t(x_i, N^{-1}\sum_{j=1}^{N} a_j)^{b_i} \left\{1 - t(x_i, N^{-1}\sum_{j=1}^{N} a_j)\right\}^{1-b_i} > 0.
\end{aligned} \tag{A.23}$$

This completes the proof of Lemma 2.2. □

Proof of Lemma 3.1 The proof is very similar to that of Lemma 2.2, and is thus omitted. □

Proof of Lemma 3.3 We prove (3.18). By the quadratic Taylor formula, we have if $|h| \leq 2\epsilon_0$ and N is large enough that

$$\begin{aligned}
|T_N(p_N + h) - T_N(p_N) - hr_N| &\leq (1/2)h^2 \sup_{0\leq\theta\leq 1} |T_N''(p_N + \theta h)| \\
&\leq a_0 h^2,
\end{aligned} \tag{A.24}$$

for some constant a_0, in view of (3.17). If $|h| > 2\epsilon_0$, then, since

$$|\tilde{T}_N(p_N + h) - \tilde{T}_N(p_N) - hr_N| \le 3, \tag{A.25}$$

it is enough to choose $A_0 \ge a_0$ such that $A_0(2\epsilon_0)^2 \ge 3$. □

Proof of Lemma 3.4 It is similar to that of Lemma 3.3 except that we make use of a linear Taylor expansion rather than a quadratic one. □

Lemma A.5 *We have*

$$|p_{m+1} - p_m| \le m^{-1} J_c J_{obs}^{-1}(1 + O(1)) \quad (m \to \infty). \tag{A.26}$$

Proof of Lemma A5 By definition, $L'_m(p_m) = L'_{m+1}(p_{m+1}) = 0$. By a linear Taylor expansion of $L'_m(p)$,

$$\begin{aligned} L''_m(p_m)(p_{m+1} - p_m) \;+\; O(|p_{m+1} - p_m|^2) &= -(\partial/\partial p)\log h(x_{m+1,p})|_{p=p_{m+1}} \\ &= -(f_1 - f_2)(x_{m+1})h^{-1}(x_{m+1}, p_{m+1}). \end{aligned} \tag{A.27}$$

Since
$m^{-1}L''_m(p_m) \to -J_{obs}$ a.s.
and
$|f_1 - f_2|(x)h^{-1}(x,p) \le p^{-1}(1-p)^{-1} = J_c(p)$ for all x and p in $(0,1)$, the proof is complete. □

Proof of Lemma 4.1 Assertion (i) is straightfoward since under (H6) $\beta(m) = O(m^{-b})$ for some positive b. Assertion (ii) results from the following inequality:

$$e(m) = 1 - R(m) \ge cst.c(m) \ (m \to \infty). \tag{A.28}$$

We now turn to (iii). Since $1 - R(m) = e(m) \ge cst.m^{-\mu}$ for some positive cst.,

$$\begin{aligned} \log\left\{\prod_{k=1}^{m} R(k)\right\} + (3/2)\log m &= \sum_{k=1}^{m} \log(1 - e(k)) + (3/2)\log m \\ &\sim -\sum_{k=1}^{m} e(k) + (3/2)\log m \\ &\le -cst.(1-\mu)^{-1}m^{1-\mu} + (3/2)\log m \\ &\to -\infty \quad \text{as } m \to \infty. \end{aligned}$$

The proof of (4.13) is similar. $\square$

References

Agrawala, A. K. (1970). Learning with a probabilistic teacher. *IEEE Transactions on Information Theory*, **IT 16**, 373-379.

Biscarat, J-C. (1992). Almost sure convergence of a class of stochastic algorithms. Rapport Technique LSTA.

Biscarat, J-C., Celeux, G. and Diebolt, J. (1992). Stochastic versions of the EM algorithm. Rapport Technique LSTA 154.

Broniatowski, M., Celeux, G. and Diebolt, J. (1983). Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste. *Data Analysis and Informatics*, **3**, 359-374. North-Holland.

Celeux, G. and Diebolt, J. (1985). The SEM Algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quaterly*, **2**, 73-82.

Celeux, G. and Diebolt, J. (1986a). L'algorithme SEM : Un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *Revue de Statistique Appliquée*, **34**, 2, 35-52.

Celeux, G. and Diebolt, J. (1986b). Comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les mélanges de lois de probabilité. Rapport de Recherche INRIA 563.

Celeux, G. and Diebolt, J. (1987). A Probabilistic teacher algorithm for iterative maximum likelihood estimation. *Classification and Related Methods of Data Analysis*, editor: H.H. Bock, 617-623. North-Holland.

Celeux, G. and Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem. *Stochastics and Stochastics Reports.* (To appear)

Chauveau, D. (1991). Algorithmes EM et SEM pour un mélange caché et censuré. *Revue de Statistique Appliquée*, **39**, 4. (To appear)

Davisson, L. D. and Schwarz, S. C. (1970). Analysis of decision-directed receiver with unknown priors. *IEEE Transactions on Information Theory*, **IT 16**, 270-276.

Davydov, Y. A. (1973). Mixing conditions for Markov chains. *Theory of Probability and Applications*, **18**, 312-328.

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

Diebolt, J. and Robert, C. P. (1992). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society*, Ser. B. (To appear)

Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. 2. New York: Wiley.

Kazakos, D. (1977). Recursive estimation of prior probabilities using a mixtures. *IEEE Transactions on Information Theory*, **IT 23**, 203-211.

Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, Ser. B **44**, 226-233.

Makov, U. E. and Smith, A. F. (1977). A Quasi-Bayes unsupervised learning procedure for priors. *IEEE Transactions on Information Theory*, **IT 23**. 761-764.

Odell, P. L. and Basu, J. P. (1976). Concerning several methods for estimating crop acreages using remote sensing data. *Communications in Statistics (A)*, **5**, 1091-1114.

Owen, J. R. (1975). A Bayesian sequential procedure for quantal response in the context of adaptative mental testing. *Journal of the American Statistical Association*, **70**, 351-356.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**, 195-239.

Silverman, B. W. (1980). Some asymptotic properties of the probabilistic teacher. *IEEE Transactions on Information Theory*, **IT 26**, 246-249.

Sundberg, R. (1976). An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Communications in Statistics Simul. Comput. (B)*, **5**. 55-64.

Shorack, G. R. and Wellner J. A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association*, **82**, 528-550.

Titterington, D. M. (1984). Recursive parameter estimation using incomplete data *Journal of the Royal Statistical Society*, Ser. B **46**, 257-267.

Titterington, D. M., Smith A. F. and Makov U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley. *Journal of the Royal Statistical Society*, Ser. B **46**, 257-267.

Wei, G.C.G. and Tanner, M. A. (1990). A Monte-Carlo Implementation of the EM algorithm and the poor man 's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699-704.

Windham, M. P. and Cutler, A. (1991). Information ratios for validating cluster analyses. Working paper, Utah State University (Logan).

Wu, C. F. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, **11**, 95-103.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique.

ISSN 0249-6399