Semiparametric hidden Markov models: Identifiability and estimation

Jörn Dannemann[1], Hajo Holzmann[2] and Anna Leister[2]

[1]: *Institut für Mathematische Stochastik, Georg-August-Universität Göttingen, Germany*

[2]: *Fakultät für Mathematik und Informatik, Philipps-Universität Marburg, Germany*

We review the theory on semiparametric hidden Markov models (HMMs), that is, HMMs for which the state-dependent distributions are not fully parametrically, but rather semi- or nonparametrically specified. We start by reviewing identifiability in such models, where by exploiting the dependence much stronger results can be achieved than for independent finite mixtures. We also discuss estimation, in particular in an algorithmic fashion by using appropriate versions or modifications of the Baum-Welch (or EM-) algorithm. We present some simulation results and give an application to modeling bivariate financial time series, where we compare parametric with nonparametric fits for the state-dependent distributions as well as the resulting state-decoding.

Keywords: hidden Markov models, semiparametric modeling, log-concave densities, financial log-returns, EM-algorithm, finite mixture models

1 Introduction

A hidden Markov model (HMM) consists of an observed process $(Y_t)_{t\in\mathbb{N}}$ as well as a latent (unobserved) process $(X_t)_{t\in\mathbb{N}}$, such that 1. the $(Y_t)_{t\in\mathbb{N}}$ are independent given the $(X_t)_{t\in\mathbb{N}}$, 2. the conditional distribution of Y_s given the $(X_t)_{t\in\mathbb{N}}$ depends on X_s only and 3. $(X_t)_{t\in\mathbb{N}}$ is a finite-state Markov chain. We assume that $(X_t)_{t\in\mathbb{N}}$ is time-homogeneous. The cardinality K of the state space $\mathcal{X} = \{1,\dots,K\}$ of $(X_t)_{t\in\mathbb{N}}$ is then called the number of states of the HMM, and we denote the entries of the transition probability matrix (t.p.m.) by $(\alpha_{j,k})_{j,k=1,\dots,K}$. The t.p.m. is assumed to be irreducible and aperiodic, and its unique stationary distribution is denoted by π. The conditional distributions of Y_s given $X_s = k$, $k = 1,\dots,K$, are called the state-dependent distributions. We assume that they are independent of s. Since the marginal distributions of the $(Y_t)_{t\in\mathbb{N}}$ are finite mixtures in the state-dependent distributions, HMMs are also called Markov-dependent mixtures.

HMMs and related latent state models have important applications e.g. in speech recognition [40], ion channel modeling [5], biological sequence alignment [25], finance [36], and time series modeling in general [54].

Extensions of HMMs to non-finite state space as well as models where the $(Y_t)_{t\in\mathbb{N}}$ have a conditional time-series structure are discussed in [15, 24].

For statistical modeling, the state-dependent distributions are most often assumed to belong to some parametric family such as the normal distributions. In this case, the statistical theory is quite well-developed, see [44, 8, 24, 23] for theory on maximum likelihood estimation, or [27] for Bayesian methods.

Recently, there has been some interest in a semi- or even fully nonparametric specification of the state-dependent distributions, cf. [42, 43, 39, 17, 21] for some applications of such models. We shall call the resulting HMMs semiparametric HMMs. Semiparametric modeling may be of interest

a. to guard against misspecification,
b. as part of lack of fit tests for a parametric family of state-dependent distributions,
c. theoretically.

In this article we review the results on identifiability and estimation in such semiparametric HMMs. In Section 2 we discuss identifiability, and in particular review recent results which show the additional identifying power of a dependent regime, as compared to the case of independent mixtures. Section 3 is devoted to estimation in semiparametric HMMs. While there are few estimation methods which are theoretically justified, the popular EM-algorithm can readily be extended to such models. Thus, from an algorithmic point of view, the fitting of such models is relatively straightforward. Finally, we give an application to the modeling of a bivariate time series of financial log-returns, where we compare the performance of a simple parametric HMM with normal state-dependent distributions with that of a nonparametric alternative.

2 Identifiability of semiparametric hidden Markov models

When studying HMMs with semiparametrically specified component distributions, a major issue is identifiability. In the parametric case, [44] shows that full identifiability up to label switching follows from identifiability of the marginal finite mixtures (see [49, 34, 35] for results and references), since by an old result of [50] this implies identifiability of product distributions.

For HMMs with finite observational space, generic identifiability of the parameters, that is, identifiability except on a subset of the parameters space of Lebesgue measure zero, is achieved in [2], see also [47]. Note that identifiability completely fails for independent mixtures, thus, these results are a clear indication that using the dependence in an HMM, by considering the joint distribution of successive observations, much stronger results can be obtained.

When turning to semi - or nonparametrically specified component distributions, most research focuses on independent finite mixtures (sometimes with addition of covariates). Here, identifiability is most often only possible in specific, somewhat restrictive situations, in particular for location mixtures, often additional with the assumption of symmetric components, and mainly for two components. See [32, 33, 11, 51, 9, 10, 12, 37, 38, 31, 3, 4, 14, 13]. An exception are mixtures of product distributions, for which identifiability in the fully nonparametric case is possible under some assumptions, see [30, 41, 29, 2].

For HMMs, much stronger results can be obtained. From the joint distribution of two successive observations (Y_1, Y_2), in [28] the authors obtain full identifiability of the number of states K and of the state-dependent distributions $F(\cdot - \mu_k)$, $k = 1, \dots, K$ without further assumptions on F such as symmetry ($\mu_1 = 0$ for normalization), if the t.p.m. has full rank. Using the methods developed in [2], [26] obtained a general identifiability result. They showed that if the number of components K is known, and the t.p.m. has full rank K, and the component distributions are linearly independent, then from the joint distribution of three successive observations (Y_1, Y_2, Y_3) all parameters are identified.

[1] have the final word on nonparametric identifiability of HMMs. They show that if the t.p.m. has full rank, and if the state-dependent distributions are all distinct (not necessarily linearly independent), then all parameters including the number of components are fully nonparametrically identified (up to label switching).

3 Estimation

Concerning estimation, we may distinguish two general approaches. The first is to construct estimators, based on some particular identification result, for which theoretical properties such as consistency, rates of convergence and asymptotic normality can be established. The only method designed specifically for HMMs, in a situation which is not identified for independent mixtures seems to be the location model in [28].

The second approach which we shall focus on below is based on the EM-algorithm (see [22, 48]) or variants thereof. Theoretically, the EM-algorithm serves to compute the maximum likelihood (ML) estimator. However, since the theory of nonparametric ML estimation for HMMs still needs to be developed, estimation via the EM algorithm for semiparametric HMMs is currently a rather heuristic, numerical approach.

If the family of state-dependent distributions allows for a weighted nonparametric maximum likelihood estimator, an actual EM-algorithm may be formulated. In the mixture context, see [20, 16, 53]. Otherwise, the M-step may be modified to include e.g. kernel smoothing, such as in [9, 45, 18, 7],

For HMMs the EM-algorithm is also called the Baum-Welch algorithm, see [6]. [21] presents the following nonparametric EM-algorithm. Assume that the Markov chain is parametrized in all transition probabilities, without further restrictions. The parameter vector is denoted by

$$\vartheta = \big(\alpha_{1,1}, \dots, \alpha_{1,K-1}, \alpha_{2,1}, \dots, \alpha_{K,K-1}; f_1, \dots, f_K\big)^T,$$

where the f_k are the state-dependent densities. Introduce the notation $x_1^n = (x_1, \dots, x_n)^T$. The log-likelihood is then given by

$$L_n(\vartheta) = \log\Big(\sum_{x_1^n \in \mathcal{X}_1^n} p_n\big(x_1^n, Y_1^n; \vartheta\big)\Big),$$

$$p_n\big(x_1^n, Y_1^n; \vartheta\big) = \pi_{x_1}(\vartheta) \prod_{t=1}^{n-1} \alpha_{x_t, x_{t+1}}(\vartheta) \prod_{t=1}^{n} f_{x_t}(y_t; \vartheta).$$

For the EM-algorithm, consider the objective function

$$
\begin{aligned}
Q\big(\vartheta, \vartheta'\big) &= E_{\vartheta'}\big(\log(p_n(X_1^n, Y_1^n; \vartheta))|Y_1^n\big) \\
&= \sum_{k=1}^{K} \phi_{1|n}(k; \vartheta') \log \pi_k(\vartheta) + \sum_{t=1}^{n-1} \sum_{j,k=1}^{K} \phi_{t,t+1|n}(j,k;\vartheta') \log \alpha_{j,k}(\vartheta) \\
&+ \sum_{t=1}^{n} \sum_{k=1}^{K} \phi_{t|n}(k;\vartheta') \log f_k(Y_t),
\end{aligned}
$$

where

$$
\phi_{t|n}(k;\vartheta') = P_{\vartheta'}\big(X_t = k|Y_1^n\big), \qquad \phi_{t,t+1|n}(j,k;\vartheta') = P_{\vartheta'}\big(X_t = j, X_{t+1} = k|Y_1^n\big)
$$

which can be efficiently computed from the forward and backward probabilities, based on the parameter ϑ', see [15, 21] for a complete treatment. The formal M-step is now to compute the maximizer

$$
\vartheta^{(l+1)} = \operatorname{argmax}_{\vartheta} Q\big(\vartheta, \vartheta^{(l)}\big),
$$

which can be performed separately for the parameters of the t.p.m. as well as the state-dependent densities. For the t.p.m., nothing changes as compared to the fully parametric case. The explicit solution

$$
\alpha_{j,k}^{(l+1)} = \frac{\sum_{t=1}^{n-1} \phi_{t,t+1|n}(j,k;\vartheta^{(l)})}{\sum_{t=1}^{n-1} \sum_{k'=1}^{K} \phi_{t,t+1|n}(j,k';\vartheta^{(l)})}
$$

is available for fixed starting distribution, otherwise (e.g. in the stationary case), a numerical solution must be obtained. Now consider the state-dependent densities. If there are no structural parameters and a weighted MLE is available in $\mathcal{F}_k$, then one obtains

$$
f_k^{(l+1)} = \operatorname{argmax}_{f_k \in \mathcal{F}_k} \sum_{t=1}^{n} \phi_{t|n}(k;\vartheta^{(l)}) \log f_k(Y_t).
$$

Major examples are the classes of log-concave densities and of monotone densities, see [21]. A further example is the class of finite mixtures of some parametric family as state-dependent densities, see [36, 26, 52].

A prime example involving a structural parameter is a location-scale model, in which case one needs to resort to a ECM-algorithm (see [46]).

If the state-dependent densities, or the structural parameter f belong to some more general nonparametric class for which no nonparametric MLE is available, one may resort to EM-like algorithms. Here, the M-step is suitably replaced, and in some instances one obtains the decent property for a modified objective function.

Dannemann [21] considers a penalized version of the objective function, where the penalty is on the state-dependent densities. For an appropriate choice, this gives spline-type estimators of the state-dependent densities.

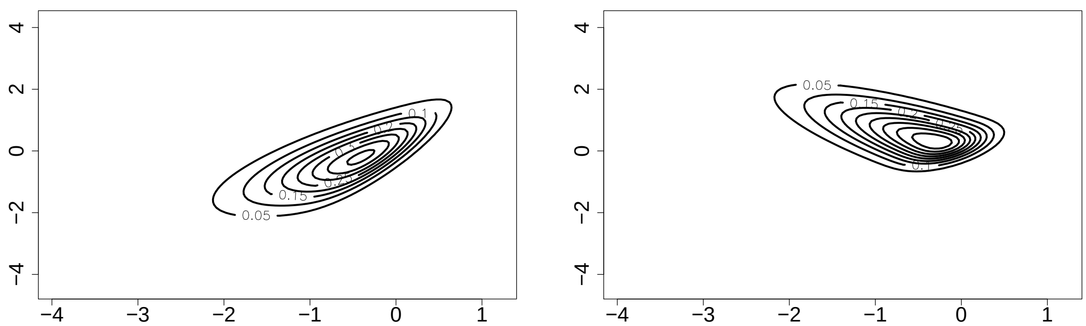
The M-step concerning the state-dependent densities can also be substituted by a smoothing step, see [39] for kernels or [17] for a wavelet-based approach. [26] propose a variant of the kernel method, for which they also show the ascent property for a suitable objective function.

4 Simulation

In this section we illustrate the flexibility of the nonparametric approach using simulated data. We generate 5000 bivariate observations from a two-state HMM with state dependent skew-normal distributions (contour plots are given in Figure 1) and transition probability matrix

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

If we use normal state dependent distributions we obtain a reasonable estimate of the



(a) State 1

(b) State 2

Figure 1: True parameter - Contour plot

transition probability matrix

$$\hat{A} = \begin{pmatrix} 0.74 & 0.25 \\ 0.49 & 0.51 \end{pmatrix},$$

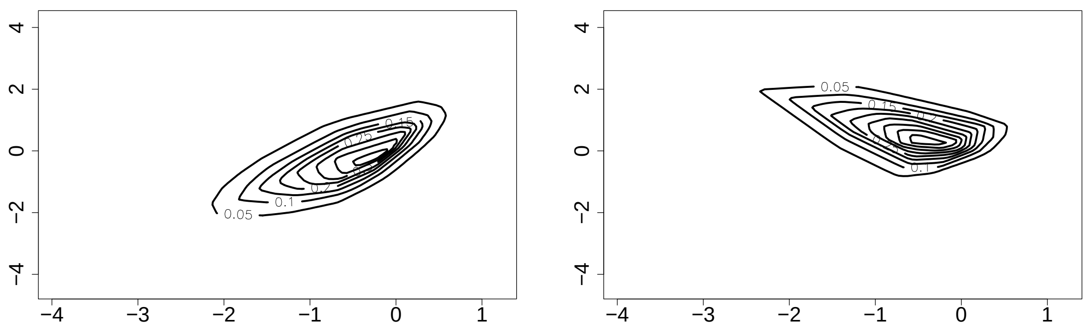
however, the state-dependent distribution cannot capture the skewness. When using a nonparametric approach assuming the state-dependent distributions to be log-concave based on the estimation procedure from [19] we estimate the transition probability matrix

$$\hat{A} = \begin{pmatrix} 0.73 & 0.27 \\ 0.37 & 0.63 \end{pmatrix},$$

contour plots for the state-dependent distributions are given in Figure 2

5 Application

In this section we use the EM-algorithm from section 3 to illustrate two choices of state dependent densities for a bivariate time series of financial logreturns. The states of the hidden Markov model may be interpreted as different volatility states.



(a) State 1

(b) State 2

Figure 2: Logconcave estimation - Contour plot

We use a database of 2926 logreturns from the stocks of Deutsche Bank and Munich RE on a daily basis from 21st January 2000 to 23rd April 2013[1] and fit a three state hidden Markov model using the EM-algorithm.
First we consider a parametric model with state-dependent normal distributions. The estimate for the transition probability matrix and the corresponding stationary distribution are given by

$$A = \begin{pmatrix} 0.807 & 0.193 & 8 \times 10^{-5} \\ 0.392 & 0.575 & 0.033 \\ 0.056 & 0.477 & 0.467 \end{pmatrix}, \qquad \pi = \begin{pmatrix} 0.659 & 0.321 & 0.020 \end{pmatrix}.$$

The estimated state dependent parameters are

$$\mu_1 = \begin{pmatrix} -1.80 \times 10^{-4} \\ 1.49 \times 10^{-4} \end{pmatrix}, \qquad \Sigma_1 = \begin{pmatrix} 1.8 \times 10^{-4} & 0.8 \times 10^{-4} \\ 0.8 \times 10^{-4} & 1.0 \times 10^{-4} \end{pmatrix},$$
$$\mu_2 = \begin{pmatrix} -2.4 \times 10^{-4} \\ -2.7 \times 10^{-4} \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 9.4 \times 10^{-4} & 5.0 \times 10^{-4} \\ 5.0 \times 10^{-4} & 6.7 \times 10^{-4} \end{pmatrix},$$
$$\mu_3 = \begin{pmatrix} 64.4 \times 10^{-4} \\ 52.9 \times 10^{-4} \end{pmatrix}, \qquad \Sigma_3 = \begin{pmatrix} 54.8 \times 10^{-4} & 29.0 \times 10^{-4} \\ 29.0 \times 10^{-4} & 39.6 \times 10^{-4} \end{pmatrix}.$$

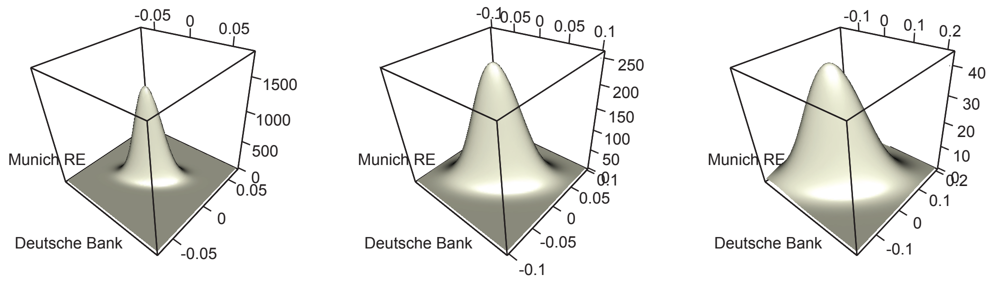
We observe that state 1 represents a state with low volatility whereas state three correspond to a high volatility state. The estimated densities are shown in Figure 3.

Next, we assume the state-dependent densities to be log-concave (without structural parameter) and compute the MLE by using the nonparametric version of the EM algorithm to obtain the estimated densities.

Here, the estimates from the parametric fit are used as starting values for the EM-algorithm in a nonparametric setting.

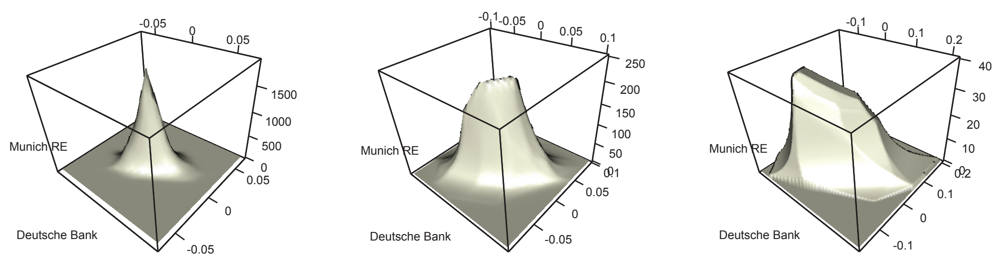
The estimation procedure for multivariate log-concave densities is described in [20]. For

[1] Data access from `http://de.finance.yahoo.com/`



(a) State 1 (b) State 2 (c) State 3

Figure 3: Parametric estimation



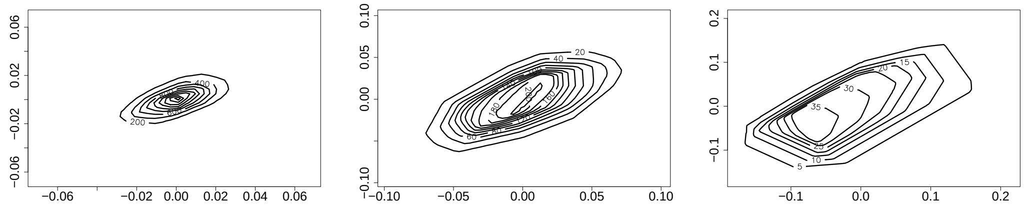
(a) State 1 (b) State 2 (c) State 3

Figure 4: Logconcave estimation

the evaluation of the maximum likelihood estimator in the M-step we use the **R** package *LogConcDEAD* introduced in [19]. The resulting transition probability matrix and stationary distribution are

$$A = \begin{pmatrix} 0.860 & 0.140 & 0.000 \\ 0.325 & 0.644 & 0.031 \\ 0.046 & 0.482 & 0.472 \end{pmatrix}, \qquad \pi = \begin{pmatrix} 0.689 & 0.294 & 0.017 \end{pmatrix}.$$

The estimated densities are shown in Figures 3 and 4, contour plots for the log-concave fit are given in Figure 5. The nonparametric fit is somewhat skewed, otherwise, the dispersion is similar to that in the parametric case. In Figure 6 we plot the time series together with a state-decoding using global decoding with the Viterbi-algorithm (see [15]), both from the nonparametric as well as from the parametric fit. The state-decoding based on the nonparametric fit has slightly less transitions than in the parametric case. However, both are still quite similar, which indicates that the parametric fit is reasonable.



(a) State 1 (b) State 2 (c) State 3

Figure 5: Logconcave estimation

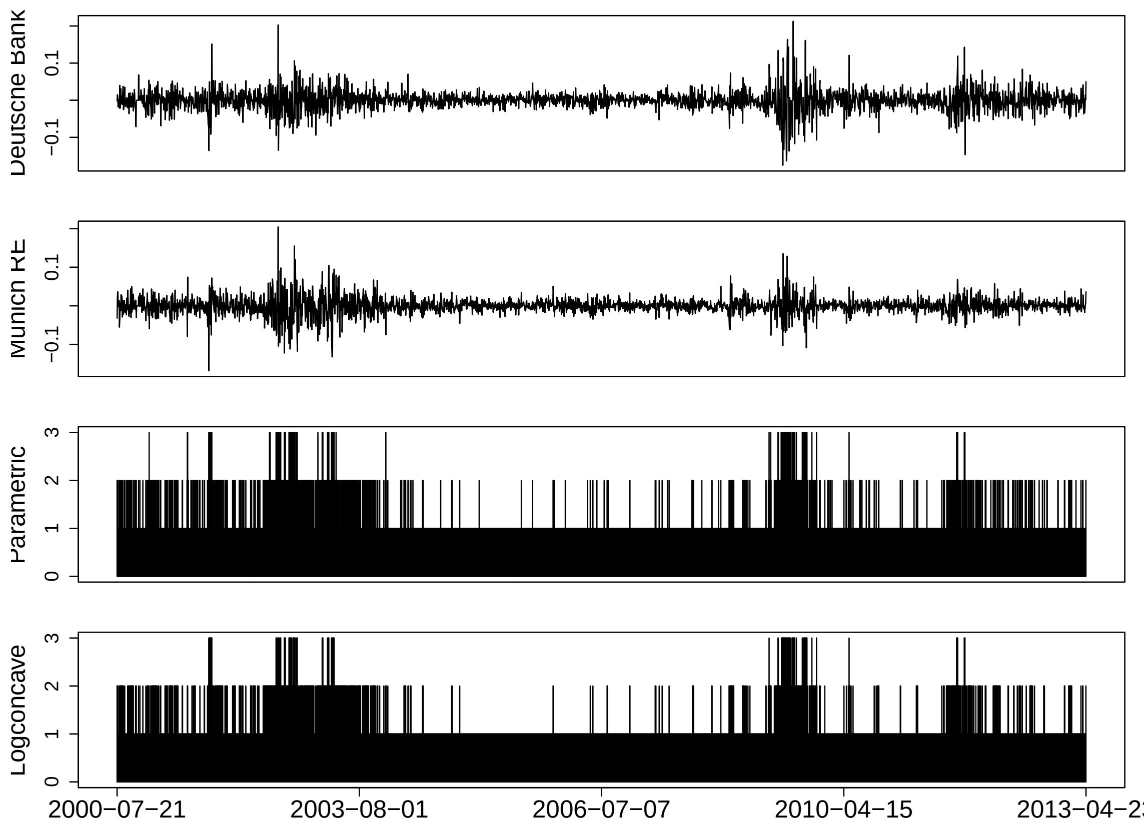


Figure 6: Upper two figures: Series of log-returns of the two stocks. Lower two figures: Global decoding using the Viterbi-algorithm based on parameter estimates from parametric fit (upper figure) and nonparametric log-concave fit (lower figure).

Acknowledgements

Anna Leister and Hajo Holzmann gratefully acknowledge financial support from the DFG, grant Ho 3260/3-2. The authors would like to thank two reviewers for helpful comments.

References

[1] Alexandrovich, G. and Holzmann, H. (2014). Nonparametric identification of hidden markov models. ArXiv:1404.4210 [math.ST].

[2] Allman, E. S., Matias, C. and Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Stat.*, **37** 3099–3132.

[3] Balabdaoui, F. and Butucea, C. (2014). On location mixtures with pólya frequency components. Preprint.

[4] Balabdaoui, F. and Doss, C. (2014). Inference for a mixture of symmetric distributions under log-concavity. Preprint.

[5] Ball, F. G. and Rice, J. A. (1992). Stochastic models for ion channels: Introduction and bibliography. *Math. Biosci.*, **112** 189–206.

[6] Baum, L., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, **41** 164–171.

[7] Benaglia, T., Chauveau, D. and Hunter, D. R. (2009). An em-like algorithm for semi- and nonparametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, **18** 505–526.

[8] Bickel, P. J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Stat.*, **26** 1614–1635.

[9] Bordes, L., Chauveau, D. and Vandekerkhove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Stat. Data Anal.*, **51** 5429–5443.

[10] Bordes, L., Delmas, C. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scand. J. Stat.*, **33** 733–752.

[11] Bordes, L., Kojadinovic, I. and Vandekerkhove, P. (2013). Semiparametric estimation of a two-component mixture of linear regressions in which one component is known. *Electron. J. Stat.*, **7** 2603–2644.

[12] Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Stat.*, **34** 1204–1232.

[13] Butucea, C., Nguyepe Zumpe, R. and Vandekerkhove, P. (2014). Semiparametric topographical mixture models with symmetric errors. *Preprint*.

[14] Butucea, C. and Vandekerkhove, P. (2013). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*.

[15] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in hidden Markov models.* New York, NY: Springer.

[16] Chang, G. T. and Walther, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Stat. Data Anal.*, **51** 6242–6251.

[17] Couvreur, L. and Couvreur, C. (2000). Wavelet-based non-parametric hmm's: theory and applications. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol. 1. 604–607 vol.1.

[18] Cruz-Medina, I. and Hettmansperger, T. (2004). Nonparametric estimation in semi-parametric univariate mixture models. *J. Stat. Comput. Simulation*, **74** 513–524.

[19] Cule, M., Gramacy, R. B. and Samworth, R. (2009). Logconcdead: An r package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software*, **29** 1–20.

[20] Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72** 545–607.

[21] Dannemann, J. (2012). Semiparametric hidden markov models. *Journal of Computational and Graphical Statistics*, **21** 677–692.

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

[23] Douc, R., Moulines, E., Olsson, J. and van Handel, R. (2011). Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Stat.*, **39** 474–513.

[24] Douc, R., Moulines, E. and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Stat.*, **32** 2254–2304.

[25] Durbin, R., Eddy, S. R., Krogh, A. and Mitchison, G. (1998). *Biological sequence analysis. Probabilistic models of proteins and nucleic acids.* Cambridge: Cambridge University Press.

[26] Gassiat, E., Cleynen, A. and Robin, S. (2013). Finite state space non parametric Hidden Markov Models are in general identifiable. *ArXiv e-prints*. `1306.4657`.

[27] Gassiat, E. and Rousseau, J. (2013). About the posterior distribution in hidden markov models with unknown number of states. To appear: Bernoulli.

[28] Gassiat, E. and Rousseau, J. (2013). Non parametric finite translation mixtures with dependent regime. *ArXiv e-prints*. `1302.2345`.

[29] Hall, P., Neeman, A., Pakyari, R. and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika*, **92** 667–678.

[30] Hall, P. and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Stat.*, **31** 201–224.

[31] Henry, M., Kitamura, Y. and Salanie, B. (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Papers 1767, Cowles Foundation for Research in Economics, Yale University.

[32] Hohmann, D. and Holzmann, H. (2013). Semiparametric location mixtures with distinct components. *Statistics*, **47** 348–362.

[33] Hohmann, D. and Holzmann, H. (2013). Two-component mixtures with independent coordinates as conditional mixtures: nonparametric identification and estimation. *Electron. J. Stat.*, **7** 859–880.

[34] Holzmann, H., Munk, A. and Gneiting, T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scand. J. Stat.*, **33** 753–763.

[35] Holzmann, H., Munk, A. and Stratmann, B. (2004). Identifiability of finite mixtures – with applications to circular distributions. *Sankhyā*, **66** 440–449.

[36] Holzmann, H. and Schwaiger, F. (2014). Hidden markov models with state-dependent mixtures: minimal representation, model testing and applications to clustering. *Statistics and Computing* 1–16. URL `http://dx.doi.org/10.1007/s11222-014-9481-1`.

[37] Hunter, D. R., Wang, S. and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Stat.*, **35** 224–251.

[38] Hunter, D. R. and Young, D. S. (2012). Semiparametric mixtures of regressions. *J. Nonparametric Stat.*, **24** 19–38.

[39] Jin, N. and Mokhtarian, F. (2006). A non-parametric hmm learning method for shape dynamics with application to human motion recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 2. 29–32.

[40] Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics*, **33** 251–272.

[41] Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, **77** 135–175.

[42] Lambert, M. F., Whiting, J. P. and Metcalfe, A. V. (2003). A non-parametric hidden markov model for climate state identification. *Hydrology and Earth System Sciences*, **7** 652–667.

[43] Lefèvre, F. (2003). Non-parametric probability estimation for hmm-based automatic speech recognition. *Computer Speech & Language*, **17** 113 – 136. New Computational Paradigms for Acoustic Modeling in Speech Recognition.

[44] Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes Appl.*, **40** 127–143.

[45] Levine, M., Hunter, D. and Chauveau, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika*, **98** 403–416.

[46] Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, **80** 267–278.

[47] Petrie, T. (1969). Probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, **40** pp. 97–115.

[48] Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, **26** 195–239.

[49] Teicher, H. (1963). Identiability of finite mixtures. *Ann. Math. Stat.*, **34** 1265–1269.

[50] Teicher, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Stat.*, **38** 1300–1302.

[51] Vandekerkhove, P. (2013). Estimation of a semiparametric mixture of regressions model. *J. Nonparametric Stat.*, **25** 181–208.

[52] Volant, S., Bérard, C., Martin-Magniette, M.-L. and Robin, S. (2013). Hidden markov models with mixtures as emission distributions. *Statistics and Computing* 1–12.

[53] Walther, G. (2009). Inference and modeling with log-concave distributions. *Statistical Science*, **24** 319–327.

[54] Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov models for time series. An introduction using R.*