Chapter 4

Triplet Markov chains and image segmentation

A number of Markov models have been shown to be remarkably effective for a variety of modelization problems and treatment of a wide range of phenomena. The use of these models is very much on the increase in economics, finance, genomics, ecology, communications, signal and image processing, etc. In particular, hidden Markov models (HMM) are well known for their effectiveness for treating the problem of segmentation, which is among the most prominent and difficult problems in image processing. In such a model, the hidden data, which model the desired segmented image, are considered as the realization of a Markov process, which may be a field, a tree or a chain. More generally, HMMs are used to treat other inverse problems in imagery such as noise removal or contour detection – see Chapter 1. The distribution $p(x)$ of the hidden process X, referred to as the *a priori* distribution, can generally be interpreted as a *regularization* tool for the unobservable image which is being determined. The distribution of Y, conditional on X, is known as the data-driven distribution. Thus the various extensions to HMM presented in this chapter are even more general tools, applicable to a diverse range of inverse problems in imagery and vision.

The aim of this chapter, which takes a pragmatic approach accessible to readers who are not familar with HMMs, is to present a number of generalizations of HMMs which have emerged since the year 2000. The first generalization involves the introduction of "pairwise" Markov models (PMMs), in which the Markovian nature of the (hidden process, observed process) pair is considered. PMMs are strictly more general than HMMs, and it turns out that this property results in a greater efficiency of the treatments. A second stage of generalization involves the introduction of an auxiliary process, and considering the (hidden process, auxiliary process, observed process)

Chapter written by Wojciech PIECZYNSKI.

triplet. These are known as triplet Markov models (TMM), which show great potential in terms of their generality. In particular, they make it possible to treat non-stationary data, or even "long-memory" data, modeled as a semi-Markov process.

We will restrict ourselves to a discussion of the simplest Markov processes, which are chains. Some of the properties discussed can easily be extended to Markov fields and trees, and the possibilities of generalizing the other properties are interesting avenues of research.

All the treatments are discussed in a non-supervised context, giving them immediate value to an interested user.

4.1. Introduction

When we are interested in determining numerical quantities $x = (x_1 \ldots, x_n)$ which are not directly accessible, based on observable numerical quantities $y = (y_1, \ldots, y_n)$, probability theory offers a rigorous framework which makes it possible to derive rigorous results that are generally effective – at times spectacularly so. The (x, y) pair is treated as a realization of two random processes (X, Y), with $X = (X_1 \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$, and the links between x and y are modeled by the probability distribution $p(x, y)$ for the (X, Y) pair. The power of probabilistic methods stems from the fact that, despite the absence of deterministic links between x and y, it is possible to propose methods of finding values for x which are optimal "on average", or "in the long term", when the problem is treated a "large" number of times. Some of these methods can be implemented in the case where it is possible to calculate the X_i distributions conditional on $Y = y$, which we will write as $p(x_i|y)$, something which will be among the main objectives of the various models studied in this chapter. Suppose that each X_i takes its values from a finite set $\Omega = \{\omega_1, \omega_2\}$, and each Y_i takes its values from $\mathbb{R}$. The (X, Y) pair then takes is values from $\Omega^n \times \mathbb{R}^n$. Suppose that its distribution $p(x, y)$ is given by $p(x, y) = p(x)p(y|x)$, with $p(x)$ being a distribution over Ω^n and $p(y|x)$ being densities over R^n with respect to Lebesgue measure. When n is not too large, the marginal distributions $p(x_i, y)$ giving the desired distributions $p(x_i|y)$ can be calculated from $p(x, y)$, which can then be used in its most general form. However, this calculation requires 2^{n-1} operations, and rapidly becomes impossible as n increases. It is then necessary to restrict ourselves to distributions $p(x, y)$ which take particular forms that allow the calculation of $p(x_i|y)$, $1 \leq i \leq n$ for large n. In the standard model, which we will call a "hidden Markov chain with independent noise" (HMC-IN), $p(x, y)$ is given by:

$$p(x, y) = p(x_1)p(x_2|x_1) \ldots p(x_n|x_{n-1})p(y_1|x_1) \ldots p(y_n|x_n). \tag{4.1}$$

We will retain this type of notation, which is very concise and useful, noting that it contains notations which are familiar in the treatment of signals and images. For

example, suppose that $\Omega = \mathbb{R}$, let $X_0 = \zeta_0$, and consider the following equations:

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_i = AX_{i-1} + \zeta_i \end{cases}, \tag{4.2}$$

with the variables $\varepsilon_1, \ldots, \varepsilon_n, \zeta_0, \zeta_1, \ldots, \zeta_n$ being centered and mutually independent. This is a standard AR first-order hidden process with additive noise, which belongs to the set of models described in Equation (4.1).

The term HMC-IN stems from the fact thats, firstly, the hidden chain X is a Markov chain, and secondly the random variables $Y_1, \ldots, Y_n$ are independent conditional on $X = x$. Although the HMC-IN model appears to be a fairly sophisticated model, it is one of the simplest models that allows us to take useful account of the mutual dependence between the various variables making up the two processes $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$. In fact, the Markovian distribution $p(x) = p(x_1)p(x_2|x_1)\ldots p(x_n|x_{n-1})$ for X is in a sense the simplest of the distributions which allow the introduction of the dependence between the variables $X_1, \ldots, X_n$. The distribution $p(y|x) = p(y_1|x_1)\ldots p(y_n|x_n)$ is also very simple. Nevertheless, this type of model, which is very widely used, can give spectacular results in a number of different fields. Its attraction lies in the possibility of calculating, with a number of elementary operations proportional to n, several quantities of interest including $p(x_i|y)$. By setting $\alpha_i(x_i) = p(y_1, \ldots, y_{i-1}, y_i, x_i)$, which we will refer to as the "forward probability", and $\beta_i(x_i) = p(y_{i+1}, \ldots, y_n|x_i)$, which we will call the "backwards probability", we can demonstrate the well-known result that: $p(x_i, y) = \alpha_i(x_i)\beta_i(x_i)$. In addition α_i and β_i can be calculated recursively by:

$$\alpha_1(x_1) = p(x_1, y_1), \text{ and}$$

$$\alpha_{i+1}(x_{i+1}) = \sum_{x_i \in \Omega} \alpha(x_i)p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}) \text{ for } 2 \leq i \leq n-1, \tag{4.3}$$

$$\beta_n(x_n) = 1, \text{ and}$$

$$\beta_i(x_i) = \sum_{x_{i+1} \in \Omega} \beta_{i+1}(x_{i+1})p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}) \text{ for } 1 \leq i \leq n-1, \tag{4.4}$$

and the recursive nature of these operations allows $p(x_i, y)$ to be calculated for very large n (several million...).

This ability to extend the calculation, as well as the fact that $p(x_i|y)$, which depends on all the components $y_1, \ldots, y_n$ of y, uses all the available information, is what has given rise to the success, and often great effectiveness, of the HMC-IN model. Its applications are countless, and thousands of papers have been published on this subject. We draw the reader's attention to a few recent publications in the

fields of bioscience [KOS 01, NIC 02, NUE 07], climatology [BEL 00], communications [CAP 05], ecology [LEB 06], econometrics and finance [GRE 00, THO 02], handwriting recognition [CHE 94], music [RAP 99], and images [GIO 97, MAI 01]. However, this well-established model remains relatively rudimentary, and its simplicity, particularly in terms of the distribution $p(y|x) = p(y_1|x_1) \dots p(y_n|x_n)$, can prove to be inadequate in a number of situations. To solve this problem, it is possible to use more general models, known as pairwise Markov chains or PMCs [DER 04, PIE 03]. In these, the Markovian nature of the paired process $Z = (X, Y)$ is directly assumed. This retains the same ability to recover the hidden signal, while permitting more complete modelization of $p(y|x)$. Finally, PMCs can in turn be generalized to triplet Markov chains (TMCs [PIE 02]), in which a third random chain $U = (U_1, \dots, U_n)$ is introduced, which may or may not have a physical significance, and the Markovian nature of the triplet $T = (X, U, Y)$ is considered. Finally, we should mention "hidden Markov chains" (HMCs) which we will define as PMCs $Z = (X, Y)$ where X is a Markov chain (the hidden process is Markovian, which leads to the name HMC).

Finally, four families of models are obtained with strictly increasing degrees of generality: HMC-IN, HMC, PMC and TMC. We note that in addition to PMC and TMC, HMMs have been generalized in other directions; however, to our knowledge, all such extensions retain the Markovian nature of X and hence are still HMMs. PMCs form a family which is strictly larger than HMC because PMCs exist such that X is not Markovian [PIE 07]. Similarly, TMCs $T = (X, U, Y)$ exist such that (X, Y) is not a Markov chain. Through the flexibility available in choosing the third process U, TMCs offer a wide range of possibilities for the definition – and use – of the $p(x, y)$ distributions, and the aim of this chapter is to describe recent advances in the use of these models. In particular, we will see that it is possible to use $p(x, y)$ distributions which cannot be expressed analytically, which is probably one of the most surprising aspects of TMC. The various modelizations and treatments presented in this chapter can be used in many applications; in fact, any of the HMC, PMC and TMC models can be applied anywhere that standard HMC-IN models – or their standard extensions – can be applied. The aim of this chapter is to describe these various extensions, to explain their interest for solving a number of practical problems where standard models reach their limits, and to present some results of their application to statistical image segmentation.

The organization of the chapter is as follows. PMCs are introduced in the next section. Section 4.3 is dedicated to describing recent uses of copulas in the framework of PMCs [BRU 05a, BRU 05b]. This section, which aims to show the rich possibilities inherent in the modeling of $p(y|x)$ noise, can if desired be skipped on first reading. Section 4.4 covers the estimation of parameters from incomplete data. We will present the method of "iterative conditional estimation" (ICE), widely used due to its ease of implementation and good behavior, with PMCs and TMCs. This section is independent of the subsequent ones, and can also be skipped by a reader who is mostly interested in PMC and TMC models. Triplet Markov chains are then introduced in

Section 4.5, and the subsequent sections discuss various applications. In Section 4.6, we will describe the use of TMCs for the treatment of non-stationary processes, and then Section 4.7 treats the use of TMCs in the modeling of the semi-Markovian nature of the hidden process. Section 4.8 presents various possibilities of simultaneous modeling of various aspects discussed in the previous sections, through the use of TMCs which make use of auxiliary "multivariate" chains. Finally, the ninth and final section contains conclusions and future directions.

4.2. Pairwise Markov chains

4.2.1. *General model*

Consider $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$, two random processes, where each X_i draws its values from a finite ensemble $\Omega = \{\omega_1, \ldots, \omega_k\}$ and each Y_i takes its values from the set $\mathbb{R}$ of real numbers. Let $Z = (Z_1, \ldots, Z_n)$, with $Z_i = (X_i, Y_i)$. The process Z will be referred to as the pairwise Markov chain (PMC) if it is itself Markovian, and as a "hidden Markov chain" (HMC) if it is a PMC such that X is Markovian. Thus Z is a PMC if and only if its distribution takes the following form (in order to keep our notation compact, we will use z_i in place of (x_i, y_i) wherever possible, and z in place of (x, y)):

$$p(z) = p(z_1)p(z_2|z_1)\ldots p(z_n|z_{n-1}). \tag{4.5}$$

Given that the transitions in 4.5 can be written:

$$p(z_{i+1}|z_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i), \tag{4.6}$$

we can see that the standard HMC-IN model given by (4.1) is a PMC (4.5) where the transitions satisfy $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$ and $p(y_{i+1}|x_{i+1}, x_i, y_i) = p(y_{i+1}|x_{i+1})$. We can also see that the first of these equalities is a sufficient condition for a PMC to be a HMC; by integrating (4.5) with respect to $y_n, \ldots, y_1$, we arrive at $p(x) = p(x_1)p(x_2|x_1)\ldots p(x_n|x_{n-1})$, which indicates that the hidden chain $X = (X_1, \ldots, X_n)$ is Markovian. As mentioned in the previous section, we are interested in the possibilities of calculating *a posteriori* laws $p(x_i|y)$. The "forward" probabilities α_i and "backward" probabilities β_i of the previous section are first extended to PMCs by setting $\alpha_i(x_i) = p(y_1, \ldots, y_{i-1}, z_i)$ and $\beta_i(x_i) = p(y_{i+1}, \ldots, y_n|z_i)$. In a manner analogous to standard proofs, we can show that they are recursively calculable:

$$\alpha_1(x_1) = p(z_1), \text{ et } \alpha_{i+1}(x_{i+1}) = \sum_{x_i \in \Omega} \alpha_i(x_i)p(z_{i+1}|z_i) \text{ for } 2 \leq i \leq n-1, \tag{4.7}$$

$$\beta_n(x_n) = 1, \text{ et } \beta_i(x_i) = \sum_{x_{i+1} \in \Omega} \beta_{i+1}(x_{i+1})p(z_{i+1}|z_i) \text{ for } 1 \leq i \leq n-1. \tag{4.8}$$

As in the standard case of HMC-INs, we then have $p(x_i, y) = \alpha_i(x_i)\beta_i(x_i)$, giving $p(x_i|y)$. We can see that the PMC model, whose main advantage is better modeling of complex phenomena, makes it possible to calculate the distribution $p(x_i, y)$ of interest, in a process whose complexity is equivalent to that used in the HMC-IN case.

We saw earlier that the greater generality of PMCs compared to HMC-INs can be understood in the general case by comparing the transitions $p(z_{i+1}|z_i)$. In the case of stationary chains, we can show two necessary and sufficient conditions for a PMC to be a HMC. We will say that a PMC is stationary if $p(z_i, z_{i+1})$ does not depend on $1 \leq i \leq n-1$, and if $p(z_i, z_{i+1}) = p(z_{i+1}, z_i)$ for all $1 \leq i \leq n-1$. We then have the following result, a proof of which can be found in [PIE 07]:

Proposition 4.2.1
Let $Z = (X, Y)$ be a stationary PMC. Then the following three conditions are equivalent:

(i) X is a Markov chain;

(ii) for all $2 \leq i \leq n$, $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$;

(iii) for all $1 \leq i \leq n$, $p(y_i|x) = p(y_i|x_i)$

Condition (ii) helps us to understand, in the context of stationary chains, the advantage of PMCs compared to HMCs in local terms: as soon as we assume the Markovian nature of the hidden process X – which is always the case in the various standard extensions of HMC-INs – it is impossible to express any differences in $p(y_i|x_i, x_{i-1})$ as x_{i-1} varies. With the help of condition (iii) we can go even further down this avenue: in a HMC the distribution $p(y_i|x)$ cannot depend on any x_j different to x_i. Thus the modeling of noise is more complete in PMCs, where the $p(y_i|x)$ distributions depend on all the components $x_1, \ldots, x_n$ of x. It has also been observed through various simulations that this greater richness of the model can result in significant gains in the quality of non-supervized segmentation of data [DER 04].

Remark 4.2.1 *When we examine the various treatments which use standard HMCs, we realize that the Markovian nature of the distribution of X is never used, and is only of indirect use, allowing the Markovian nature of the* a posteriori *distribution of X, $p(x|y)$, something which is then indispensible. Furthermore, in the context of Proposition 4.2.1, assuming X to be Markovian is equivalent to imposing restrictions on the conditional distributions $p(y|x)$. In other words, in considering a PMC which is not*

an HMC, we lose the Markovian nature of X, which is not relevant to the treatment, but make significant gains in terms of the richness of the noise model. The main advantage of PMCs is thus to escape the assumption that X is Markovian, something which would otherwise restrict the flexibility of the model. We note that this assumption, although not indispensable, appears in all the models which use Markovian processes in restoration or segmentation problems, which are both hidden-variable problems. In fact, whether in the context of chains, fields, trees or general graphical models, the Markovian nature of the hidden process appears in the very name of the "hidden Markov models", and so gives every appearance of being incontrovertible.

As in the case of HMC-INs, it can be shown that for a PMC the distribution $p(x|y)$ is Markovian, with transitions given by:

$$p(x_{i+1}|x_i, y) = \frac{p(z_{i+1}|z_i)\beta_{i+1}(x_{i+1})}{\beta_i(x_i)} \tag{4.9}$$

and that the distributions $p(x_i, x_{i+1}|y)$, which will be used in Section 4.4, are given by:

$$p(x_i, x_{i+1}|y) = \alpha_i(x_i)p(z_{i+1}|z_i)\beta_{i+1}(x_{i+1}) \tag{4.10}$$

Also, note that the distribution for a stationary PMC is defined by $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. The transitions can then be written:

$$p(z_{i+1}|z_i) = \frac{p(x_i, x_{i+1})p(y_i, y_{i+1}|x_i, x_{i+1})}{\sum_{\omega\in\Omega} p(x_i, x_{i+1} = \omega)p(y_i|x_i, x_{i+1} = \omega)} \tag{4.11}$$

Finally, we observe that the distribution $p(z)$ of a PMC Z can always be written:

$$\begin{aligned} p(z) &= \frac{p(z_1, z_2)\dots p(z_{n-1}, z_n)}{p(z_2)\dots p(z_n)} \\ &= \underbrace{\frac{p(x_1, x_2)\dots p(x_{n-1}, x_n)}{p(x_2)\dots p(x_{n-1})}}_{a(x)} \\ &\times \underbrace{\frac{p(y_1, y_2|x_1, x_2)\dots p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_2|x_2)\dots p(y_{n-1}|x_{n-1})}}_{b(x,y)} . \end{aligned} \tag{4.12}$$

Thus a PMC is a HMC if and only if $p(x) = a(x)$ and $p(y|x) = b(x, y)$.

4.2.2. *Stationary Gaussian PMCs*

In an analogous way to what occurs in the original HMC-IN family, "Gaussian" models are among the simplest of the PMCs. Consider a stationary PMC, whose distribution is given by $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$ such that the conditional distributions $p(y_1, y_2|x_1, x_2)$, equal to the conditional distributions $p(y_i, y_{i+1}|x_i, x_{i+1})$ for all $i = 1, \ldots, n-1$, are Gaussian. If Z is not a PMC then we can see, by virtue of (4.11) and Proposition 4.2.1, that the transitions $p(z_{i+1}|z_i)$ do not have Gaussian distributions (the denominator is a mixture of Gaussians). As a consequence, $p(y|x)$ is not a Gaussian distribution. Furthermore, the marginal distributions $p(y_i|x)$, which are not Gaussian, depend on all $x_1, \ldots, x_n$. In order to prove this last property, we can make use of the fact that in a PMC the X and Y processes play symmetric roles. By reversing X and Y, we can introduce $\alpha_i^*(y_i) = p(x_1, \ldots, x_{i-1}, z_i)$ and $\beta_i^*(y_i) = p(x_{i+1}, \ldots, x_n|z_i)$, and show that these quantities satisfy the following recursions:

$$\alpha_1^*(y_1) = p(z_1), \text{ and}$$

$$\alpha_{i+1}^*(y_{i+1}) = \int_{\mathbb{R}} \alpha_i^*(y_i)p(z_{i+1}|z_i)dy_i \text{ for } 2 \leq i \leq n-1; \tag{4.13}$$

$$\beta_n^*(y_n) = 1, \text{ and}$$

$$\beta_i^*(y_i) = \int_{\mathbb{R}} \beta_{i+1}^*(y_{i+1})p(z_{i+1}|z_i)dy_{i+1} \text{ for } 1 \leq i \leq n-1. \tag{4.14}$$

However, contrary to the recursions in (4.7) and (4.8), which are calculable since they are defined as sums over small-sized ensembles, the recursions in (4.13), (4.14) are defined by integrals, and are not therefore analytically calculable, except for α_2^* and β_{n-1}^*.

Thus in the case of Gaussian and stationary PMCs, which are not HMCs, the noise distributions $p(y|x)$ are not Gaussian, and are impossible to calculate. However, these "difficulties" in fact are a sign of the richness, and do not cause problems with the determination of the $p(x_i|y)$ distributions.

4.3. Copulas in PMCs

The advantage of copulas is their ability to model, in a particularly elegant manner, the dependencies between random variables. More precisely, the dependence is given by a "copula" which is independent of the form of the marginal distributions. The theory of copulas is relatively well-established [NEL 98]; however, their introduction into a pairwise Markov chain, which we will briefly discuss in this section, is relatively recent [BRU 05a, BRU 05b].

4.3.1. *Copulas*

We will use the term "copula" to refer to any cumulative distribution function C for the distribution of a pair of random variables which takes its values from $[0,1]^2$, whose marginal distributions are uniform distributions over $[0,1]$. Let $h(y_1, y_2)$ be a probability density over $\mathbb{R}^2$, which we will assume to be continuous, H its cumulative distribution function, $h_1(y_1)$, $h_2(y_2)$ the marginal distribution densities, and H_1, H_2 the associated cumulative distribution functions. According to Sklar's theorem, there then exists a unique copula such that (see [NEL 98]):

$$H(y_1, y_2) = C\left(H_1(y_1), H_2(y_2)\right). \tag{4.15}$$

Differentiating (4.15) with respect to y_1, y_2 and setting $c(u,v) = \dfrac{\partial\partial C(u,v)}{\partial u \partial v}$, we have:

$$h(y_1, y_2) = h_1(y_1) h_2(y_2) c\left(H_1(y_2), H_2(y_2)\right) \tag{4.16}$$

Conversely, considering distribution functions H_1, H_2 over $\mathbb{R}$ and a copula C, (4.15) defines a distribution function H over $\mathbb{R}^2$.

Finally, we will say that using a distribution function H is equivalent to using the triplet (H_1, H_2, C), whereH_1 and H_2 are the marginal distributions of H, and C is the copula associated with H through (4.15). It then follows that it is possible to construct distributions over $\mathbb{R}^2$ by considering various marginal distributions – and various copulas – in an independent manner. In particular, it is possible, starting from (Y_1, Y_2) with a given H, to construct real, correlated random variables with any desired marginal distribution. Consider the marginal distributions H_1, H_2 associated with H, with H_1', H_2' being the desired new marginal distributions. Setting $Y_1' = (H_1')^{-1} \circ H_1(Y_1)$ and $Y_2' = (H_2')^{-1} \circ H_2(Y_2)$, the variables (Y_1', Y_2') are correlated, and have H_1', H_2' for the cumulative distribution functions of their distributions. Finally, we note that the copula C associated with H through (4.15) can also be seen as the distribution function for the copula $(H_1(Y_1), H_2(Y_2))$, whose marginal distributions are thus indeed uniform distributions over $[0,1]$.

Example 4.3.1 *Let $h(y_1, y_2)$ be the Gaussian density over $\mathbb{R}^2$ defined with both means being zero, both variances equal to 1, and with a correlation coefficient ρ. According to (4.15), the associated Gaussian copula is given by:*

$$C(u,v) = H\left(H_1^{-1}(u), H_2^{-1}(v)\right), \tag{4.17}$$

or by its derivative:

$$c(u,v) = \frac{h\left(H_1^{-1}(u), H_2^{-1}(v)\right)}{h_1\left(H_1^{-1}(u)\right) h_2\left(H_2^{-1}(v)\right)}. \tag{4.18}$$

A Gaussian copula over $\mathbb{R}^2$ is thus defined by a single real parameter ρ. As discussed above, this copula can be used to defined a distribution for a copula (Y_1, Y_2) of dependent variables, with the required marginal distributions. Let us consider two random real variables Y_1, Y_2, with marginal distribution functions F_1, F_2, with f_1, f_2 being their corresponding densities. The following density for the distribution of (Y_1, Y_2)

$$f(y_1, y_2) = f_1(y_1) f_2(y_2) \frac{h\left(H_1^{-1}\left(F_1(y_1)\right), H_2^{-1}\left(F_2(y_2)\right)\right)}{h_1\left(H_1^{-1}\left(F_1(y_1)\right)\right) h_2\left(H_2^{-1}\left(F_2(y_2)\right)\right)} \tag{4.19}$$

then defines the distribution for a copula of dependent random variables (Y_1, Y_2) such that their marginal distributions have densities f_1 and f_2. It is interesting to note that when the densities f_1 and f_2 are Gaussian, then $f(y_1, y_2)$, given by (4.19), is Gaussian. Of course, we can also consider f_1, f_2 to be Gaussian, and consider a copula other than a Gaussian copula [BRU 05a].

4.3.2. *Copulas in PMCs*

Let $Z = (X, Y)$ be a stationary PMC whose distribution is defined by $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. For all classes $\Omega = \{\omega_1, \ldots, \omega_k\}$, we then have, in the general case, $k(k-1)$ densities $p_{ij}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$. Each of these densities defines a copula and two marginal distributions over $\mathbb{R}^2$. We note that in the specific case of PMCs, the number of marginal distributions is k.

We observe that variation in the form of the marginal distribution $p(y_1|x_1 = \omega_i)$ with the class ω_i can occur in real images, as has been shown in various studies, including those in [DEL 02, GIO 97, MAI 01].

4.4. Parameter estimation

The problem of parameter estimation in triplet Markov chains is identical to that of parameter estimation in PMCs; as a result we can treat them in this simpler context. When this problem is treated in the case of traditional HMCs, the method most often used is that known as "expectation-maximization" or EM [MCL 97]. This method is based on maximization, through an iterative process, of the likelihood of the observed chain, and can give remarkable results as long as the initialization is not too far removed from the actual parameters. It also benefits from partial theoretical results concerning its asymptotic behavior; in particular, it can be shown that it produces a sequence of increasingly large likelihood values as the iterations progress. However, other than in simple cases, the explicit calculation of the various values involved in the iterations is often impossible, and it is often necessary to resort to approximations, which by nature destroy the good general characteristics of the maximum likelihood method.

In this section we will present an alternative method known as "iterative conditional estimation" or ICE [PIE 92, PIE 08]. This is based on a different principle, which makes use of the concept of conditional expectation, and does not make use of the marginal likelihood of the observed process. We note however that the maximum likelihood estimator, based on complete data, may be involved in ICE, which thus guarantees an "optimality" comparable to that offered by the EM method. Not only does this principle of ICE make it simpler, but it also makes it practical in a more general context than the EM method.

It has been applied in various complex modelizations such as triplet Markov fields [BEN 05a, BEN 05b, BEN 07a, BEN 07b], evidential hidden Markov fields [BEN 01], or non-stationary hidden Markovian or semi-Markovian chains [LAP 06, LAP 07]. We note that the first theoretical results for the asymptotic behavior of ICE, valid in the case of independent observations, have only recently been published [PIE 08], and Markovian models offer a very interesting range of possible theorietical investigations. Finally, in certain parameterizations of exponential models, ICE and EM may be equivalent [DEL 97].

We will present two variants of ICE which are well suited to PMCs, of increasing generality. The first variant is a standard one involving parametric models, and the second involves generalized mixtures and, in the context of PMCs with copulas, contains original aspects.

4.4.1. *ICE in a parametric model*

In general terms, we will consider two stochastic processes (X, Y) whose distribution depends on a parameter $\theta = (\theta_1, \ldots, \theta_s) \in \mathbb{R}^s$. Suppose that the problem is the estimation of θ from the process Y, the only observable. It is then possible to use ICE subject to the following two conditions:

(i) there exists an estimator $\hat{\theta}(X, Y)$ for θ based on complete data (X, Y);

(ii) it is possible, for all θ, to simulate X using $p(x|y, \theta)$.

It should be pointed out that these conditions are not very restrictive. In fact (i) is not really a condition at all, since if an estimator cannot be obtained that uses complete data then the problem of estimation is also insoluble based on incomplete data. Condition (ii) is almost always satisfied in problems where we are trying to obtain X from Y using Bayesian methods; such methods generally require $p(x|y, \theta)$ to be Markovian.

ICE is carried out in the following manner:

(i) θ^0 is taken as an initial value of θ;

(ii) $\theta_r^{q+1} = E[\hat{\theta}_r(X,Y)|Y = y, \theta^q]$ is set for the components θ_r of θ for which this expectation is calculable;

(iii) for the components θ_r of θ for which the above expectation is not calculable, m values $x^{q,1}, \ldots, x^{q,m}$ of X are simulated using $p(x|y, \theta^q)$, and we set:

$$\theta_r^{q+1} = \frac{\hat{\theta}_r(x^{q,1}, y) + \cdots + \hat{\theta}_r(x^{q,m}, y)}{m}.$$

A stationary PMC $(X,Y) = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ will be said to be "parametric" when the distribution $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$ depends on a set of parameters θ. We will divide the components of the vector θ into two sub-vectors. The first defines the distribution $p(x_1, x_2)$ over Ω^2, and so contains k^2 components. The second defines all the distributions $p(y_1, y_2|x_1, x_2)$, and so contains $(k-1)k$ groups of components, with each group corresponding to one $(x_1, x_2) = (\omega_i, \omega_j)$. If all these distributions are Gaussian, each group contains five real components: two means, two variances and the covaraiance. In the Gaussian case the vector θ therefore contains $k^2 + 5k(k-1)$ components.

In order to simplify the notation, we will set $p(i,j) = p(x_1 = \omega_i, x_2 = \omega_j)$, and define θ_{ij} as the group of parameters defining $p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$. In order to apply ICE, we must consider an estimator $\hat{\theta}(X,Y)$ which relies on complete data (X,Y). The components $p(i,j)$ can be estimated in a standard manner from X using frequencies:

$$\hat{p}(i,j) = \frac{1_{[x_1=\omega_i, x_2=\omega_j]} + \cdots + 1_{[x_{2n-1}=\omega_i, x_{2n}=\omega_j]}}{n}, \tag{4.20}$$

where 1_A is the indicatrix function for the set A.

The estimation of the groups θ_{ij} based on (X,Y) is possible as long as each group θ_{ij} can be estimated from a sample produced by $p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$, which we will assume is possible. Then, knowing $X = x = (x_1, \ldots, x_{2n})$, we define a sub-sample y^{ij} of $y = (y_1, \ldots, y_{2n})$ by $(y_{2l}, y_{2l+1}) \in y^{ij}$ if and only if $(x_{2l} = \omega_i, x_{2l+1} = \omega_j)$. Thus we obtain samples of the copulas (Y_{2l}, Y_{2l+1}) of distributions defined by θ_{ij}, which leads us to an estimator $\hat{\theta}_{ij}$. We will write $\hat{\theta}_{ij}(x,y)$, knowing that in this notation $\hat{\theta}_{ij}(x,y)$ only depends on y^{ij}, with this being defined using $x = (x_1, \ldots, x_{2n})$.

For example, in the Gaussian case both means are estimated using empirical means, and the covariance matrix by the empirical covariance matrix.

Since the estimator $\hat{\theta}(X,Y)$ is defined by (4.20) and the estimators $\hat{\theta}_{ij}(X,Y)$, ICE can be applied in the following manner. The calculation of the likelihood

$$E[\hat{\theta}_r(X,Y)|Y = y, \theta^q]$$

giving θ_r^{q+1} is possible for the components of the first sub-vector of θ given by $p(i,j)$. Knowing that the likelihood of an indicatrix function is the probability of the corresponding ensemble, by taking the conditional expectation of (4.20) we obtain:

$$p^{q+1}(i,j) =$$

$$\frac{p(x_1 = \omega_i, x_2 = \omega_j | y, \theta^q) + \cdots + p(x_{2n-1} = \omega_i, x_{2n} = \omega_j | y, \theta^q)}{n} \tag{4.21}$$

where the probabilities $p(x_{2n-1} = \omega_i, x_{2n} = \omega_j | y, \theta^q)$ can be calculated from (4.10), in Section 4.2. As for the groups of parameters θ_{ij}, the conditional expectation is not calculable, and step (iii) of the ECI is used, where simulations are used. We note that since the distribution for X conditional on $Y = y$ is a Markov chain with transitions calculable using (4.9) (Section 4.2), the simulations do not pose any particular problem. In practice, often a single sample x^q is used, and we set $\theta_{ij}^{q+1} = \hat{\theta}_{ij}(x^q, y)$.

4.4.2. *ICE in a generalized mixture model with copulas*

Consider a stationary PMC with a distribution defined by the probability density $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. Suppose that each of the $(k-1)k$ marginal densities $p_{ij}(y_1) = p(y_1|x_1 = \omega_i, x_2 = \omega_j)$ has a form belonging to a set of admissible forms, which will be denoted as Φ^{ij}. For example we can imagine that for two classes $\Omega = \{\omega_1, \omega_2\}$, Φ^{11} contains two elements: "Gaussian distributions" and "gamma distributions", $\Phi^{12} = \Phi^{21}$ contains three elements: "Gaussian distributions, "gamma distributions" and "beta distributions", and Φ^{22} contains two elements: "Gaussian distributions" and "K distributions". The problem is then twofold: to identify the form of each of these densities, and to estimate the parameters that define the density within the identified family. This type of problem, treated in [GIO 97] for the case of HMC-IN models, is known as "generalized mixture estimation". We will describe an approach of this type, but extended to PMCs. This will firstly generalize the approach presented in [GIO 97] and secondly generalize the approach laid out in the previous subsection, by the use of copulas.

We said in Section 4.3 that each of the $(k-1)k$ distributions $p_{ij}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$ is defined by $(k-1)k$ marginal densities and $(k-1)k$ copulas. Suppose that for each (i,j), the form of the copula associated with $p_{ij}(y_1, y_2)$ belongs to the set $\mathcal{X}^{ij}$ of admissible forms. For each (i,j) we must therefore:

(i) select from Φ^{ij} the marginal form $p_{ij}(y_1)$ and estimate its corresponding parameters;

(ii) select from $\mathcal{X}^{ij}$ the form of the copula given by $p_{ij}(y_1, y_2)$ and estimate its corresponding parameters.

Consider the case of estimation from complete data, and suppose $X = x = (x_1, \ldots, x_{2n})$ are known. As before, we define, for each (i, j), a subsample y^{ij} from $y = (y_1, \ldots, y_{2n})$ by $(y_{2l}, y_{2l+1}) \in y^{ij}$ if and only if $(x_{2l} = \omega_i, x_{2l+1} = \omega_j)$. Thus for each (i, j) we have a sample of copulas (Y_{2l}, Y_{2l+1}) whose marginal laws belong to one of the families $F_1^{ij}, \ldots, F_{r(i,j)}^{ij}$ within $\Phi^{ij} = \{F_1^{ij}, \ldots, F_{r(i,j)}^{ij}\}$. Supposing that these families are parameterized respectively by $\theta_1^{ij}, \ldots, \theta_{r(i,j)}^{ij}$ and that adequate estimators exist, we will denote their estimated values by $\hat{\theta}_1^{ij}, \ldots, \hat{\theta}_{r(i,j)}^{ij}$. These values give $r(i, j)$ "candidate" densities for the correct distribution, and one of these candidates must be chosen in order to optimize some criterion which may be, as is the case in [GIO 97], the histogram-based Kolmogorov distance. We will define H_1^{ij}, H_2^{ij} as the distribution functions of the marginal distributions that have been estimated in this way; then the distribution of the copula $\left(H_1^{ij}(Y_1), H_2^{ij}(Y_2)\right)$ is the copula we are looking for. Supposing that it belongs to one of the families $C_1^{ij}, \ldots, C_{m(i,j)}^{ij}$ within $\mathcal{X}^{ij} = \{C_1^{ij}, \ldots, C_{m(i,j)}^{ij}\}$, we can then estimate $m(i, j)$ "candidates" and chose the one which minimizes some criterion, in a manner analogous to that used to find the marginal distributions.

This whole procedure is then incorporated into ICE where, at each iteration, x is replaced by the simulated x^q.

Remark 4.4.1 *We observed in Remark 4.2.1 that the standard habit of systematically assuming the hidden process to be Markovian probably contributes to the persistent use of "hidden Markov models" (HMMs), with simple noise models. We also observe that the difficulties in studying parameter estimation in theoretical terms is probably a second reason that these models persist. For example, PMCs are briefly mentioned in [CAP 05], but the authors do not follow this line for very long, because they subsequently consider theoretical results for maximum likelihood estimation which are intimately linked to the Markovian nature of the hidden process.*

4.5. Triplet Markov chains

The problem remains exactly the same as that considered in the previous sections: recovering the hidden realization of the X chain from the observed realization of the Y chain. For the case of a triplet Markov chain, we need to introduce a third random chain $U = (U_1, \ldots, U_n)$, which here we will assume takes finite values (each U_n takes its values from a finite set $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$), and assume that the triplet $T = (X, U, Y)$ is a Markov chain. We note that U may or may not have physical significance; we also note that the TMC model is strictly more general than the PMC model; in a TMC $T = (X, U, Y)$, the chain $Z = (X, Y)$ is not necessarily Markovian. However, if we set $V = (X, U)$, the chain $T = (V, Y)$ is a PMC and we can use all

the results from the previous sections. In particular, the method of estimation of the ICE parameters is applicable, including in the context of copulas.

In Bayesian MPM restoration, the advantage of TMCs is the fact that when the cardinal of Λ is not too large, the *a posteriori* marginals $p(x_i|y)$ are calculable. In fact, $p(v_i|y) = p(x_i, u_i|y)$ are such because $T = (V, Y)$ is a PMC, and so the $p(x_i|y)$ distributions are simply given by:

$$p(x_i|y) = \sum_{u_i \in \Lambda} p(x_i, u_i|y). \tag{4.22}$$

Thus we obtain a very rich model, which can give lead to a large number of specific models. In the following sections we will discuss various situations where different TMCs are applied, sometimes in an unsupervised manner, to the restoration of hidden data. In this section we will restrict ourselves to stating a result which generalizes the various specific cases, which is very useful for understanding the nuances of specific TMC models. Then, we will present two examples of TMCs, illustrating their richness and how they differ from classical models.

The following result, which generalizes Proposition 4.2.1, allows the construction of the different specific TMCs.

Let $W = (G, H) = (G_1, H_1, \ldots, G_n, H_n)$ be a random chain, with each $W_i = (G_i, H_i)$ taking values from the ensemble given by the product $\Gamma \times \mathcal{H}$. Let γ be a σ-additive measure over Γ, and let η be a σ-additive measure over $\mathcal{H}$. We will assume W to be Markovian, and represent with the same letter p the various densities with respect to the different measures γ and η. In practice, each of the measures γ and η is either the counting measure or the Lebesgue measure; however, mixed measures, involving Dirac masses and the Lebesgue measure, can also be considered, as in [SAL 07].

We have the following result, whose proof follows the general outline of the proof of Proposition 4.2.1.

Proposition 4.5.1 *Let $W = (G, H)$ be a Markov chain satisfying the following conditions:*

(a) $p(w_i, w_{i+1})$ does not depend on $1 \leq i \leq n-1$;

(b) $p(w_i = a, w_{i+1} = b) = p(w_i = b, w_{i+1} = a)$ for all $1 \leq i \leq n-1$, a, and b.

The following three conditions are then equivalent:

(i) G is a Markov chain;

(ii) for all $2 \leq i \leq n$, $p(h_i|g_i, g_{i-1}) = p(h_i|g_i)$;

(iii) for all $1 \leq i \leq n$, $p(h_i|g) = p(h_i|g_i)$,

We will now present two examples. The first illustrates the fact that certain TMCs can have complexities comparable to that of the classical HMC-IN model, and the second generalizes classical mixture models.

Example 4.5.1 *Let $T = (X, U, Y)$ be a TMC whose distribution is given by $p(x_1, u_1, y_1)$, with transitions $p(t_{i+1}|t_i)$ of the form:*

$$p(t_{i+1}|t_i) = p(u_{i+1}|u_i)p(x_{i+1}|u_{i+1})p(y_{i+1}|u_{i+1}). \tag{4.23}$$

Such a TMC is relatively simple, while being very different from classical HMC-INs. In particular, the following properties can be shown: U is a Markov chain; none of the chains X, Y, (X, Y) are Markovian; X, Y are independent conditional on U.

This example shows that it is possible to very simply construct workable models that are as simple as HMC-INs, but very different from them. In particular, none of the three following properties which define HMC-INs: (i) X is Markovian; (ii) the random variables $Y_1, \ldots, Y_n$ are independent conditional on X ; (iii) $p(y_i|x) = p(y_i|x_i)$ for all $i = 1, \ldots, n$, are satisfied by the model in Example 4.5.1.

Example 4.5.2 *Consider the classical HMC-IN case given by (4.1), and suppose that the densities $p(y_i|x_i)$ are mixtures $p(y_i|x_i = \omega_j) = \sum_{l=1}^{m} \alpha_{lj} f_{lj}(y_i)$. Such models are known as "mixture models", and can in particular be used when the form of the distributions $p(y_i|x_i)$ is not known, in which case the latter are approximated by mixtures, for example of Gaussian distributions [DIZ 07]. Such a model is a TMC whose distribution is given by $p(x_1, u_1, y_1)$, and the transitions $p(t_{i+1}|t_i)$ have the form:*

$$p(t_{i+1}|t_i) = p(x_{i+1}|x_i)p(u_{i+1}|x_{i+1})p(y_{i+1}|u_{i+1}, x_{i+1}) \tag{4.24}$$

with $p(u_i = \lambda_l|x_i = \omega_j) = \alpha_{lj}$ and $p(y_i|u_i = \lambda_l, x_i = \omega_j) = f_{lj}(y_i)$. We note that in this example we have:

$$p(y_i|x) = p(y_i|x_i). \tag{4.25}$$

Consider now a stationary TMC, more general to that given by (4.24), whose distribution has the form:

$$p(t_{i+1}|t_i) = p(x_{i+1}, u_{i+1}|x_i, u_i)p(y_{i+1}|u_{i+1}, x_{i+1}) \tag{4.26}$$

such that $p(x_{i+1}|u_{i+1}, u_i) \neq p(x_{i+1}|u_{i+1})$ and $p(u_{i+1}|x_{i+1}, x_i) \neq p(u_{i+1}|x_{i+1})$. The chain $V = (X, U)$ is then Markovian and $p(y_i|u, x) = p(y_i|u_i, x_i)$: thus (V, Y)

is a classical HMC-IN. However, by applying Proposition 4.5.1 to $G = X$ *and* $H = U$*, we can see that condition (ii) is not satisfied, and so* X *is not a Markov chain. In a similar way it can be shown that* (X, Y) *is not Markovian either, and furthermore that the random variables* $Y_1, \ldots, Y_n$ *are not necessarily independent conditional on* X*. Finally, we will show that* $p(y_i|x)$ *depends on all* $x_1, \ldots, x_n$*. We have* $p(y_i|x) = \sum_{u_i \in \Lambda} p(y_i, u_i|x) = \sum_{u_i \in \Lambda} p(y_i|x, u_i)p(u_i|x) = \sum_{u_i \in \Lambda} p(y_i|x_i, u_i)p(u_i|x)$*, which shows that* $p(y_i|x)$ *is a mixture whose coefficients* $p(u_i|x)$ *depend on all* $x_1, \ldots, x_n$*. It can also be seen that the distributions* $p(y_i|x_i)$ *are extremely rich mixtures: for* k *possible values for each* x_n*, and* m *possible values for each* u_n*, the number of components in the mixture* $p(y_i|x_i)$ *is* mk^{n-1}*. So, setting* $x^* = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$*, we have* $p(y_i|x_i) =$
$\sum_{(x^*, u_i)} p(x^*, u_i, y_i|x_i) = \sum_{(x^*, u_i)} p(y_i|x_i, x^*, u_i)p(x^*, u_i|x_i) =$
$\sum_{(x^*, u_i)} p(y_i|x_i, u_i)p(u_i|x)p(x^*|x_i)$ *and* mk^{n-1} *is the number of possible* (x^*, u_i) *copulas.*

Thus, this model generalizes in a consistent manner, which can be useful in applications such as classical mixture models, in that $p(y_i|x_i)$ is a mixture of mk^{n-1} components instead of m.

4.6. Triplet Markov chains and non-stationarity

TMCs $T = (X, U, Y)$ can be used to treat the non-stationarity of the chain $Z = (X, Y)$. The third chain U then has a physical significance, and each state in $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$ models a particular stationarity. More specifically, we will say that a random chain $Z = (X, Y)$ is an "m-stationary TMC" if a random chain U exists with values in a set of m elements such that $T = (X, U, Y)$ is a stationary TMC. U then models the fact that there are m different stationarities, at random instances, in the chain $Z = (X, Y)$. It is then possible to consider different types of restoration. Classically, we could look for $X = x$, as described in previous sections. We could also consider the copula $(X, U) = (x, u)$, or even just the chain $U = u$. The flexibility of Bayesian approaches then makes it possible, through various loss functions, to propose various segmentation methods, which may be non-supervised, based on the Markovian nature of the conditional distribution $p(x, u|y)$.

The advantage of using TMCs in the context of data which contain a finite number of stationarities has been shown in the context of image segmentation through various experiments presented in [LAN 04, LAN 05, LAN 06], an example of which we will discuss below.

We note that the introduction of a third chain modeling the different stationarities is not original, because various "jump process" models, where the jump process fulfills this role, have previously been proposed and studied. However, in these various such models, the jump process is always assumed to be Markovian, which is, by virtue of Proposition 4.5.1, a special case – and a relatively restrictive one – of the TMCs discussed in this section.

Example 4.6.1 *Consider a 256x256 binary image of a zebra, as shown in Figure 4.1. The image is given by a square grid of pixels, where each pixel is either black or white. The grid is transformed into a sequence using a "Hilbert-Peano" scan, as suggested in [GIO 97]. We then obtain a chain $X = x$, and the observed image becomes a chain $Y = y$. The synthetic noise is Gaussian, and spatially invariant, with a variance of one for both classes, and with means of 0 and 2 respectively. Aside from the issue of stationarity, the noise thus corresponds to the classical HMC-IN model. The observed image is then segmented in two different manners: using the classical HMC-IN model, and using the triplet model with a distribution given by:*

$$p(x_1, u_1, y_1) = p(x_1, u_1)p(y_1|x_1),$$
$$p(t_{i+1}|t_i) = p(x_{i+1}, u_{i+1}|x_i, u_i)p(y_{i+1}|x_{i+1}). \tag{4.27}$$

We note that the TMC used is a relatively rudimentary one: in particular, U and Y are independent conditional on X. The classical method gives $\hat{X}^1 = \hat{x}^1$, with an error level of 7.2%, whereas the method using the TMC gives $\hat{X}^2 = \hat{x}^2$, with an error level of 3.5%. In this example we are assuming that there are three different stationarities in the image ($m = 3$). The determination of $U = u$ then gives $\hat{U} = \hat{u}$, which is also shown in Figure 4.1, and it can be seen that the three different regions of stationarity, which are the background ("black" stationarity), the relatively broad bands found on the body of the animal ("grey" stationarity), and the relatively narrow bands found on the neck and legs of the animal ("white" stationarity) have been successfully recovered.

Similar studies have been carried out in the context of images segmentation using methods based on triplet Markov fields [BEN 05a, BEN 05b, BEN 07b]. In particular, such models and associated methods of treatment appear very promising in the treatment of the important problem of texture classification [BEN 07a, BLA 08]. Finally, it should be mentioned that triplet Markov chains and fields allow the use of the Dempster-Shafer theory of evidence in a Markovian context, which in particular also allows the study of non-stationary data [LAN 05, PIE 06].

4.7. Hidden semi-Markov chains, and triplet Markov chains

In many areas of application, HMC-INs have proven insufficient, and more general models have had to be used, models known as "hidden semi-Markov chains"

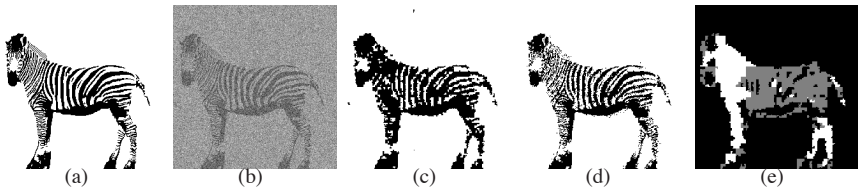


Figure 4.1.
(a) Image $X = x$ of a zebra, (b) a version $Y = y$ with independent Gaussian noise, (c) non-supervised segmentation using HMC-INs $\tau = 7.2\%$, (d) non-supervised segmentation using TMCs: $\tau = 3.5\%$, (e) segmentation $\hat{U} = \hat{u}$ into three different stationarities using TMCs. τ represents the error level.

(HSMCs). Applications exist in the segmentation of medical images [FAI 05], speech processing [MOO 04], and tracking [YU 03] (we should also mention hybrid models which combine the "CMC" and "HSMC" aspects [GUE 05]). Given that a semi-Markov distribution can be viewed as the marginal distribution of a Markovian distribution, it can be seen that HSMCs are a special case of TMCs. This observation allows many extensions to classical HSMCs to be proposed within a unified framework. We will use the rest of this section to briefly describe these.

4.7.1. *Hidden semi-Markov chains as a special case of TMCs*

Consider a process $X = (X_1, ..., X_n, ...)$, where each X_i takes values from $\Omega = \{\omega_1, ..., \omega_k\}$. X is a "semi-Markov" chain if its distribution is given by the distribution of X_1, represented as $p(x_1)$, with the set of transition matrices $(p(x_i|x_{i-1}))_{i\geq 2}$ satisfying $p(x_i|x_{i-1}) = 0$ for $x_i = x_{i-1}$, and k sets of distributions over $\mathbb{N}^*$. For each $i = 1, ..., k$, the set of corresponding distributions will be denoted by $p^*(.|x_n = \omega_i)$. A realization of the semi-Markov chain $X = (X_1, ..., X_n, ...)$ is obtained in the following manner: (i) $X_1 = x_1$ is simulated based on $p(x_1)$; (ii) a positive integer $N_1 = n_1$ is simulated based on $p^*(.|x_1)$; (iii) for $1 \leq i \leq n_1$, we set $x_i = x_1$; (iv) $X_{n_1+1} = x_{n_1+1}$ is simulated based on $p(x_{n_1+1}|x_{n_1})$; a positive integer $N_2 = n_2$ is simulated based on $p^*(.|x_{n_1})$; etc.

A Markov chain is then a special case of a semi-Markovian chain in which $p^*(.|x_n = \omega_i)$, which is the sojourn time distribution in ω_i given that the state was entered at time n, is an exponential distribution. The introduction of HSMCs was inspired by the observation that in a number of real-world phenomena the sojourn time in a given state is not exponential.

When considering a real observed process $Y = (Y_1, ..., Y_n, ...)$, the distribution for a HSMC is obtained by defining $p(y|x)$ in the same way as for HMCs: for each

n and $X = (X_1, ..., X_n)$, $Y = (Y_1, ..., Y_n)$, we set $p(y|x) = p(y_1|x_1)...p(y_n|x_n)$. It can then be shown that HSMCs are a special case of TMCs and so, at the expense of a greater complexity to the calculations, they can be treated in the same way. The chain U is such that, for each n, the remaining sojourn time of X_n in x_n is modeled by U_n. If this time is greater that 1 then we have $U_{n+1} = U_n - 1$. We therefore have the following result:

Proposition 4.7.1 *A hidden semi-Markov chain is a triplet Markov chain $T = (X, U, Y)$, with U_n taking values from $\mathbb{N}^*$, defined by $p(x_1, u_1, y_1) = p(x_1)p(u_1|x_1)p(y_1|x_1)$ and with transitions:*

$$p(x_{n+1}, u_{n+1}, y_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n, y_n)\times$$
$$p(u_{n+1}|x_n, u_n, y_n, x_{n+1})p(y_{n+1}|x_n, u_n, y_n, x_{n+1}, u_{n+1}) \tag{4.28}$$

given by (δ representing the Dirac mass):

$$p(x_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n) = \begin{cases} \delta_{x_n}(x_{n+1}) \text{ if } u_n > 1 \\ p(x_{n+1}|x_n) \text{ if } u_n = 1 \end{cases} \tag{4.29}$$

$$p(u_{n+1}|x_n, x_{n+1}, u_n, y_n) =$$
$$= p(u_{n+1}|x_{n+1}, u_n) = \begin{cases} \delta_{u_n-1}(u_{n+1}) \text{ if } u_n > 1 \\ p(u_{n+1}|x_{n+1}) \text{ if } u_n = 1 \end{cases} \tag{4.30}$$

$$p(y_{n+1}|x_n, u_n, y_n, u_{n+1}, x_{n+1}) = p(y_{n+1}|x_{n+1}) \tag{4.31}$$

Proof 4.7.1 *The proof is immediate.*

4.7.2. *Extensions to hidden semi-Markov chains*

The result given in Proposition 4.7.1 leads us to suggest two sets of generalizations to HSMCs, using a number of recent results that apply to TMCs.

In the first set, we will remain within the general model 4.28, but consider various extensions to identities 4.29-4.31.

The first generalization is obtained by replacing $p(x_{n+1}|x_n)$ in Equation 4.29 by $p(x_{n+1}|x_n, y_n)$ (for $u_n = 1$). To see why this is a useful extension, imagine that the indices of the various processes are the pixels of a line in a numerical image representing a scene consisting of three classes: "water", "desert" and "forest". Each X_n takes its values from $\Omega = \{\omega_1, \omega_2, \omega_3\}$, each Y_n takes its value from the subset of $\mathbb{R}$ of the possible numerical values, and the distribution $p(y_n|x_n)$ models, among other things, the "natural variation" (different types of sand, level of greenery in the forest,

the color of the water, etc.). For $x_n = \omega_3$ and $u_n = 1$ we know that the forest ends at n and the probability $p(x_{n+1}|x_n = \omega_3)$ determines whether the "water" or "desert" class is found at $n+1$. We can then imagine that the probability of finding "water" at $n+1$ depends on the nature of the forest at n: the greener the forest, the higher chance of finding water at $n+1$. This is an example of how replacing $p(x_{n+1}|x_n)$ with $p(x_{n+1}|x_n, y_n)$ can have a physical justification. Furthermore, it is now clear that this type of extension can improve the treatments significantly. By applying it in the simplest context, where replacing $p(x_{n+1}|x_n)$ with $p(x_{n+1}|x_n, y_n)$ turns hidden Markov chains into pairwise Markov chains, it is possible to significantly improve the quality of the non-supervised segmentation of certain images [DER 04].

Similar reasoning leads to a justification for a generalization of Equation 4.31 in which $p(y_{n+1}|x_{n+1})$ is replaced with $p(y_{n+1}|x_{n+1}, x_n)$, which is a second extension to the model described by Equations 4.29-4.31. Given that u_{n+1} represents the residual sojourn time of $X_{n+1}, X_{n+2}, ...$ in x_n, we can imagine, returning to the earlier example, that the aspect y_{n+1} of the class x_{n+1} depends on this sojourn time (an isolated tree presents mean characteristics which are different to a tree contained within a forest), and so it is the case that $p(y_{n+1}|x_n, x_{n+1}, u_{n+1}) \neq p(y_{n+1}|x_n, x_{n+1})$. We can also imagine that the mean aspect (and so the distribution of Y_{n+1}) of the class x_{n+1} is different depending on whether there is a change of class between n and $n+1$ (which is equivalent to $u_n = 1$), or not (which is equivalent to $u_n > 1$). This means that $p(y_{n+1}|x_n, u_n, x_{n+1}, u_{n+1}) \neq p(y_{n+1}|x_n, x_{n+1}, u_{n+1})$. Finally, we could also imagine that $p(u_{n+1}|x_{n+1})$ in 4.30 also depends on y_n.

When we consider these two extensions together, Equations 4.29-4.31 become:

$$p(x_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n) = \delta_{x_n}(x_{n+1}) \text{ if } u_n > 1,$$
$$\text{and } p(x_{n+1}|x_n, y_n) \text{ if } u_n = 1. \tag{4.32}$$
$$p(u_{n+1}|x_n, x_{n+1}, u_n, y_n) = p(u_{n+1}|u_n, x_{n+1}) = \delta_{u_n - 1}(u_{n+1}) \text{ if } u_n > 1,$$
$$\text{and } p(u_{n+1}|x_{n+1}, y_n) \text{ if } u_n = 1. \tag{4.33}$$
$$p(y_{n+1}|x_n, u_n, y_n, u_{n+1}, x_{n+1}) = p(y_{n+1}|x_n, u_n, u_{n+1}, x_{n+1}). \tag{4.34}$$

Extensions 4.32-4.34 to the transitions in 4.29-4.31 can thus be solidly justified; in other words, 4.29-4.31 are simplifications of real-world situations which can clearly be seen to have an impact on the quality of the treatment.

The second set of extensions is obtained by making use of the fact that the TMC $T = (X, U, Y)$ from the proposition in the previous section is also a PMC $T = (V, Y)$, with $V = (X, U)$. The classic transition from a pairwise chain to a triplet chain can then be carried out, by considering a latent process $W = (W_1, ..., W_n, ...)$, with each W_i taking its values from $\Lambda = \{\lambda_1, ..., \lambda_m\}$, and a new triplet chain $T' = (V, W, Y)$. All the cases considered in the first set we discussed can then be extended to T' by

replacing (x, u) with (x, u, w) in every transition in each of the models. The models of the first set allow the estimation of (x, u), and so those of the second set allow the estimation of (x, u, w). These various estimations give the x that is to be determined.

4.8. Auxiliary multivariate chains

The various properties of the hidden random chains modeled by TMCs, as discussed in the previous sections, can be considered all together, which leads to a large number of specific models. This involves considering an auxiliary chain U in the multivariate form $U = (U^1, ..., U^d)$, where each U^i models a particular property, which could be one of the ones discussed in the preceding sections. This rapidly leads to fairly complex models whose implementation, particularly in the case of non-supervised methods, quickly becomes rather demanding. In terms of the practical usefulness of such models, at least two questions then come up:

(i) since classical HMC-IN models have had their robustness demonstrated, is the gain from the greater generality of such models enough to justify their use, even when we assume that we are using optimal parameters?

(ii) given that the number of parameters grows rapidly in models which rely on auxiliary multivariate chains, do those estimation methods which allow non-supervised treatments not run the risk of failing, particularly in the presence of high noise levels?

It is very difficult to answer these questions in general terms; furthermore, these models have only very recently been introduced, and so there are very few practical studies published on them. Nevertheless, the results from some studies have been very positive. Next we will present results for TMCs, modeling both the semi-Markov nature of the hidden chain and its non-stationarity [LAP 06].

Consider a TMC $T = (X, U, Y)$, with $U = (U^1, U^2)$, $X = (X_1, ..., X_n)$, $U^1 = (U^1_1, ..., U^1_n)$, $U^2 = (U^2_1, ..., U^2_n)$ and $Y = (Y_1, ..., Y_n)$. The variables X_i, U^1_i, U^2_i and Y_i take their values from $\Omega = \{\omega_1, ..., \omega_k\}$, $\Lambda_1 = \mathbb{N}^*$, $\Lambda_2 = \{1, ..., m\}$ and $\mathbb{R}$ respectively. The process U^1 models the semi-Markov nature of X, and the process U^2 models its non-stationarity. The distribution of T is given by $p(t_1)$ and the transitions $p(t_{i+1}|t_i)$, which can be written in a number of different ways. Let us consider the following specific case:

$$p(t_{i+1}|t_i) = p(u^2_{i+1}|x_i, u^1_i, u^2_i)\times$$

$$\times p(x_{i+1}|u^2_{i+1}, x_i, u^1_i, u^2_i)p(u^1_{i+1}|x_{i+1}, u^2_{i+1}, x_i, u^1_i, u^2_i)p(y_{i+1}|x_{i+1}) \qquad (4.35)$$

with:

$$p(u^2_{i+1}|x_i, u^1_i, u^2_i) = \delta_{u^2_i}(u^2_{i+1}) \text{ if } u^1_i > 1,$$

$$\text{and } p(u^2_{i+1}|u^2_i) \text{ if } u^1_i = 1; \qquad (4.36)$$

$$p(x_{i+1}|u_{i+1}^2, x_i, u_i^1, u_i^2) = \delta_{x_i}(x_{i+1}) \text{ if } u_i^1 > 1,$$
$$\text{and } p(x_{i+1}|u_{i+1}^2, x_i) \text{ if } u_i^1 = 1; \tag{4.37}$$
$$p(u_{i+1}^1|x_i, x_{i+1}, u_i^1, u_i^2, u_{i+1}^2) = \delta_{u_i^1-1}(u_{i+1}^1) \text{ if } u_i^1 > 1,$$
$$\text{and } p(u_{i+1}^1|x_{i+1}, u_{i+1}^2) \text{ if } u_i^1 = 1. \tag{4.38}$$

We will also assume that for $u_i^1 = 1$ in (4.37) we have $p(x_{i+1} = x_i|u_{i+1}^2, x_i) = 0$; the realization $U_i^1 = u_i^1$ then represents the exact remaining sojourn time for the chain X in state x_i.

This type of model was investigated in [LAP 06], and the first numerical results are promising. These results involve a model that is slightly different to the preceding one, with the difference being that $\Lambda_1 = \{1, ..., l\}$ is considered to be finite and, for $u_i^1 = 1$ in (4.37), it is assumed that $p(x_{i+1} = x_i|u_{i+1}^2, x_i)$ can be non-zero; the realization $U_i^1 = u_i^1$ represents the minimum sojourn time of the chain X in state x_i. Such a model is similar to a very classical type of hidden Markov chain model. The chain $V = (X, U^1, U^2)$ is a finite discrete chain, with each V_i taking its values from $\Omega \times \Lambda_1 \times \Lambda_2 = \{\omega_1, ..., \omega_k\} \times \{1, ..., l\} \times \{1, ..., m\}$, and noise is introduced in the classical manner through $p(y_i|v_i) = p(y_i|x_i)$. With T assumed stationary, its distribution is defined by $p(t_1, t_2) = p(v_1, v_2)p(y_1|x_1)p(y_2|x_2)$. In the Gaussian case, we then have, in the most general case, $(klm)^2$ parameters defining $p(v_1, v_2)$, as well as k means, and k variances defining the distributions $p(y_i|x_i)$.

We note that the assumption $p(y_i|v_i) = p(y_i|x_i)$ implies that the chains Y and (U^1, U^2) are independent conditional on the chain X, which might, *a priori*, be expected to lead to difficulties in estimating (U^1, U^2). However, no such difficulty appears in practice. We will next present two sets of results.

In the first, we will simulate a triplet Markov chain $T = (X, U, Y)$ satisfying (4.35-4.38), with (4.37) modified as discussed earlier. Then, the realization $Y = y$ will be segmented using three MPM methods, which will be based on:

– classical hidden Markov chains with independent noise (HMC-IN);

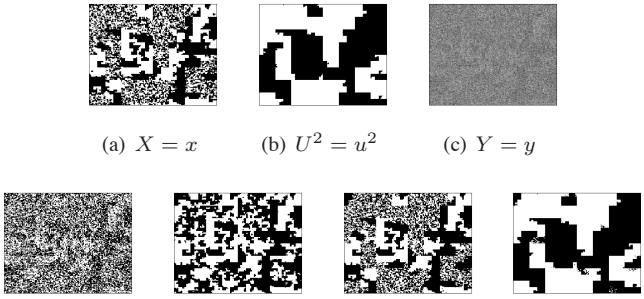
– hidden semi-Markov chains with independent noise (HSMC-IN);

– the true TMC model; this is a non-stationary hidden semi-Markov chain (NSHSMC-IN).

The aim of this investigation is to test whether the first two methods are able to approach the optimal solution, given by the third method, in terms of the quality of their solutions.

The TMC model to be considered is as follows. Let $k = 2$, $l = 5$ and $m = 2$. The two Gaussians modeling the noise have the same common variance, equal to 1,

and means of 1 and 1.5. In addition, we will take $p(u_{i+1}^2|u_i^2) = \begin{pmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{pmatrix}$ in (4.36),
$p(x_{i+1}|u_{i+1}^2 = 0, x_i) = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$ and $p(x_{i+1}|u_{i+1}^2 = 1, x_i) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$ in (4.37), and $p(u_{i+1}^1|x_{i+1}, u_{i+1}^2) = 0.1$ in (4.38).



(a) $X = x$ (b) $U^2 = u^2$ (c) $Y = y$

(d) HMC-IN, error of 34% (e) HSCM-IN, error of 22% (f) NSHSMC-IN, error of 17% (g) $\bar{u}^2$

Figure 4.2. *Realization of (X, U^2, Y), and non-supervised segmentations with HMC-IN, HSMC-IN and NSHSMC. $\bar{u}^2$ is the estimate of $U^2 = u^2$ with NSHSMC-IN.*

The parameters that define the respective distributions of X required for the calculation of segmentations based on HMC-IN and HSMC-IN models are estimated from X. The results shown in Figure 4.2 show the method based on the classical HMC-IN gives fairly mediocre results, both visually and quantitatively, with an error level of 34%. These results are improved by the method based on the HSMC-IN model; visually, the two stationarities can start to be distinguished, and the error level falls to 22%. This first comparison shows that the semi-Markov nature of the hidden data cannot be ignored, despite the very good robustness of classical HMC-IN models. It is also interesting to note that this first comparison was made in the context of non-stationary data, which is not the case for either of the two models being compared. Finally, the optimal segmentation using the NSHSMC-IN model gives very good visual results, with an error level of 17%. This example thus demonstrates a simple situation where classical HMC-IN models do not work; furthermore, the simultaneous use of two auxiliary chains leads to improved results compared to those obtained by using only one of the chains.

In the second series we will consider data which does not correspond to any of the three models listed above, and the aim is to show that there are situations in which the use of TMCs, which use the most complete and most complex model of the three

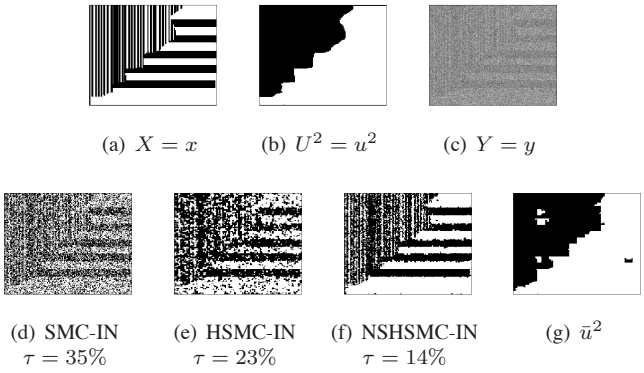
types, is useful. The parameters required for the implementation of the various methods will be estimated using ICE.

We will consider a hand-drawn image, represented in Figure 4.3. We will distinguish two types of stationarities – or "textures" – which will be considered as a realization u^2 of the auxiliary chain U^2, and also shown in Figure 4.3. The image of the classes $X = x$ then has noise added in the form of two Gaussians with means of 1 and 1.5, and variances both equal to 1, which gives the observed image $Y = y$. As earlier, the latter is then segmented by three non-supervised MPM Bayesian methods, with the parameters being estimated by ICE type methods, based on the three HMC-IN, HSMC-IN and NSHSMC-IN models. We can see that the trend from the previous set remains: the use of the HMC-IN model gives poor results, with an error level of 35%, HSMC-IN improves on this with an error level of 23%, and NSHSMC-IN gives still better results, with an error level of 14%. In addition, the three methods of estimation of the noise parameters give comparable results, shown in Table 4.1. From this we can conclude that despite the increasing generality of the models, which implies a growing number of parameters to be estimated, and hence an increasing difficulty of estimating them in a practical manner, the quality of the estimate obtained in all three models is comparable.

Thus this second series of experiments has helped confirm the existence of situations in which questions (i) and (ii), put forward at the start of the section, can be answered. It is clear that non-supervised treatments based on HMC-IN can be ineffective, and that NSHSMC-IN models can reliably produce impressive solutions. It is also interesting to observe that the results based on HSMC-INs have an intermediate level of quality. This demonstrates that the addition of successive auxiliary chains improves the results obtained for data which does not, *a priori*, follow any model. In terms of the behavior of the ICE, we note the good estimation of the noise parameters. Finally, we should observe that the Peano scan applied to the set of pixels produces a non-stationary chain, even in the case of "visually" stationary images, which reinforces the importance of these results.

4.9. Conclusions and outlook

In this chapter we have presented a number of recent results involving "triplet Markov chain" models (TMC) $T = (X, U, Y)$, where the observed process Y takes continuous values, whereas the hidden process X, as well as the auxiliary process U take values from finite sets. Such a model offers a very wide variety of specific models. Some of these are extensions of classical models; others allow the treatment of specific cases which are important in the real world, such as non-stationary random processes. TMCs of increasing complexity also allow a unified treatment of problems which combine a number of different classical aspects. In particular, we presented results of



(a) $X = x$ (b) $U^2 = u^2$ (c) $Y = y$

(d) SMC-IN $\tau = 35\%$ (e) HSMC-IN $\tau = 23\%$ (f) NSHSMC-IN $\tau = 14\%$ (g) $\bar{u}^2$

Figure 4.3. *Realization of (X, U^2, Y), and non-supervised segmenations using HMC-IN, HSMC-IN and NSHSMC. τ is the error level, and $\bar{u}^2$ is the estimate of $U^2 = u^2$ using NSHSMC-IN.*

Classe	HMC-IN		HSMC-IN		NSHSMC-IN	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
ω_1	0.84	0.91	1.09	1.04	0.9	0.94
ω_2	1.65	0.89	1.46	1.02	1.49	0.99
τ	35%		23%		14%	

Table 4.1.
Estimation of noise parameters, corresponding to the data in Figure 4.3, using the ECI method in the HMC-IN, HSMC-IN and NSHSMC-IN models. τ is the error level obtained by applying the MPM Bayesian method using the estimated parameters.

segmentation of non-stationary hidden semi-Markov chains. Finally, copula theory, recently introduced in the context of pairwise Markov models [BRU 05a, BRU 05b], offers the possibility of highly detailed noise models. In all the cases considered, the general "iterative conditional estimation" (ICE) learning method allows the parameters for a wide range of models to be estimated, and thus lays the foundation for non-supervised Bayesian segmentation methods.

We note that similar studies are currently being carried out on models where the three chains X, U and Y are all continuous, and a number of promising results have been published already [DES 03a, DES 03b, FQU 06]. The addition of U allows classical filtering or smoothing problems to be considered within more general models, with the aim of providing better correspondence to real data. In particular, it can be

shown that all these treatments, whether exact due to the Gaussian nature of the processes of interest [DES 03a, FQU 06], or approximated through specific filtering or smoothing techniques [DES 03b], can easily be extended to new models. As in the case considered in the current chapter, the introduction of the auxiliary chain leads to highly general models, in which (X, Y) is not necessarily Markovian [DES 03a].

Other possibilities exist of mixing chains with continuous values and chains with discrete values, and can lead to new models and associated treatments, which may provide better correspondence with particular situations, thereby improving the results obtained using classical approaches. We refer the reader to the example considered in [BEN 07a], where the chain X is discrete whereas the chains U and Y are continuous. Finally, the choice between "discrete" and "continuous" is not exhaustive, and we could consider chains formed from variables which can take "mixed" distributions, consisting of both a continuous and a discrete part. The first models of this type were proposed in the context of independent data in [CAI 93]; then, a number of hidden Markov models with "mixed states" were introduced (fields [PIE 94], chains [GER 02], and trees [MON 03]). These various models are normally used for fuzzy segmentation [CAR 06, RUA 00, RUA 02, SAL 06]; we should also mention a more general interpretation of the continuous part of the mixed measure, recently used to model textures in [BOU 06]. All these models can be extended to triplet models, and the first study of this type, which treats the non-stationarity, was very promising [SAL 07].

The "Markov chain" models studied in this chapter can relatively directly be extended to "Markov tree" models, which also have many possible applications, notably in multi-resolution imaging. The first "Markov pair" type models [DES 05, DES 06, MON 03], which can be easily generalized to Markov "triplets", can therefore easily benefit from the various generalizations studied, or simply suggested, in the context of chains. In addition, a certain number of considerations discussed in this chapter in the context of Markov chains can easily be adapted to Markov fields. Triplet Markov fields [BEN 05a] thus offer the possibility of modeling and treating non-stationary fields [BEN 05b], or treatments of the important problem of texture classification [BEN 05b, BLA 08]. We note that a link with the "theory of evidence", not discussed in this chapter, was also established and studied in [BEN 05a, PIE 06]. Other models described in this chapter in the context of Markov chains, in particular those involving multivariate auxiliary fields or mixed probability measures, are among the most promising prospects. We note that these different considerations can be extended to general graphical Markovian models. Furthermore, these additions can be envisaged in the context of "partially Markov" models, as proposed (for chains) in [PIE 05]. The first studies of these models, which allow the introduction of "long memory" noise, have produced encouraging results [LAN 08, LAP 07].

Finally, we note the possibility of considering multivariate observations, including those of different natures, as in [BRI 06, BRI 08], which further increases the range of original triplet Markov models that can be constructed.

4.10. Bibliography

[BEL 00] BELLONE E., HUGUES J., GUTTORP P., "A hidden Markov model for downscaling synoptic athmospheric patterns to precipitation amounts", *Climate research*, vol. 15, num. 1, p. 1-15, 2000.

[BEN 01] BENDJEBBOUR A., DELIGNON Y., FOUQUE L., SAMSON V., PIECZYNSKI W., "Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context", *Trans. on Geoscience and Remote Sensing*, vol. 39, num. 8, p. 1789-1798, 2001.

[BEN 05a] BENBOUDJEMA D., "Champs de Markov triplets et segmentation bayésienne non supervisée d'images", Thesis, National Telecommunications Institute, Evry, France, 2005.

[BEN 05b] BENBOUDJEMA D., PIECZYNSKI W., "Unsupervised image segmentation using triplet Markov fields", *Computer Vision and Image Understanding*, vol. 99, num. 3, p. 476-498, 2005.

[BEN 07a] BEN MABROUK M., PIECZYNSKI W., "Unsupervised segmentation of random discrete data using triplet Markov chains", International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA), Chania, Crete, May 2007.

[BEN 07b] BENBOUDJEMA D., PIECZYNSKI W., "Unsupervised statistical segmentation of non stationary images using triplet Markov fields", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, num. 8, p. 1367-1378, 2007.

[BLA 08] BLANCHET J., FORBES F., "Triplet Markov fields for the supervised classification of complex structure data", *IEEE Trans. on Pattern Analysis and Machine*, vol. 30, num. 6, p. 1055-1067, 2008.

[BOU 06] BOUTHEMY P., HARDOUIN C., PIRIOU G., YAO J.-F., "Mixed-state auto-models and motion texture modeling", *Journal of Mathematical Imaging and Vision*, vol. 25, num. 3, p. 387-402, 2006.

[BRI 06] BRICQ S., COLLET C., ARMSPACH J.-P., "Triplet Markov chains for 3D MRI brain segmentation using a probabilistic atlas", Proceedings of IEEE International Symposium on Biomedical Imaging, ISBI'06, Arlington, Virginia, USA, 6-9 April, 2006.

[BRI 08] BRICQ S., COLLET C., ARMSPACH J.-P., "Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains", *Medical Image Analysis. Elsevier*, vol. doi:10.1016/j.media.2008.03.001, 2008.

[BRU 05a] BRUNEL N., "Sur quelques extensions des chaînes de Markov cachées et couples, application à la segmentation non supervisée des signaux radar", Thesis, Paris University VI.

[BRU 05b] BRUNEL N., PIECZYNSKI W., "Unsupervised signal restoration using hidden Markov chains with copulas", *Signal Processing*, vol. 85, num. 12, p. 2304-2315, 2005.

[CAI 93] CAILLOL H., HILLION A., PIECZYNSKI W., "Fuzzy random fields and unsupervised image segmentation", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 34, num. 4, p. 801-910, 1993.

[CAP 05] CAPPÉ O., MOULINES E., RYDEN T., *Inference in hidden Markov models*, Springer, Series in Statistics, New York, 2005.

[CAR 06] CARINCOTTE C., DERRODE S., BOURENNANE S., "Unsupervised change detection on SAR images using fuzzy hidden Markov chains", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, num. 2, p. 432-441, 2006.

[CHE 94] CHEN M., KUNDU A., ZHOU J., "Off-Line Handwritten Work Recognition Using a Hidden Markov Model Type Stochastic Network", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, num. 5, 1994.

[DEL 97] DELMAS J.-P., "An equivalence of the EM and ICE algorithm for exponential family", *IEEE Trans. on Signal Processing*, vol. 45, num. 10, p. 2613-2615, 1997.

[DEL 02] DELIGNON Y., PIECZYNSKI W., "Modeling non-Rayleigh speckle distribution in SAR images", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, num. 6, p. 1430-1435, 2002.

[DER 04] DERRODE S., PIECZYNSKI W., "Signal and Image segmentation using pairwise Markov chains", *IEEE Trans. on Signal Processing*, vol. 52, num. 9, p. 2477-2489, 2004.

[DES 03a] DESBOUVRIES F., PIECZYNSKI W., "Modèles de Markov Triplet et Filtrage de Kalman", *Comptes Rendus de l'Académie des Sciences-Mathématiques*, vol. 336-8, p. 667-70, 2003.

[DES 03b] DESBOUVRIES F., PIECZYNSKI W., "Particle Filtering in pairwise and triplet Markov chains", Grado-Gorizia, Italy, Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003), June 8-11 2003.

[DES 05] DESBOUVRIES F., LECOMTE J., "Multiscale Bayesian restoration in pairwise Markov trees", *IEEE Trans. on Automatic Control*, vol. 50, num. 8, p. 1185-1190, 2005.

[DES 06] DESBOUVRIES F., LECOMTE J., PIECZYNSKI W., "Kalman filtering in pairwise Markov trees", *Signal Processing*, vol. 86, num. 5, p. 1049-54, 2006.

[DIZ 07] DI ZIO M., GUAMERA U., ROCCI R., "A mixture of mixture models for a classification problem: the unity measure error", *Computational Statistics & Data Analysis*, vol. 51, num. 5, p. 2573-2585, 2007.

[FAI 05] FAISAN S., THORAVAL L., ARMSPACH J., METZ-LUTZ M.-N., HEITZ F., "Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models", *IEEE Trans. on Medical Imaging*, vol. 24, num. 2, p. 263-276, 2005.

[FQU 06] AIT-EL FQUIH B., DESBOUVRIES F., "Kalman filtering in triplet Markov chains", *IEEE Trans. on Signal Processing*, vol. 54, num. 8, p. 2957-63, 2006.

[GER 02] GERMAIN M., VOORONS M., BOUCHER J.M., BENIE G.B., "Fuzzy statistical classification method for multiband image fusion", *Proceedings of the Fifth International Conference on Information Fusion*, vol. 1, p. 178-184, 2002.

[GIO 97] GIORDANA N., PIECZYNSKI W., "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, num. 5, p. 465-475, 1997.

[GRE 00] GREGOIR S., LENGLART F., "Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden Markov model", *Journal of forecasting*, vol. 19, num. 2, p. 81, 2000.

[GUE 05] GUEDON Y., "Hidden hybrid Markov/semi-Markov chains", *Computational Statistics & Datta Analysis*, vol. 49, num. 3, p. 663-688, 2005.

[KOS 01] KOSKI T., *Hidden Markov models for bioinformatics*, Kluwer Academic Publishers, 2001.

[LAN 04] LANCHANTIN P., PIECZYNSKI W., "Unsupervised non stationary image segmentation using triplet Markov chains", Advanced Concepts for Intelligent Vision Systems (ACIVS 04), Brussels, Belgium, 31 August-3 September, 2004.

[LAN 05] LANCHANTIN P., PIECZYNSKI W., "Unsupervised restoration of hidden non stationary Markov chain using evidential priors", *IEEE Trans. on Signal Processing*, vol. 53, num. 8, p. 3091-3098, 2005.

[LAN 06] LANCHANTIN P., "Chaînes de Markov Triplets et segmentation non supervisée des signaux", Thesis, National Telecommunications Institute, Evry, 2006.

[LAN 08] LANCHANTIN P., LAPUYADE-LAHORGUE J., PIECZYNSKI W., "Unsupervised Segmentation of triplet Markov chains hidden with long-memory noise", *Signal Processing*, vol. 88, num. 5, p. 1134-1151, 2008.

[LAP 06] LAPUYADE-LAHORGUE J., PIECZYNSKI W., "Unsupervised Segmentation of hidden semi-Markov non stationary chains", Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MaxEnt2006, Paris, France, July 8-13, 2006.

[LAP 07] LAPUYADE-LAHORGUE J., PIECZYNSKI W., "Partially Markov models and unsupervised segmentation of semi-Markov chains with long-dependent noise", International Symposium on Applied Stochastic Models and Data Analysis, Chania, Crete, May 2007.

[LEB 06] LE BER F., BENOÎT M., SCOTT C., MARI J.-F., MIGNOLET C., "Studying crop sequences with CarrotAge, a HMM-based data mixing software", *Ecological Modelling*, vol. 191, num. 1, p. 170-185, 2006.

[MAI 01] MAÎTRE H., *Traitement des images RSO*, Hermès, Collection IC2, Paris, 2001.

[MCL 97] MCLACHLAN G., KRISHNAN T., *EM algorithm and extensions*, John Wiley & Sons, Series in Probabilities and Statistics, New York, 1997.

[MON 03] MONFRINI E., LECOMTE J., DESBOUVRIES F., PIECZYNSKI W., "Image and Signal Restoration using Pairwise Markov Trees", IEEE Workshop on Statistical Signal Processing (SSP 2003), Saint Louis, Missouri, 28 September- 1st October, 2003.

[MOO 04] MOORE M., SAVIC M., "Speech reconstruction using a generalized HSMM (GHSMM)", *Digital Signal Processing*, vol. 14, num. 1, p. 37-53, 2004.

[NEL 98] NELSEN R., *An introduction to Copulas*, n° 139 in Lecture notes in Statistics, Springer-Verlag, London, 1998.

[NIC 02] NICOLAS P., BIZE L., MURI-MAJOUBE F., HOEBEKE M., RODOLPHE F., DUSKO EHRLICH S., PRUM B., BESSIÈRES P., "Mining bacillus subtilis chromosome heterogeneities using hidden Markov models", *Nucleic Acid Research*, vol. 30, 2002.

[NUE 07] NUEL G., PRUM B., *Analyse statistique des séquences biologiques: modélisation markovienne, alignements et motifs*, Hermès, Collection Bioinformatique, Paris, 2007.

[PIE 92] PIECZYNSKI W., "Statistical image segmentation", *Machine Graphics and Vision*, vol. 1, num. 1/2, p. 261-268, 1992.

[PIE 94] PIECZYNSKI W., CAHEN J. M., "Champs de Markov flous cachés et segmentation d'images", *Revue de Statistique Appliquée*, vol. 42, num. 3, p. 13-31, 1994.

[PIE 02] PIECZYNSKI W., HULARD C., VEIT T., "Triplet Markov Chains in hidden signal restoration", SPIE's International Symposium on Remote Sensing, 22-27 September, Crete, 2002.

[PIE 03] PIECZYNSKI W., "Pairwise Markov chains", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, num. 5, p. 634-639, 2003.

[PIE 05] PIECZYNSKI W., "Copules gaussiennes dans les chaînes triplet partiellement de Markov", *Comptes Rendus de l'Académie des Sciences-Mathématiques*, vol. 341, num. 3, p. 189-194, 2005.

[PIE 06] PIECZYNSKI W., BENBOUDJEMA D., "Multisensor triplet Markov fields and theory of evidence", *Image and Vision Computing*, vol. 24, num. 1, p. 61-69, 2006.

[PIE 07] PIECZYNSKI W., "Multisensor triplet Markov chains and theory of evidence", *International Journal of Approximate Reasoning*, vol. 45, num. 1, p. 1-16, 2007.

[PIE 08] PIECZYNSKI W., "Sur la convergence de l'estimation conditionnelle itérative", *Comptes Rendu de l'Académie des Sciences-Mathématiques.*, vol. 346, num. 7-8, p. 457-460, 2008.

[RAP 99] RAPHAEL C., "Automatic segmentation of acoustic musical signals using hidden Markov models", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, num. 4, p. 360, 1999.

[RUA 00] RUAN S., JAGGI C., XUE J., FADILI J., BLOYET D., "Brain tissue classification of magnetic resonance images using partial volume modeling", *IEEE Trans. on Medical Imaging*, vol. 19, num. 12, p. 1179-1187, 2000.

[RUA 02] RUAN S., MORETTI B., FADILI J., BLOYET D., "Fuzzy Markovian segmentation in application of magnetic resonance images", *Computer Vision and Image Understanding*, vol. 85, p. 54-69, 2002.

[SAL 06] SALZENSTEIN F., COLLET C., "Fuzzy Markov random fields versus chains for multispectral image segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, num. 11, p. 1753-1767, 2006.

[SAL 07] SALZENSTEIN F., COLLET C., LECAM S., HATT M., "Non stationary fuzzy Markov chains", *Pattern Recognition Letters*, vol. 28, num. 16, p. 2201-2208, 2007.

[THO 02] THOMAS L., ALLEN D., MORKEL-KINGSBURY N., "A hidden Markov chain model for the term structure of bond credit risk spreads", *International Review of Financial Analysis*, vol. 11, num. 3, p. 311-329, 2002.

[YU 03] YU S.-Z., KOBAYASHI H., "A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking", *Signal Processing*, vol. 83, num. 2, p. 235-250, 2003.