# UNSUPERVISED RESTORATION OF GENERALIZED MULTISENSOR HIDDEN MARKOV CHAINS

*Nathalie Giordana and Wojciech Pieczynski*
Département Signal et Image
Institut National des Télécommunications
9 rue Charles Fourier, 91000 Evry cedex France
e-mail: Nathalie.Giordana@int-evry.fr Wojciech.Pieczynski@int-evry.fr

## ABSTRACT

This work addresses the problem of generalized multisensor Hidden Markov Chain estimation with application to unsupervised restoration. A Hidden Markov Chain is said to be "generalized" when the exact nature of the noise components is not known; we assume however, that each of them belongs to a finite known set of families of distributions. The observed process is a mixture of distributions and the problem of estimating such a "generalized" mixture thus contains a supplementary difficulty: one has to label, for each state and each sensor, the exact nature of the corresponding distribution. In this work we propose a general procedure with application to estimating generalized multisensor Hidden Markov Chains.
**Key words :** multisensor data, generalized mixture estimation, Hidden Markov Chains, Bayesian restoration, unsupervised restoration.

## 1 INTRODUCTION

Hidden Markov Chains are well known as an efficient tool for treating numerous concrete problems. Such models have been successfully applied to speech processing problems [5], script recognition problems, image processing problems [1] and others. In a more general manner, one can envisage the use of theses models once the problem is to estimate some discrete phenomenon from the observed "noisy", i.e., continuous phenomenon. The noise is generally modelled as a realization of a Gaussian random variable. The unsupervised restoration of Hidden Markov Chains has been studied in the Gaussian case in [1], which is the origin of the present work. Nevertheless the Gaussian model can be unsuited to describe reality and one has to consider the use of other noise types; moreover, it is desirable to be able to find automatically the right nature of the noise for each class and each sensor. Such generalized mixtures have been previously studied assuming that the components lie in the Pearson system. An adaptation of the classical SEM algorithm [2] can be used to estimate such mixtures [3]. This work lies within the scope of this general problem. The organization of the paper is as follows:

In the next section we adress the generalized mixture estimation problem in a general setting and present ICE [4] and tests based methods of its estimation. The third section is devoted to the Hidden Markov Chain model. In section 4 we present a particular method for Generalized Hidden Markov Chain model estimation, based on the Kolmogorov Smirnov test. Section 5 is devoted to unsupervised restoration and presents some results.

## 2 GENERALIZED MIXTURE ESTIMATION

Let us consider a finite set $S$ and random variables $(X, Y) = ((X_s)_{s \in S}, (Y_s)_{s \in S})$. $X = (X_{s \in S})$ is the class random process: thus each $X_s$ takes its values in a finite set of classes $\Omega = \{\omega_1, \ldots, \omega_k\}$. $Y = (Y_{s \in S})$ is the observed process and each $Y_s$ takes its values in $\mathbb{R}^m$, where $m$ is the number of sensors; thus $Y_s = (Y_s^1, \ldots, Y_s^m)$. The distribution of $X$ depends on a parameter $\alpha$ and is denoted by $\pi_\alpha$. The random variables $Y_s^1, \ldots, Y_s^m$ are independent conditionally on $X$ and all distributions of $Y$ conditional on $X$ are given by $k$ distributions of $Y_s$ conditional on $X_s = \omega_1, \ldots, \omega_k$, respectively. The latter distributions are given by densities $f_1, \ldots, f_k$ with respect to the Lebesgue measure. The problem of mixture estimation is to find $\alpha$ and $f_1, \ldots, f_k$ from $Y = y$. In the "classical" mixture case the general form of $f_i$ is known and these densities depend on a parameter $\beta$ which has to be estimated from $Y = y$. For instance, if each $f_i$ is Gaussian, $\beta$ contains $k \times m$ means and $k \times m$ variances. In the "generalized" mixture case the general form of densities is not known exactly. However, the form of each $f_i$ is in a given finite set of forms. Let $\Psi = \{F_1, \ldots, F_M\}$ be a set of families of distributions. Thus each $f_i$ belongs to one of the families $F_1, \ldots, F_M$ and we do not know which one it is. The problem of finding the densities is then two-fold: for each $f_i$, find the family $F_l$ to which $f_i$ belongs and find the parameter which fixes $f_i$ in $F_l$. We propose a general algorithm called ICE-TEST based on the ICE algorithm ([4]), which in fact comprises a family of generalized mixture estimation methods. We assume the following:

(H1) One disposes of an estimator $\hat{\alpha} = \hat{\alpha}(X)$ of $\alpha$ from $X$;

(H2) It is possible to simulate realizations of $X$ according to its distribution conditional on $Y$;

(H3) Each family $F_l$ of $\Psi$ is parametrized with a parameter $\beta^l$;

(H4) One disposes of $M$ estimators $\hat{\beta}^1, \ldots, \hat{\beta}^M$ such that if a sample $z = (z_1, \ldots, z_r)$ is produced by a distribution $f_{\beta^j}$ in $F_j$, then $\hat{\beta}^j = \hat{\beta}^j(z)$ estimates $\beta^j$;

(H5) One disposes of a test which, given any distribution density $f_j$, checks the hypothesis "$f = f_j$" against the hypothesis "$f \neq f_j$".

For the multisensor case we add the following:
(H6) $Y_s^1, \ldots, Y_s^m$ are independent conditionally on $X_s$. Thus each density $f_i$ on $\mathbb{R}^m$ is given by $m$ densities $f_i^1, \ldots, f_i^m$ on $\mathbb{R}$ and $f_i = f_i^1 \times \ldots \times f_i^m$

The ICE-TEST algorithm is an iterative method: at step $q$, let $\alpha^q$ and $f_1^q, \ldots, f_k^q$ be current prior parameters and current densities. The updating is:

1. Simulate $x^q$, a realization of $X$, according to its $\alpha^q$ and $f_1^q, \ldots, f_k^q$ based distribution conditionnal to $Y = y$.
2. Calculate $\alpha^{q+1} = E_q[\hat{\alpha}(X) \mid Y = y]$, where $E_q[\,.\mid Y = y]$ is $\alpha^q$ and $f_1^q, \ldots, f_k^q$ based conditional expectation. If this calculation is impossible, calculate $\alpha^{q+1} = \hat{\alpha}(x^q)$
3. for $i = 1, \ldots, k$ consider $S_i = \{s \in S \mid x_s^q = \omega_i\}$. Let $y_i^q = (y_s)_{s \in S_i} = (y_s^1, \ldots, y_s^m)_{s \in S_i}$ and $y_i^{q,r} = (y_s^r)_{s \in S_i}$. For each sensor $r = 1, \ldots, m$ and each class $i = 1, \ldots, k$ calculate $M$ parameters $\beta_i^{1,r} = \hat{\beta}^1(y_i^{q,r}), \ldots, \beta_i^{M,r} = \hat{\beta}^M(y_i^{q,r})$.
4. for $r = 1, \ldots, m$ and $i = 1, \ldots, k$, consider that $y_i^{q,r} = (y_s^r)_{s \in S_i}$ is issued from a density $f$ and perform $M$ tests (for $j = 1, \ldots, k$) "$f = f_{\beta_i^j}$" against the hypothesis "$f \neq f_{\beta_i^j}$". Take $\beta_i^{r,q+1}$ as the first parameter giving the positive answer to the test.
5. For each class $i = 1, \ldots, k$ take $f_i^{q+1} = f_{\beta_1^{1,q+1}} \times \ldots \times f_{\beta_k^{M,q+1}})$ and update $(f_1, \ldots, f_k)$ with $(f_1^{q+1}, \ldots, f_k^{q+1})$.

We call our method ICE-TEST because it can be seen as a generalization of ICE. The latter method is a general method of estimation in case of hidden data, including classical mixture estimation.

## 3 HIDDEN MARKOV CHAINS

This section is devoted to a brief review of the Hidden Markov Chain model. We describe it in the mono-sensor case; the generalization to the multisensor case is immediate when the observations for a given class in different sensors are independent, by replacing in all formulas $f_i(y_s)$ by $f_i(y_s^1, \ldots, y_s^m) = f_i^1(y_s^1) \times \ldots \times f_i^m(y_s^m)$.
A sequence of random variables $X = (X_n)_{n \in N}$ taking their values in $\Omega = \{\omega_1, \ldots, \omega_k\}$ is a Markov random chain if it verifies for every $n \geq 1$:

$$P(X_{n+1} = \omega_{i_{n+1}} \mid X_n = \omega_{i_n} \ldots X_1 = \omega_{i_1}) = P(X_{n+1} = \omega_{i_{n+1}} \mid X_n = \omega_{i_n}) \qquad (1)$$

Then the distribution of $X = (X_n)_{n \in N}$ is given by the distribution of $X_1$, called the initial distribution, and a sequence of transition matrices $a_{ij} = P(X_{n+1} = \omega_j \mid X_n = \omega_i)$. In what follows we will assume that

$$c_{ij} = P(X_n = \omega_i, X_{n+1} = \omega_j) \qquad (2)$$

does not depend on $n$. Thus the initial distribution is given by

$$\pi_i = P(X_1 = \omega_i) = \sum_{j=1}^{k} c_{ij} \qquad (3)$$

and we have just one transition matrix $A = [a_{ij}]$, with

$$a_{ij} = \frac{c_{ij}}{\sum_{j=1}^{k} c_{ij}} \qquad (4)$$

Following the general hypotheses of Section 2, we will assume that the random variables $Y = (Y_s)_{s \in S}$ are independent conditionally on $X$ and that the distribution of each $Y_s$ conditional on $X$ is equal to its distribution conditional on $X_s$. We still denote by $f_1, \ldots, f_k$ the distribution densities of $Y_s$ conditional on $X_s = \omega_1, \ldots, X_s = \omega_k$, respectively. We denote by $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ the realizations of $X$ and $Y$, respectively.
Let us consider the Forward-Backward probabilities which will be used in estimation and restoration stages:

$$F_t(\omega_i) = P[X_t = \omega_i, Y_1 = y_1, \ldots, Y_t = y_t]$$
$$B_t(\omega_i) = P[Y_{t+1} = y_{t+1}, \ldots, Y_n = y_n \mid X_t = \omega_i]$$

$$F_1(\omega_j) = N_1 \pi_j f_j(y_1)$$
$$F_t(\omega_j) = N_t \left( \sum_{i=1}^{k} F_{t-1}(\omega_i) a_{ij} \right) f_j(y_t) \quad \text{for} \quad t > 1 \qquad (5)$$

$$B_n(\omega_i) = 1$$
$$B_t(\omega_i) = N_{t+1} \sum_{i=1}^{k} a_{ij} B_{t+1}(\omega_i) f_j(y_{t+1}) \quad \text{for} \quad t < n \qquad (6)$$

Here $N_t$ is a normalizing constant :

$$N_t = \left( \sum_{j=1}^{k} F_t(\omega_j) \right)^{-1}$$

## 4 ESTIMATION OF GENERALIZED HIDDEN MARKOV CHAINS

A Hidden Markov Chain is a type of general model described in section 2. Assuming (H6), we show in this section that the hypotheses (H1)-(H5) assumed in section 2 are also verified and then we develop the formulas adapted to the model.

1. The parameter $\alpha$ is here $c_{i,j}$, which can be estimated by

$$\hat{c}_{i,j} = \frac{\sum_{t=0}^{n-1} 1_{[X_t=\omega_i, X_{t+1}=\omega_j]}}{n-1} \quad (7)$$

Thus (H1) is verified.

2. The distribution of $X = (X_1, \ldots, X_n)$ conditional to $Y = y$, where $n$ is finite and fixed, is a distribution of a nonstationary Markov chain. The initial distribution and the transition matrix at time $t$ are given by

$$\pi_i^t = \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{k} \phi_t(i,j) \quad (8)$$

$$a_{ij}^t = \frac{\sum_{t=1}^{n-1} \phi_t(i,j)}{\sum_{t=1}^{n-1} \sum_{j=1}^{k} \phi_t(i,j)} \quad (9)$$

where $\phi_t$ is defined with (5) and (6):

$$\phi_t(i,j) = \frac{P(X_t=\omega_i, X_{t+1}=\omega_j, Y=y)}{P(Y=y)} = \frac{F_t(\omega_i) a_{ij} f_j(y_{t+1}) B_{t+1}(\omega_j)}{\sum_{l=1}^{k} f_l(y_{t+1}) \sum_{j=1}^{k} F_t(\omega_j) a_{jl}} \quad (10)$$

Thus simulations are possible; (H2) is verified.

3. There are numerous families verifying (H3). As an example let us consider $\Psi = \{\text{N,I,III}\}$; we discribe below these densities and their parameter estimations:
- type N: distribution $N(m, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(x-m)^2}{2\sigma^2}\right) \quad x \in \mathbb{R}$$

$$m = \mu_1 \quad \sigma^2 = \mu_2$$

- type I: distribution $B(p, q, a_1, a_2)$, $x \in [a_1\, a_2]$, $p, q > 0$

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(x-a_1)^{p-1}(a_2-x)^{q-1}}{(a_2-a_1)^{p+q-1}}$$

$$p, q = \frac{1}{2} r \left(1 \pm (r+2)\sqrt{\frac{\gamma_1}{(r+2)^2\gamma_2 + 16\,(r+1)}}\right)$$

$p > q$ if $\mu_3 > 0$ otherwise $q > p$

$$r = \frac{6\,(\gamma_2 - \gamma_1 - 1)}{(6 + 3\gamma_1 - 2\gamma_2)}$$

$$\Delta = a_2 - a_1 = \frac{1}{2}\sqrt{\mu_2}\sqrt{(r+2)^2\,\gamma_2 + 16\,(r+1)}$$

$$a_1 = \mu_1 - \Delta \frac{p}{p+q}$$

- type III: distribution $\Gamma(p, q, a_1)$, $x \geq a_1$, $p, q > 0$

$$f(x) = \frac{1}{\Gamma(p)\, q^p} (x-a_1)^{p-1} \exp\left(\frac{(a_1 - x)}{q}\right)$$

$$p = \frac{4\,\mu_2^3}{\mu_3^2} \quad q = \frac{\mu_3}{2\mu_2} \quad a_1 = \mu_1 - \frac{2\mu_2^2}{\mu_3}$$

Where $\mu_1$, $\mu_2$, $\mu_3$, $\mu_4$ are the first four moments, $\gamma_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\gamma_2 = \frac{\mu_4}{\mu_2^2}$.

4. To estimate the first four moments, one can use empirical moments

$$\hat{\mu_i}(X,Y) = \frac{\sum_{t=1}^{n} y_t 1_{[X_t=\omega_i]}}{\sum_{t=1}^{n} 1_{[X_t=\omega_i]}} \quad (11)$$

$$\hat{\mu}_{i,p}(X,Y) = \frac{\sum_{j=1}^{n} (y_j - \hat{\mu_i})^p 1_{[X_j=\omega_i]}}{\sum_{j=1}^{n} 1_{[X_j=\omega_i]}} \quad p > 1 \quad (12)$$

(H4) is verified.

5. We propose the use of the Kolmogorov-Smirnov test and accordingly call this algorithm the ICE-KOLM algorithm. The test runs as follows :
Let $y = (y_1, \ldots, y_n)$ be a sample issued from a density $f$. Let

$$\hat{G}_n(t) = \frac{1}{n} \sum_{t=1}^{n} 1_{]-\infty \quad y_i]}(t)$$

the empirical distribution function, $G_j$ the distribution function associated with the distribution $P_j$ (represented by a density $f_j$) and set $K_n = \sup_x \left|\hat{G}_n(x) - G_j(x)\right|$. For a given $\epsilon > 0$, let $W = \{y \in \mathbb{R}^n \,/ K_n(y) \leq c\}$ where $c$ is defined by $P_j(W) = \epsilon$ (for an $\epsilon$, $c$ is given by Kolmogorov's table). The hypothesis "$f = f_j$" is then accepted if $y \in W$, and rejected otherwise.

(H5) is verified.

## 5 UNSUPERVISED RESTORATION OF MARKOV CHAINS

### 5.1 Restoration

At the end of the estimation stage, the prior distribution and conditional distributions have been calculated; so we can perform the restoration. We have chosen the MPM algorithm which maximizes the marginal posterior probability:

$$X_t = \omega_j \Leftrightarrow P(X_t = \omega_j \mid Y_1 = y_1 \dots Y_n = y_n) = \max_{i \in \{1 \dots k\}} P(X_t = \omega_i \mid Y_1 = y_1 \dots Y_n = y_n) = \max_{i \in \{1 \dots k\}} F_t(\omega_i) B_t(\omega_i) \quad (13)$$

The unsupervised restoration runs as follows:

- Estimate $(\pi_i, a_{ij}, f_i)$
- For each $s = 1, \dots, n$:
  - Calculate $F_i, B_i$ using (5) and (6);
  - Classify $X_s = \omega_i$ where $\omega_i$ maximizes the posterior probability according to (13).

### 5.2 Numerical results

Let us consider two Markov Chains with two classes, chain A and chain B of different homogeneity. Some previous numerical results show that the homogeneity of the class can have a strong influence on the behavior of estimation and restoration stages.
- Chain A homogeneous. The transition matrix and initial distribution are:

$$A = \begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix} \qquad \pi = (0.5\ ,\ 0.5)$$

- Chain B non homogeneous. The transition matrix and initial distribution are:

$$A = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \qquad \pi = (0.5\ ,\ 0.5)$$

The sensor 1 is corrupted with $\Gamma(4;10;100)$ and N(180;400) and the sensor 2 is corrupted with $\Gamma(8;10;100)$ and B(7;3;0;200). We compare the ICE-KOLM algorithm with the classical one called ICE-GAUS which assumes that all densities are Gaussian. Results of estimation and kind of distribution detected are presented in table 1. The classification error rates after restoration are presented in table 2.
In view of the results presented in table 1, we observe that ICE-KOLM finds the correct distributions for chain A. In case of chain B, ICE-KOLM fails to detect the Beta distribution in sensor 2. The good estimation of the parameters improves the efficiency of the unsupervised Gaussian MPM restoration (see table 2). Moreover, the error rates of the MPM based on the ICE-KOLM estimates are very close to the error rates of the MPM based on the true parameters.

| Chain A | Sensor 1 | Sensor 2 |
|---|---|---|
| Real distribution | $\Gamma(4,10,100)$<br>N(180,400) | $\Gamma(8;10;100)$<br>B(7;3;0;200) |
| ICE-GAUS | N(140,379)<br>N(180,405) | N(179;855)<br>N(140;764) |
| ICE-KOLM | $\Gamma(3,5;12;98)$<br>N(180;400) | $\Gamma(10,2;8;95)$<br>B(6,3;2,3;36;198) |

| Chain B | Sensor 1 | Sensor 2 |
|---|---|---|
| Real distribution | $\Gamma(4,10,100)$<br>N(180,400) | $\Gamma(8;10;100)$<br>B(7;3;0;200) |
| ICE-GAUS | N(137,251)<br>N(180,397) | N(171;871)<br>N(139;810) |
| ICE-KOLM | $\Gamma(3,8;11;100)$<br>N(181;374) | $\Gamma(6,7;10;116)$<br>N(142;822) |

Table 1: 2 sensor generalized mixture estimation

| | Chain A | Chain B |
|---|---|---|
| Real distribution | 0.46% | 6.8% |
| ICE-GAUS | 3.14% | 9.1% |
| ICE-KOLM | 1.45% | 7.9% |

Table 2: Error rates after restoration

## References


[1] Benmiloud.B et Pieczynski.W. Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images. *Traitement du Signal*, 12(5):433–454, 1995.

[2] Celeux.G et Diebolt.J. L'algorithme SEM : Un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités. *Revue de Statistique Appliquée*, XXXIV(2), 1986.

[3] Delignon.Y, Quelle.HC. and Marzouki.A. Unsupervised Bayesian segmentation of SAR images using the Pearson system. *Proceedings of IGARSS'93 Tokyo*, pages 1538–1540, 1993.

[4] Pieczynski.W. Champs de Markov cachés et Estimation Conditionnelle Itérative. *Traitement du Signal*, 11(2):141–153, 1994.

[5] Rabiner.L.R, Levinson.S.E and Sondhi.M.M. On the use of hidden Markov model for speaker independant recognition of isolated words from a medium size vocabulary. *Bell systemTechnical Journal*, april 1984.