

the distribution defined by

$$p(x_1, u_1, y_1, x_2, u_2, y_2) = p(u_1, u_2) p(y_1 | u_1) p(x_1 | u_1) p(x_2 | u_2) p(y_2 | u_2). \quad (3.1)$$

Then the transitions are

$$p(x_2, u_2, y_2 | x_1, u_1, y_1) = p(u_2 | u_1) p(x_2 | u_2) p(y_2 | u_2). \quad (3.2)$$

We call this model “simplified TMM” (STMM) and we observe that a STMM (*cf.* Fig. 5) is not a PMM; for example, we have $p(u_2 | x_1, y_1, y_2, x_2) \neq p(u_2 | y_2, x_2)$. We announced earlier that the objective of the paper was to provide a numerical comparison among different PMMs; however, we also provide a comparison between STMM and the classic HMM-IN.

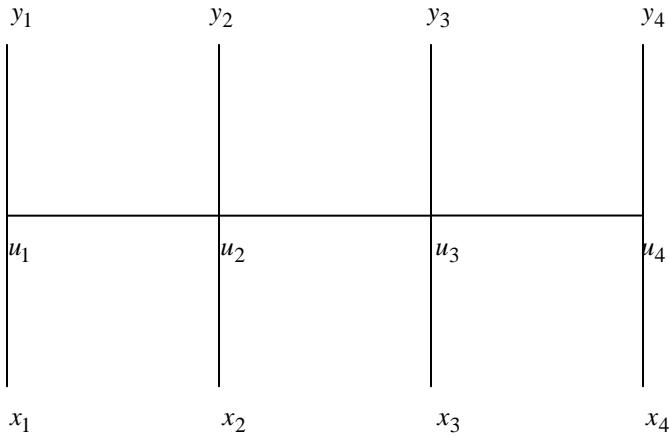


Fig. 5. The dependence graph of the STMM.

Remark 2

The TMMs are strictly more general than PMMs. In fact, one can use the result from [9] for the case of the stationary reversible systems. Let $\mathbf{T} = (\mathbf{X}, \mathbf{U}, \mathbf{Y})$ be a TMM; we set $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ to make appear $\mathbf{T} = (\mathbf{U}, \mathbf{Z})$ as a PMM. Thus, $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ is Markovian if and only if $p(u_2 | x_1, y_1, y_2, x_2) = p(u_2 | y_2, x_2)$. Thus, we obtain a reversible stationary TMM $\mathbf{T} = (\mathbf{X}, \mathbf{U}, \mathbf{Y})$ with non Markovian (\mathbf{X}, \mathbf{Y}) by choosing a distribution such that $p(u_2 | x_1, y_1, y_2, x_2) \neq p(u_2 | y_2, x_2)$.

4 Experiments

4.1 Pairwise Markov models

We present different experiments to compare PMMs-CN, PMMS-IN, HMMs-CN and HMMs-IN from Section II. We decide to set $\Omega = \{\omega_1, \omega_2\}$ for the sake of simplicity.

We begin with sampling data points of the most inclusive model (PMM-CN), for which we choose

$$p(x_1, x_2) = 1_{[x_1=x_2]}(0.5 - \varepsilon) + 1_{[x_1 \neq x_2]}\varepsilon;$$

$$p(y_1, y_2 | x_1, x_2) = \mathcal{N} \left(\begin{bmatrix} \mu_1^{x_1, x_2} \\ \mu_2^{x_1, x_2} \end{bmatrix}, \begin{bmatrix} (\sigma_1^{x_1, x_2})^2 & \rho \\ \rho & (\sigma_2^{x_1, x_2})^2 \end{bmatrix} \right).$$

The coefficients ε (the probability of regime-switching) and ρ (the conditional correlation) depend on the experimental setting. The values of the remaining parameters are per each pair (x_1, x_2) are in the table below.

(x_1, x_2)	$\mu_1^{x_1, x_2}$	$\mu_2^{x_1, x_2}$	$\sigma_1^{x_1, x_2}$	$\sigma_2^{x_1, x_2}$
(ω_1, ω_1)	-5	-5	14	14
(ω_1, ω_2)	-3	3	7	9
(ω_2, ω_1)	3	-3	9	7
(ω_2, ω_2)	5	5	20	20

Next, we compute \mathbf{X} from \mathbf{Y} by using the PMM backward-forward algorithm for the “projections” of the known parameters into each sub-model. Finally, we compute a relative error rate, referring to the PMMs-CN (in percents):

$$\tau = \frac{\text{err}_{\text{model}} - \text{err}_{\text{PMM-CN}}}{\text{err}_{\text{PMM-CN}}}. \quad (4.1)$$

For example, if the relative error rate reaches 100%, then it means that the reference model decreases the misclassification percentage by a half when compared to the proposal one.

The values in Table 1 are the relative error rates (in percents) of the three models compared to PMMs-CN for various probabilities of regime-switching changes. Each value is averaged over the values of the correlation coefficient.

The values in Table 2 are the relative error rates of the three models compared to PMMs-CN for various correlation coefficients. Each value is averaged over the

values of the probabilities of regime-switching.

We simulate 40 random chains for each pair of (ϵ, ρ) and 1000 elements per each chain.

ϵ	Model compared to PMM-CN		
	HMM-IN	HMM-CN	PMM-IN
0.025	40.7	4.9	35.7
0.125	34.9	11.0	18.5
0.275	55.5	13.6	21.1
0.425	57.2	13.3	48.8
Average	46.8	11.6	27.0

Table 1. Relative error rates (4.1) of the three models for varying probability of regime-switching.

ρ	Model compared to PMM-CN		
	HMM-IN	HMM-CN	PMM-IN
0.05	19.3	10.2	0.3
0.25	22.3	10.1	3.6
0.75	56.9	12.4	37.9
0.95	142.0	17.6	112.7
Average	60,1	12,6	38,6

Table 2. Relative error rates (4.1) of the three models for varying noise correlation coefficient.

These tables can be seen as “cuts” of the 3-D surface plot from Fig. 6.

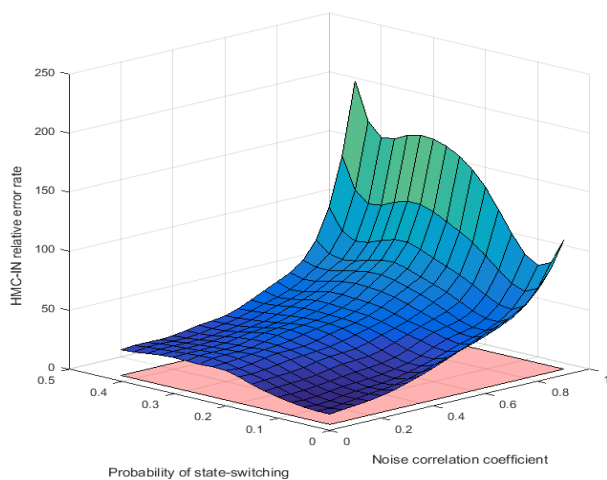


Fig. 6. Surface plot of the relative error rate. The height is a function which assigns to each pair (ϵ, ρ) the corresponding relative error (4.1).

4.2 STMM compared to the HMM-IN

In this sub-section we investigate if the STMM from Section 3 is competitive with the HMM-IN.

We decide to set $\Omega = \{\omega_1, \omega_2\}$ for \mathbf{X} , $\Lambda = \Omega$ for \mathbf{U} . We choose the state space distribution:

$$p(u_1, u_2) = 1_{[u_1=u_2]}0.49 + 1_{[u_1 \neq u_2]}0.01;$$

$$p(x_1|u_1) = 1_{[x_1=u_1]}0.99 + 1_{[x_1 \neq u_1]}0.01;$$

Regarding the observation space, we have

$$p(y_1|u_1 = \omega'_1) = N(y_1, 10, \sigma);$$

$$p(y_1|u_1 = \omega'_2) = N(y_1, 20, \sigma).$$

Then we compute the error rates relative to the HMM-IN- and STMM-based MPM state estimations for various values of σ . We present our results in Fig. 7. They appear promising enough to be worth researching.

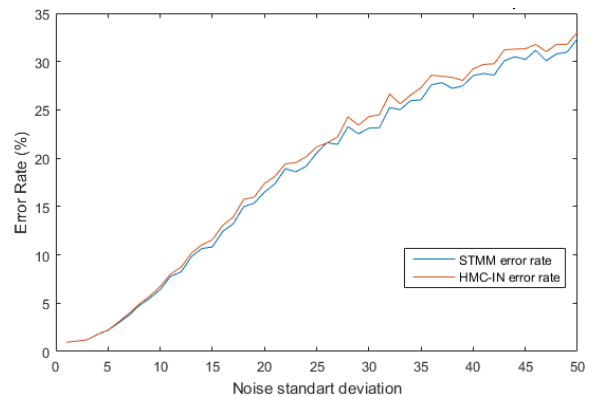


Fig. 7. Comparison between the performances of STMM and HMM-IN for various noise levels.

5 Conclusion

The primary objective of the paper is to compare efficiencies of Bayesian classifiers based on four different models: the classic HMM, the general PMM, and two intermediary models. Different results of experiments, some of which are presented in the paper, show that the PMMs potentially outperform the HMMs in “real-world” applications. Indeed, the PMM allows reducing the misclassification ratio by 10%-30% and even more. Such a gap is particularly visible if the observation noise is heavily correlated and if the hidden chain is too far from being Markovian. We also studied an example of a simplified triplet Markov chain, as simple as a HMM but very different from the latter.

The further work will include exploring more advanced TMMs, PMMs [12, 13] and their inter-comparisons with the on a similar methodology basis.

References:

- [1] O. Cappé, E. Moulines, T. Rydén, *Inference in Hidden Markov Models*, Springer Verlag, 2005.
- [2] Y. Ephraim and N. Merhav, Hidden Markov processes, *IEEE Transactions on information theory*, Vol. 48, No. 6, 2002, pp. 1518-1569.
- [3] R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286.
- [4] R. Bhar and S. Hamori, *Hidden Markov Models: Applications to Financial Economics*, Springer Science & Business Media, 2006
- [5] R. Mamon and R. Elliott, *Hidden Markov Models in Finance*, Springer, 2007.
- [6] T. Koski, *Hidden Markov Models for Bioinformatics, Computational Biology*, Springer Science & Business Media, 2001.
- [7] M. Vidyasagar, *Hidden Markov Processes: Theory and Applications to Biology, Princeton Series in Applied Mathematics*, Princeton University Press, 2014.
- [8] S. Derrode and W. Pieczynski, Unsupervised classification using hidden Markov chain with unknown noise copulas and margins, *Signal Processing*, vol. 128, 2016, pp. 8-17.
- [9] S. Derrode and W. Pieczynski, Signal and image segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, Vol. 52, No. 9, 2004, pp. 2477-2489.
- [10] L. Baum, T. Petrie, G. Soules and N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *The annals of mathematical statistics*, vol. 41, No. 1, 1970, pp. 164-171.
- [11] P. Lanchantin, J. Lapuyade-Lahorgue and W. Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise, *Signal Processing*, Vol. 91, No. 2, 2011, pp. 163-175.
- [12] S. Carincotte, S. Derrode and S. Bourennane, Unsupervised change detection on SAR images using fuzzy hidden Markov chains, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 2, 2006, pp. 432-441.
- [13] F. Salzenstein, C. Collet, S. Cam and M. Hatt, Non-stationary fuzzy Markov chain, *Pattern Recognition Letters*, vol. 28, no 16, 2007, pp. 2201-2208.