Countable Partially Exchangeable Mixtures

Cecilia Prosdocimi,
Dipartimento di Economia e Finanza, Università LUISS,
viale Romania 32, 00197 Roma, Italy,
email: cprosocimi@luiss.it

Lorenzo Finesso,
Istituto di Ingegneria Biomedica, Consiglio Nazionale delle Ricerche,
corso Stati Uniti 4, 35127 Padova, Italy,
email: lorenzo.finesso@isib.cnr.it

May 22, 2014

Abstract

Partially exchangeable sequences representable as mixtures of Markov chains are completely specified by de Finetti’s mixing measure. The paper characterizes, in terms of a subclass of hidden Markov models, the partially exchangeable sequences with mixing measure concentrated on a countable set, both for discrete valued and for Polish valued sequences.

1 Introduction

In the Hewitt-Savage generalization of de Finetti’s theorem, Polish space valued exchangeable sequences are shown to be in one to one correspondence with mixtures of i.i.d. sequences, and ultimately with the mixing measure defining the mixture. The mixing measure thus acts as a model of the exchangeable sequence and its properties shed light on the random mechanism generating the sequence. In this regard [5], using the connection with the Markov moment problem, characterizes the subclass of discrete valued exchangeable sequences whose mixing measures are absolutely continuous, and
have densities in $L^p$. It is also of interest to characterize the subclass for which the mixing measure is singular. A contribution in this direction has been given by [3], where it is proved that a discrete valued exchangeable sequence has de Finetti mixing measure concentrated on a countable set if and only if it is a hidden Markov model (HMM), i.e. a probabilistic function of a homogeneous Markov chain.

The class of partially exchangeable sequences are in one to one correspondence with mixtures of Markov chains. As noted in [5], the results on the regularity of the mixing measure carry to partially exchangeable sequences. On the other hand, to the best of our knowledge, no results have been reported concerning partially exchangeable sequences with singular mixing measures.

The goal of the present paper is to contribute a first result in this direction, characterizing, in the spirit of [3], countable mixtures of Markov chains. Our results hold both for discrete and for Polish valued sequences. This has required the development of a few special results for HMMs, previously not available in the literature. Of independent interest are Proposition 1 on the rows of the array of the successors of an HMM, and most of the Appendix, on properties of sequences of stopping times with respect to the filtration generated by the state and the output of an HMM. In the Polish valued case it has also been necessary to first extend [3] to show the equivalence between exchangeable HMMs and countable mixtures of Polish valued i.i.d. sequences.

In Section 2 we review the basic notions in the setup most convenient for our purpose. The reader should be aware of the fact that slightly different notions of partial exchangeability coexist in the literature for discrete valued sequences. We review the original definition, introduced in [2] and elaborated in [8]. The latter paper clarifies the relationship with the alternative definition given in [4]. Section 3 concerns discrete valued sequences, Section 4 Polish valued sequences. In Section 3.1 we constructively prove Proposition 1, which is instrumental in the balance of the paper. In Section 3.2 we characterize discrete valued countable mixtures of Markov chains. In Section 4 we extend the result of [3] to exchangeable sequences with values in a Polish space. This allows a characterization as HMMs of Polish valued partially exchangeable sequences with singular mixing measure. Few final remarks and a hint at possible extensions are in Section 5. The Appendix contains some results on sequences of stopping times for the HMM, unavailable elsewhere in the literature, and a technical result for a special class of
HMMs representable as mixtures of i.i.d. sequences.

2 Preliminaries

\((Y_n)_{n \geq 0}\) denotes a sequence of random variables on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), taking values in a Polish space \(S\) endowed with the Borel \(\sigma\)-field \(\mathcal{S}\). The generic element of \(S\) is denoted \(y\), and \(y_0^N = y_0y_1\ldots y_N\) is an element (a string) of the \(N + 1\)-th fold Cartesian product \(S^{N+1}\), likewise \((Y_0^N = y_0^N)\) is the event \((Y_0 = y_0, Y_1 = y_1,\ldots, Y_N = y_N)\).

**Exchangeable and partially exchangeable sequences.** An infinite sequence of random variables \((Y_n)\) is exchangeable if it is distributionally invariant under finite permutations (see [8] page 24). An infinite sequence of random variables \((Y_n)\) is partially exchangeable if its successors array \(V\) is distributionally invariant under finite permutations within each of its rows. For the precise definition of the successors array and for notations we refer to [9], Section 3 for discrete state space and Section 4 for Polish state space. Note that the rows of the successors array of a partially exchangeable sequence are exchangeable.

**Mixtures of i.i.d. sequences and of Markov chains.** The set of all probability measures on \((S, \mathcal{S})\) is denoted \(\mathcal{M}_S\) and it is equipped with the \(\sigma\)-field generated by the maps \(p \mapsto p(A)\), varying \(p\) in \(\mathcal{M}_S\) and \(A\) in \(\mathcal{S}\).

**Definition 1.** \((Y_n)\) is a mixture of i.i.d. sequences if there exists a random variable \(\tilde{p}\) with values in \(\mathcal{M}_S\) such that for any \(A_0,\ldots, A_N \in \mathcal{S}\)

\[
\mathbb{P}(Y_0 \in A_0,\ldots, Y_N \in A_N \mid \tilde{p}) = \tilde{p}(A_0)\ldots \tilde{p}(A_N) \quad \mathbb{P} - \text{a.s.} \tag{1}
\]

**Definition 2.** A mixture of i.i.d. sequences is countable (finite) if the random variable \(\tilde{p}\) is countably (finitely) valued.

Let \(H\) be a countable (finite) set of indices. For a countable (finite) mixture of i.i.d. sequences, denote with \((p_h(\cdot))_{h \in H}\) the possible values taken by \(\tilde{p}\). Integrating over \(\Omega\), equation (1) reads

\[
\mathbb{P}(Y_0 \in A_0,\ldots, Y_N \in A_N) = \sum_{h \in H} \mu_hp_h(A_0)\ldots p_h(A_N), \tag{2}
\]

with \(\mu_h := \mathbb{P}(\tilde{p} = p_h) > 0\) and \(\sum_{h \in H} \mu_h = 1\).
Partially exchangeable sequences are in one to one correspondence with mixtures of Markov chains whose transition kernels \( k_t(\cdot, \cdot) : S^* \times S^* \rightarrow [0, 1] \) are constant, with respect to the first variable, on elements \( (E_j)_{j \geq 0} \) of a partition of \( S^* = S \cup \delta \) where \( \delta \) is a fictitious state, and \( S^* \) is the Borel \( \sigma \)-field of \( S^* \). For the precise definitions and notations refer again to [8]. In short terms, let \( T^* \) be the set of kernels \( t : \mathbb{N}_0 \times S^* \rightarrow [0, 1] \) and, for any \( t \in T^* \), define

\[
k_t(y, A) := \sum_{j \geq 0} I_{E_j}(y) t(j, A),
\]

where \( I_{E_j}(\cdot) \) is the indicator function of \( E_j \). We will consider mixtures of Markov chains, say \( (Y_n) \), with fixed initial state \( y_0 \in E_1 \), i.e. satisfying

**Condition 1.** \( (Y_0 = y_0) = \Omega, \ y_0 \in E_1 \).

We are ready to give the following

**Definition 3.** \( (Y_n) \) satisfying Condition 1 is a mixture of homogeneous Markov chains if there exists a random element \( \bar{t} \) with values in \( T^* \), such that, for any choice of \( A_1, \ldots, A_N \in S^* \),

\[
\mathbb{P}(Y_1 \in A_1, \ldots, Y_N \in A_N | \bar{t}) = \int_{A_1} \ldots \int_{A_N} k_{\bar{t}}(y_0, dy_1) \ldots k_{\bar{t}}(y_N-1, dy_N) \quad \mathbb{P}\text{-a.s.}
\]

(4)

Note that, by definition of \( k_{\bar{t}} \),

\[
k_{\bar{t}}(y_0, dy_1) \ldots k_{\bar{t}}(y_N-1, dy_N) = \bar{t}(1, dy_1) \ldots \bar{t}(j_N-1, dy_N),
\]

where \( j_n \) is the element of the partition to which \( y_n \) belongs.

**Definition 4.** A mixture of Markov chains is countable (finite) if the random element \( \bar{t} \) is countably (finitely) valued.

For a countable (finite) mixture of Markov chains equation (4) reads

\[
\mathbb{P}(Y_1 \in A_1 \ldots Y_N \in A_N) = \sum_{h \in H} \mu_h \int_{A_1} \ldots \int_{A_N} t_h(1, dy_1) \ldots t_h(j_N-1, dy_N),
\]

(5)

where \( H \) is a countable (finite) set, \( t_h \) are the possible values taken by \( \bar{t} \), and \( \mu_h := \mathbb{P}(\bar{t} = t_h) \).
Remark 1. To define mixtures of Markov chains when $S^*$ is discrete, less technicalities are needed since transition kernels reduce to transition matrices. When this is the case

$$\mathbb{P}(Y_1^N = y_1^N | \tilde{P}) = \tilde{P}_{y_0y_1}\tilde{P}_{y_1y_2}\ldots\tilde{P}_{y_{N-1}y_N} \quad \mathbb{P}\text{-a.s.,}$$

(6)

where $\tilde{P} = (\tilde{P}_{i,j})_{i,j \in S^*}$ is a random matrix varying on the set $\mathcal{P}^*$ of the transition matrices. The analog of equation (5) is

$$\mathbb{P}(Y_1^N = y_1^N) = \sum_{h \in H} \mu_h P_h^{y_0y_1}\ldots P_h^{y_{N-1}y_N},$$

(7)

where $(P^h)_{h \in H}$ are the possible values taken by $\tilde{P}$.

Representation theorems. The classic de Finetti’s representation theorem characterizes exchangeable sequences as mixtures of i.i.d. sequences. The statement, the proof, and an extensive discussion of the ramifications of the theorem can be found in [1] and [9]. Partially exchangeable sequences are in one-to-one correspondence with mixtures of Markov chains, as proven by the representation theorem of [8]. An earlier representation theorem, for discrete state space and based on a slightly different notion of partial exchangeability, was given in [4].

3 Countable Markov mixtures with discrete state space

In this section $S$ is a discrete set.

3.1 The successors array of hidden Markov models

The definition and some useful properties of HMMs are given in the appendix. The following result will be instrumental later in the paper and it is also of independent interest.

**Proposition 1.** Let $(Y_n)$ be a HMM with recurrent underlying Markov chain, then each row of $(V_{y,n})$ is a HMM with recurrent underlying Markov chain.

A homogeneous Markov chain $(X_n)$ is recurrent if $\mathbb{P}(X_n = x \text{ i.o. } n \mid X_0 = x) = 1$. Such Markov chains have no transient states but possibly more than one recurrence class.
Proof. Denote with $\mathcal{X}$ the discrete state space of the Markov chain $(X_n)_{n \geq 0}$, underlying the process $(Y_n)$, and let, for any $x \in \mathcal{X}$ and $y \in S$,

$$f_x(y) := \mathbb{P}(Y_n = y \mid X_n = x)$$

be the read-out distributions. Fix $y \in S$. To prove the theorem, we construct a recurrent Markov chain $(W^n_y)_{n \geq 1}$, such that $(V^n_{y,n})_{n \geq 1}$ is a HMM with underlying Markov chain $(W^n_y)$. The Markov chain $(W^n_y)$ is based on the values $(X_n)$ immediately following some random times $(\tau^n_y)$. More precisely, define inductively the random times of the $n$-th visit of $(Y_n)$ to the state $y$:

$$\tau^n_y := \inf\{t \geq 0 \mid Y_t = y\},$$

$$\tau^n_{y,n} := \inf\{t > \tau^n_{y,n-1} \mid Y_t = y\},$$

with the usual convention $\inf\emptyset = +\infty$. The random times $(\tau^n_y)$ are stopping times with respect to the filtration spanned by $(Y_n)$, and so are the times $(\tau^n_{y,n} + 1)$. Define the sequence

$$W^n_y := \begin{cases} \varepsilon & \text{for } \tau^n_{y,n} = +\infty \\ X_{\tau^n_{y,n}+1} & \text{for } \tau^n_{y,n} < +\infty \end{cases} \tag{8}$$

where $\varepsilon$ is a fictitious state. The sequence $W^n_y$ is either identically equal to $\varepsilon$, or it never hits it since the $\tau^n_{y,n}$ are either all finite or all infinite\footnote{If $Y_n = y$, for some finite $n$, then $X_n = x$, for some $x$ such that $f_x(y) > 0$. By the recurrence, $X$ hits $x$ infinitely many times, and thus $Y$ hits $y$ infinitely many times.}. We first check that $(W^n_y)$ is a Markov chain, then we verify that it is recurrent, and finally we show that $(V^n_{y,n})$ is a HMM with underlying Markov chain $(W^n_y)$.

If the case $W^n_y \equiv \varepsilon$ obtains, $(W^n_y)$ is a recurrent Markov chain. Otherwise a direct computation gives, for any $x_1, \ldots, x_N \in \mathcal{X}$,

$$\mathbb{P}(W^n_N = x_N \mid W^n_{N-1} = x_{N-1}, \ldots, W^n_1 = x_1)$$

$$= \mathbb{P}(X_{\tau^n_N + 1} = x_N \mid X_{\tau^n_{N-1} + 1} = x_{N-1}, \ldots, X_{\tau^n_1 + 1} = x_1)$$

$$= \mathbb{P}(X_{\tau^n_N + 1} = x_N \mid X_{\tau^n_{N-1} + 1} = x_{N-1}) = \mathbb{P}(W^n_N = x_N \mid W^n_{N-1} = x_{N-1}),$$

where Remark \footnote{If $Y_n = y$, for some finite $n$, then $X_n = x$, for some $x$ such that $f_x(y) > 0$. By the recurrence, $X$ hits $x$ infinitely many times, and thus $Y$ hits $y$ infinitely many times.} in the Appendix applies, as $(\tau^n_{y,n} + 1)$ are hitting-times of $\mathcal{X} \times y$ (see Definition \footnote{If $Y_n = y$, for some finite $n$, then $X_n = x$, for some $x$ such that $f_x(y) > 0$. By the recurrence, $X$ hits $x$ infinitely many times, and thus $Y$ hits $y$ infinitely many times.} in the Appendix). Thus $(W^n_y)$ is Markov. Since

$$\mathbb{P}(W^n_y = x \text{ i.o. } n \mid W^n_1 = x) = \mathbb{P}(X_{\tau^n_{y,n} + 1} = x \text{ i.o. } n \mid X_{\tau^n_{y,n} + 1} = x),$$
to check the recurrence of $(W_n^y)$ we have to verify that, for all $x \in \mathcal{X}$,
\[ \mathbb{P}(X_{\tau^y_n+1} = x \text{ i.o. } n \mid X_{\tau^y_1+1} = x) = 1. \] (9)

Choose $\bar{x} \in \mathcal{X}$ such that $f_\bar{x}(y) > 0$ and $\mathbb{P}(X_{n+1} = x \mid \tau^y_n = x) > 0$, (there exists at least one such $\bar{x}$). Define the auxiliary sequences of stopping times with respect to the $\sigma$-field spanned by $(X_n, Y_n)$:
\[ \sigma^x_{1,y} := \inf\{t \geq 0 \mid X_t = \bar{x}, Y_t = y\}, \]
\[ \sigma^x_{n,y} := \inf\{t > \sigma^x_{n-1,y} \mid X_t = \bar{x}, Y_t = y\}. \]
The stopping times $(\sigma^x_{n,y})$ are finite whenever $(\tau^y_n)$ are finite, and the sequence $(\sigma^x_{n,y})$ is a random subsequence of $(\tau^y_n)$, (with $\sigma^x_{n,y} > \tau^y_1$ for any $n > 1$), thus
\[ (X_{\sigma^x_{n,y}+1} = x) \subseteq \bigcup_{m \geq n} (X_{\tau^y_{m+1}} = x), \] (10)
and trivially
\[ (X_{\sigma^x_{0,y}+1} = x \text{ i.o. } n) \subseteq (X_{\tau^y_n+1} = x \text{ i.o. } n). \] (11)

The events $(X_{\sigma^x_{n,y}+1} = x)_{n}$ are independent under the law $\mathbb{P}(. \mid X_{\tau^y_1+1} = x)$, since $\{(X_{\sigma^x_{n,y}+1} = x), (X_{\tau^y_1+1} = x)\}$ is a $\mathbb{P}$-independent set. In fact, for any fixed sequence $m_1 < \cdots < m_n \in \mathbb{N}$,
\[
\mathbb{P}(X_{\sigma^x_{m_1,n+1}} = x, X_{\sigma^x_{m_2,n-1}+1} = x, \ldots, X_{\sigma^x_{m_1+1}+1} = x, X_{\tau^y_{m_1+1}} = x)
\]
\[
= \mathbb{P}(X_{\sigma^x_{m_1+1}} = x, X_{\sigma^x_{m_2}+1} = \bar{x}, \ldots, X_{\sigma^x_{m_1+1}+1} = x, X_{\sigma^x_{m_1}+1} = \bar{x}, X_{\tau^y_{m_1+1}} = x)
\]
\[
= \mathbb{P}(X_{\sigma^x_{m_1,y}+1} = x \mid X_{\sigma^x_{m_1+1}} = \bar{x}) \times \cdots \times \mathbb{P}(X_{\sigma^x_{m_1+1}+1} = x \mid X_{\sigma^x_{m_1+1}} = \bar{x})
\]
\[
\times \mathbb{P}(X_{\sigma^x_{m_1+1}} = \bar{x} \mid X_{\tau^y_{m_1+1}} = x)\mathbb{P}(X_{\tau^y_{m_1+1}} = x)
\]
\[
= \mathbb{P}(X_{\sigma^x_{m_1,y}+1} = x) \cdots \mathbb{P}(X_{\sigma^x_{m_1+1}+1} = x) \mathbb{P}(X_{\tau^y_{m_1+1}} = x),
\]
where the second equality follows by Lemma 4 in the Appendix, and the first and last equality follow noting that $X_{\sigma^x_{m_1, n}} = \bar{x}$ by definition of $\sigma^x_{n,y}$.\(^3\)

\(^3\)The case $m_1 = 1$ and $\sigma^x_{m_1,y} = \tau^y_{1} + 1$ needs some care, but the goal is to apply the Borel-Cantelli lemma, therefore the initial events do not matter.
The events \( (X_{\sigma_n^{x,y+1} = x}) \) are equiprobable, with strictly positive probability. By the Borel-Cantelli lemma
\[
P(X_{\sigma_n^{x,y+1} = x} \text{ i.o. } n \mid X_{\tau^{y+1} = x} = x) = 1.
\] (12)

Equations (11) and (12) taken together give
\[
P(X_{\tau^{y+1} = x} \text{ i.o. } n \mid X_{\tau^{y+1} = x} = x) \geq P(X_{\sigma_n^{x,y+1} = x} \text{ i.o. } n \mid X_{\tau^{y+1} = x} = x) = 1.
\]

Condition (9) is satisfied, thus the recurrence of \((W_y^n)\) is proved. We now prove that \((V_{y,n})\) is a HMM with underlying Markov chain \((W_y^n)\). Set
\[
P(V_{y,n} = \delta \mid W_y^n = \varepsilon) = 1.
\]

For \( \varepsilon \neq x \in X \) and \( \delta \neq \bar{y} \in S \), the couple \((W_y^n, V_{y,n})\) inherits the read-out distributions of \((X_n, Y_n)\):
\[
P(V_{y,n} = \bar{y} \mid W_y^n = x) = P(Y_{\tau_n^{y+1} = \bar{y} } \mid X_{\tau_n^{y+1} = x}) = f_x(\bar{y}),
\] (13)

this can be seen disintegrating the stopping time \( \tau_n^{y+1} + 1 \),
\[
\begin{align*}
P(Y_{\tau_n^{y+1} = \bar{y}} \mid X_{\tau_n^{y+1} = x}) \\
= \sum_{m \in \mathbb{N}} P(\tau_n^{y} + 1 = m, Y_m = \bar{y} \mid X_{m} = x) \\
= \sum_{m \in \mathbb{N}} P(Y_m = \bar{y} \mid \tau_n^{y} + 1 = m, X_{m} = x) P(\tau_n^{y} + 1 = m, X_{m} = x) \\
= \sum_{m \in \mathbb{N}} P(Y_m = \bar{y} \mid X_{m} = x) P(\tau_n^{y} + 1 = m, X_{m} = x) \\
= f_x(\bar{y}) \sum_{m \in \mathbb{N}} P(\tau_n^{y} + 1 = m, X_{m} = x) = f_x(\bar{y}) P(X_{\tau_n^{y+1} = x}),
\end{align*}
\]

where the third equality follows by the conditional independence of the observations. To show that \((V_{y,n})n\) is a HMM, it is enough to check that the observations are conditionally independent given the Markov chain \((W_y^n)\), i.e.
\[
P(V_{y,1} = y_1, \ldots, V_{y,N} = y_N \mid W_1^{y} = x_1, \ldots, W_N^{y} = x_N) \\
= \prod_{n=1}^{N} P(V_{y,n} = y_n \mid W_n^{y} = x_n),
\]
as it follows from the direct computation,

\[
\mathbb{P}(V_{y,1} = y_1, \ldots, V_{y,N} = y_N \mid W_1^y = x_1, \ldots, W_N^y = x_N) \\
= \mathbb{P}(Y_{y1+1} = y_1, \ldots, Y_{yN+1} = y_N \mid X_{y1+1} = x_1, \ldots, X_{yN+1} = x_N) \\
= \prod_{n=1}^{N} \mathbb{P}(Y_{yn+1} = y_n \mid X_{yn+1} = x_n) = \prod_{n=1}^{N} \mathbb{P}(V_{yn} = y_n \mid W_n^y = x_n),
\]

where the second equality is a direct consequence of Remark 5 of the Appendix. The sequence \((V_{y,n})_n\) is therefore a HMM with recurrent underlying Markov chain, and this concludes the proof of the proposition.

An easier proof (see [10]) of the above proposition holds true if HMMs are equivalently defined as deterministic functions of Markov chains. The need for a more general proof arises since for Polish valued sequences, which are used in Section 4 of the paper, the latter definition of HMM does not make sense.

### 3.2 Representation of countable Markov mixtures

The paper [3] contains a characterization of countable mixtures of i.i.d. sequences, linking HMMs to the class of exchangeable sequences. The main result of [3] can be rephrased as follows

**Theorem 1.** (Dharmadhikari) Let \((Y_n)\) be an exchangeable sequence. \((Y_n)\) is a countable mixture of i.i.d. sequences if and only if \((Y_n)\) is a HMM with recurrent underlying Markov chain.

Note that in the original formulation of Theorem 1 the stationarity of the underlying Markov chain is one of the hypotheses, but close inspection of the proof in [3] reveals that only the absence of transient states is required. The aim of this section is to extend the above theorem to partially exchangeable sequences, i.e. to characterize countable mixtures of Markov chains. The analog of Theorem 1 for mixtures of Markov chains is as follows.

**Theorem 2.** Let \((Y_n)\) be a partially exchangeable sequence satisfying Condition 1. \((Y_n)\) is a countable mixture of Markov chains if and only if \((Y_n)\) is a HMM with recurrent underlying Markov chain.
Proof. We first prove that if \((Y_n)\) is a HMM, then it is a countable mixtures of Markov chains, i.e., in the notations of Remark 1 \(\tilde{P}\) takes countably many values. By the partial exchangeability of \((Y_n)\), the sequence \((V_{y,n})\) is exchangeable for any \(y \in S\), and thus a mixture of i.i.d. sequences. As proved \textit{e.g.} in Lemma 2.15 of [1] or in Proposition 1.1.4 of [9],
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{V_{y,n}}(\cdot) = \tilde{P}_y(\cdot) \quad \mathbb{P} - a.s., \tag{14}
\]
where the limit has to be interpreted in the weak convergence topology, and where \(\tilde{P}_y\) is the random variable with values in \(\mathcal{M}_S\) corresponding to \(\tilde{P}\) in Definition 1. It follows from the proof of Theorem 1 in [8], that the random measure \(\tilde{P}_y\) is the \(y\)-th row of the random element \(\tilde{P}\). By Proposition 1 each row \((V_{y,n})\) is a HMM, and by Theorem 1 it is thus a countable mixture of i.i.d. sequences. Thus \(\tilde{P}_y\) takes countably many values, and so does the \(y\)-th row of \(\tilde{P}\). Repeating the same reasoning for each \(y\), we conclude that \(\tilde{P}\) takes countably many values.

We now prove the converse: if \((Y_n)\) is a countable mixture of Markov chains, i.e. \(\tilde{P}\) takes countably many values, then \((Y_n)\) is a HMM. Let us indicate with \(\{P^h\}_{h \in H}\), where \(H\) is a countable set, the possible values of \(\tilde{P}\), and let the finite distributions of \((Y_n)\) be as in (7). We now construct a Markov chain \((X_n)\), and a sequence \((\tilde{Y}_n)\) according to Definition 5. Proving that \((\tilde{Y}_n)\) has the same distributions of \((Y_n)\), we get that \((Y_n)\) is a HMM with underlying Markov chain \((X_n)\). Let us start with the Markov chain. Let \(P\) be the direct sum of the transition matrices \(P^h\):
\[
P := \begin{pmatrix}
P^1 & \emptyset & \emptyset & \ldots \\
\emptyset & P^2 & \emptyset & \ldots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
\]
Let \((X_n)\) be the Markov chain with state space \(\mathbb{S} \times H\), transition matrix \(P\), and initial distribution \(\pi\), with
\[
\pi(y, h) = \begin{cases}
\mu_h & \text{for } y = y_0 \\
0 & \text{for } y \neq y_0,
\end{cases}
\]
with \(y \in S\) and \(h \in H\), and \(\mu_h := \mathbb{P}(\tilde{P} = P^h)\). We have to show that \((X_n)\) is recurrent. By Theorem 1 in [8], \((Y_n)\) is conditionally recurrent, therefore\footnote{\textit{e.g.} ordering the states in first lexical order as follows: \((1, 1), (2, 1), \ldots, (1, 2), (2, 2), \ldots\)}
the matrices \( \{P^h\} \) in the mixture correspond to recurrent chains. Since \( P \) is the direct sum of such matrices, \((X_n)\) is recurrent.

Consider now a sequence \((\tilde{Y}_n)\), jointly distributed with \((X_n)\), with fixed initial state \(\tilde{Y}_0 = y_0\), conditionally independent given \((X_n)\), and with read-out distributions defined as follows

\[
f_{(x,h)}(y) := \mathbb{P}(\tilde{Y}_n = y \mid X_n = (x,h)) = \delta_{x,y},
\]

where \(\delta\) is the Kronecker symbol. By its very definition \((\tilde{Y}_n)\) satisfies the properties in Definition 5. Let us compute the finite distributions of \((\tilde{Y}_n)\), for any \(y_1^N \in (S^*)^N\), (recall \(\tilde{Y}_0\) is fixed at \(y_0\)),

\[
\mathbb{P}(\tilde{Y}_1^N = y_1^N) = \sum_{(x_0^N,h_0^N)} \mathbb{P}(\tilde{Y}_0^N = y_0^N, X_0^N = (x_0^N, h_0^N))
\]

\[
= \sum_{(x_0^N,h_0^N)} \mathbb{P}(X_0 = (x_0, h_0)) \prod_{n=0}^{N} \mathbb{P}(\tilde{Y}_n = y_n \mid X_n = (x_n, h_n)) \times \prod_{n=0}^{N-1} \mathbb{P}(X_{n+1} = (x_{n+1}, h_{n+1}) \mid X_n = (x_n, h_n))
\]

\[
= \sum_{(x_0^N,h_0^N)} \pi(x_0, h_0) \prod_{n=0}^{N} f_{(x_n,h_n)}(y_n) \prod_{n=0}^{N-1} P_{(x_n,h_n)(x_{n+1},h_{n+1})}
\]

\[
= \sum_{h \in H} \mu_h P^h_{y_0 y_1} \cdots P^h_{y_{N-1} y_N},
\]

where the second equality follows by conditional independence of \((\tilde{Y}_n)\) given \((X_n)\), and the fourth follows by the definition of the read-out densities and by the block structure of \(P\). Comparing (7) with the last expression, we have that \((Y_n)\) and \((\tilde{Y}_n)\) have the same distributions, thus \((Y_n)\) is a HMM and the theorem is proved.

\[\square\]

Note that, for \(S\) finite, in Theorem\[\ref{theo:finite}\] as well as in Theorem\[\ref{theo:finite}\] the state space of the underlying Markov chain is finite if and only if the mixture is finite.
4 Countable Markov mixtures with Polish state space

In this section $S$ is a Polish space.

4.1 The successors array

The proposition below is the analog of Proposition 1 for uncountable state space $S$.

**Proposition 2.** Let $(Y_n)$ be a HMM with recurrent underlying Markov chain, then each row of $(V_{j,n})$ is a HMM with recurrent underlying Markov chain.

**Proof.** Let $(X_n)$ be the underlying Markov chain of $(Y_n)$. Consider the partition $\mathcal{E} = (E_j)_{j \geq 0}$ of $S^*$, and for any element $E_j$ of the partition define

$$
\tau_j^1 := \inf\{t \geq 0 \mid Y_t \in E_j\}, \quad \tau_j^n := \inf\{t > \tau_j^{n-1} \mid Y_t \in E_j\}.
$$

The proof can be carried out exactly as the proof of Proposition 1, substituting $\tau_j^n$ there with $\tau_j^E$, and $\sigma_j^n$ with $\sigma_j^E$, defined below

$$
\sigma_j^{\bar{x},E} := \inf\{t \geq 0 \mid X_t = \bar{x}, Y_t \in E_j\},
$$

$$
\sigma_j^{\bar{x},E} := \inf\{t > \sigma_j^{\bar{x},y} \mid X_t = \bar{x}, Y_t \in E_j\},
$$

where $\bar{x} \in \mathcal{X}$ is such that $f_{\bar{x}}(X_j) > 0$ and $\mathbb{P}(X_{n+1} = x \mid X_n = \bar{x}) > 0$, $(x$ has the same role as in equation (9)). □

4.2 Representation of countable mixtures

The main result of the section is the characterization of countable mixtures of Markov chains, given in Theorem 4 below. The result for Markov chains is based on the corresponding characterization of countable mixtures of i.i.d. sequences which, for discrete state spaces, was derived in [3] and is reported in this paper as Theorem 1. To the best of our knowledge the analogue of Theorem 1 for general state space is not available in the literature. As a preliminary result we thus give, in Theorem 3 below, the needed extension of Theorem 1. Note that the result of [3] can not be directly generalized to uncountable state spaces as it relies on a definition of HMM unsuitable for general spaces.
4.2.1 Representation of countable i.i.d. mixtures

**Theorem 3.** Let \((Y_n)\) be an exchangeable sequence. \((Y_n)\) is a countable mixture of i.i.d. sequences if and only if \((Y_n)\) is a HMM with recurrent underlying Markov chain.

**Proof.** By Remark 2 in the Appendix, if \((Y_n)\) is a countable mixture of i.i.d., then it is a HMM. Let us prove the converse. Let \((Y_n)\) be an exchangeable HMM, with underlying Markov chain \((X_n)\) with initial distribution \(\pi\) and transition matrix \(P\). \((X_n)\) is recurrent, thus it has no transient states, but possibly more than one recurrence class. As noted in [3], by the exchangeability of \((Y_n)\), one can substitute \(P\) with the Cesàro limit \(P^* := \lim_{n \to \infty} 1/n \sum_{k=1}^{n} P^k\), where \(P^k\) is the \(k\)-power of \(P\). By the ergodic theorem \(P^*\) has a block structure, being the direct sum of matrices \(P_h\) with identical rows, one block \(P_h\) for each recurrence class. By Lemma 6 of the Appendix, \((Y_n)\) is a countable mixture of i.i.d. sequences. \(\square\)

4.2.2 Representation of countable Markov mixtures

**Theorem 4.** Let \((Y_n)\) be a partially exchangeable sequence satisfying Condition 1. \((Y_n)\) is a countable mixtures of Markov chains if and only if \((Y_n)\) is a HMM with recurrent underlying Markov chain.

**Proof.** Refer to the notations of Definition 3. If \((Y_n)\) is a HMM with recurrent Markov chain, the proof can be carried out as for Theorem 2 noting that \(1/N \sum_{n=1}^{N} I_{Y_n \in S_j}\) converges, for \(N \to \infty\), to a probability measure \(\theta_j\) on \(S\), which plays the role of \(\tilde{t}(j, \cdot)\), i.e. \(\theta_j(\cdot) = \tilde{t}(j, \cdot)\). We also need to use Proposition 2 instead of Proposition 1 and Theorem 3 instead of Theorem 1.

We now prove the converse. Let \(\tilde{t}\) takes countably many values, denoted with \((t_h)_{h \in H^*}\) with \(h\) varying on a countable set \(H\), and let

\[
\mu_h = \mathbb{P}(\tilde{t} = t_h).
\]  

(15)

The finite distributions of \((Y_n)\) are as in equation (5). We will construct a sequence \((\tilde{Y}_n)\) with the same distributions of \((Y_n)\), so that \((Y_n)\) is a HMM. The construction that worked for the proof of Theorem 2 cannot be used here. \((Y_n)\) takes values in an uncountable state space while we need an HMM with a discrete underlying Markov chain. Consider thus a Markov chain \((X_n)\) taking
values in $H \times \mathbb{N}_0 \times \mathbb{N}_0$, with initial distribution and transition probabilities as follows (by Condition 1, the initial value $y_0 \in E_1$

$$P(X_0 = (h, i, j)) = \begin{cases} \mu_h \delta_{j,1} t_{h,i}(i,dy_0) & \text{for } t_h(i, dy_0) \neq 0 \\ 0 & \text{for } t_h(i, dy_0) = 0. \end{cases}$$

$$P\left(X_n = (h_n, i_n, j_n) \mid X_{n-1} = (h_{n-1}, i_{n-1}, j_{n-1})\right) = \delta_{h_{n-1},h_n} \delta_{i_{n-1},i_n} \delta_{j_{n-1},j_n} t_{h_n}(j_{n-1}, E_{j_n}).$$

The components $(h_n, i_n, j_n)$ of $X_n$ represent respectively: the index of the running chain in the mixture, the discretized value of $Y_{n-1}$, and the discretized value of $Y_n$. The Markov chain $(X_n)$ is recurrent: the kernels $t_h(\cdot, \cdot)$ correspond to recurrent Markov chains by Theorem 4 in [8]. Consider now a sequence $(\tilde{Y}_n)$ jointly distributed with $(X_n)$, with fixed initial value $\tilde{Y}_0 = y_0$, conditionally independent given $(X_n)$, and with read-out distributions defined as follows

$$P(\tilde{Y}_n \in dy_n \mid X_n = (h, i, j)) = \begin{cases} \mathbb{I}_{E_j}(y_n) t_{h,i}(i,dy_0) & \text{for } t_h(i, E_j) \neq 0 \\ 0 & \text{for } t_h(i, E_j) = 0. \end{cases}$$

Note that, by its very definition, $(\tilde{Y}_n)$ satisfies the properties in Definition [5].
The distributions of \((\tilde{Y}_n)\) are computed as follows

\[
P(\tilde{Y}_1 \in dy_1, \ldots, \tilde{Y}_N \in dy_N) \\
= \sum_{h_0, i_0, j_0} \mathbb{P}(\tilde{Y}_0 \in dy_0, \ldots, \tilde{Y}_N \in dy_N, X_0 = (h_0, i_0, j_0), \ldots, X_N = (h_N, i_N, j_N)) \\
= \sum_{h_0, i_0, j_0} \mathbb{P}(\tilde{Y}_0 \in dy_0, \ldots, \tilde{Y}_N \in dy_N | X_0 = (h_0, i_0, j_0), \ldots, X_N = (h_N, i_N, j_N)) \\
\times \mathbb{P}(X_0 = (h_0, i_0, j_0), \ldots, X_N = (h_N, i_N, j_N)) \\
= \sum_{h_0, i_0, j_0} \mathbb{P}(X_0 = (h_0, i_0, j_0)) \prod_{n=0}^{N} \mathbb{P}(\tilde{Y}_n \in dy_n | X_n = (h_n, i_n, j_n)) \\
\times \prod_{n=1}^{N} \mathbb{P}(X_n = (h_n, i_n, j_n) | X_{n-1} = (h_{n-1}, i_{n-1}, j_{n-1})) \\
= \sum_{h_0, i_0, j_0} \mu_{h_0} \delta_{j_0,1} t_{h_0}(i_0, E_{j_0}) \prod_{n=0}^{N} \mathbb{I}_{E_{j_n}}(y_n) t_{h_n}(i_n, dy_n) \\
\times \prod_{n=1}^{N} \delta_{h_{n-1}, h_n} \delta_{i_n, j_{n-1}} t_{h_n}(j_{n-1}, E_{j_n}) \\
= \sum_{h \in H, i_0} \mu_{h} t_{h}(i_0, E_1) \prod_{n=1}^{N} t_{h}(l_{n-1}, dy_n) \\
= \sum_{h \in H} \mu_{h} t_{h}(1, dy_1) \ldots t_{h}(l_{N-1}, dy_n),
\]

where \(l_n\) indicates the element of the partition to which \(y_n\) belongs, and recalling \(l_0 = 1\). Integrating the expression above over \(A_1, \ldots, A_N\), and comparing it with equation (5), one concludes that the distributions of \((\tilde{Y}_n)\) coincide with those of \((Y_n)\), therefore proving that \((Y_n)\) is a HMM.

5 Concluding remarks

Throughout the paper we referred to the notion of partial exchangeability originally given by de Finetti and to the corresponding representation theorem as given in [8]. For discrete state space partial exchangeability can be defined in a slightly different way, and a representation theorem in this
alternative framework is proved in [4]. According to [4], a sequence of random variables is partially exchangeable if the probability is invariant under all permutations of a string that preserves the first value and the transition counts between any couple of states. A characterization of countable mixtures of Markov chains can be given also in the setup of [4], using different mathematical tools. The result is in [7], but for a complete proof see [10]. By the same token the characterization of countable mixtures of Markov chains of order $k$ holds true, for the proof see [10]. Unfortunately the approach of [7] and [10] does not readily generalize to Polish state space.

Based on the results in [8], a de Finetti’s type representation theorem for mixtures of semi-Markov processes have been proved in [6]. The authors are confident that a characterization of countable mixtures of semi-Markov processes in terms of HMMs can be given properly adapting the proof of Proposition 2 and Theorem 4.

6 Appendix

6.1 Hidden Markov models

Let $S$ be a Polish space endowed with the Borel $\sigma$-field $S$.

Definition 5. Let $(\tilde{Y}_n)_{n \geq 0}$ be a random sequence with values in $(S, \mathcal{S})$ and such that, for some random sequence $(X_n)_{n \geq 0}$ on a discrete space $\mathcal{X}$, it holds that

- $(X_n)$ is a homogeneous Markov chain,
- $(\tilde{Y}_N)$ is conditionally independent given $(X_n)$, i.e.

$$
P(\tilde{Y}_0 \in dy_0, \ldots, \tilde{Y}_N \in dy_N \mid X_0^N = x_0^N) = \prod_{n=0}^{N} P(\tilde{Y}_n \in dy_n \mid X_n = x_n).$$

A Hidden Markov Model (HMM) is any random sequence $(Y_n)$ sharing finite distributions with $(\tilde{Y}_n)$, i.e. such that

$$P(Y_0 \in y_0, \ldots Y_N \in dy_N) = P(\tilde{Y}_0 \in dy_0, \ldots \tilde{Y}_N \in dy_N)$$

for all $N$. 

16
A HMM is characterized by the initial distribution $\pi$ on $\mathcal{X}$, by the transition matrix $P = (P_{ij})_{i,j \in \mathcal{X}}$ of the Markov chain $(X_n)$, and by the read-out distributions $f_x(dy)$ where

$$f_x(dy) := P(\tilde{Y}_n \in dy \mid X_n = x).$$

We refer to the sequence $(X_n)$ as the ”underlying Markov chain” of the HMM, and to $(Y_n)$ as the output sequence.

Remark 2. A countable mixture of i.i.d. sequences as in equation (2) is trivially the output $(Y_n)$ of an HMM: take as $(X_n)$ the Markov chain with values in $H$, with identity transition matrix, initial distribution $(\mu_1, \ldots, \mu_h, \ldots)_{h \in H}$, and read-out distributions $P(Y_n \in dy \mid X_n = h) = p_h(dy)$.

6.1.1 Strong Markov and strong conditional independence for HMMs

This section contains a few useful properties of HMMs.

Lemma 1. (Splitting property) Let $(Y_n)$ be a HMM with underlying Markov chain $(X_n)$. Then the couple $(X_n, Y_n)$ is a Markov chain, moreover

$$P(X_N = x, Y_N \in dy \mid X_N^{N-1} = x_N^{N-1}, Y_N^{N-1} \in dy_N^{N-1}) = P(X_N = x, Y_N \in dy \mid X_{N-1} = x_{N-1}).$$

Proof.

$$P(X_N = x, Y_N \in dy_N \mid X_N^{N-1} = x_N^{N-1}, Y_N^{N-1} \in dy_N^{N-1})$$

$$= \frac{P(X_N = x_N, Y_N \in dy_N)}{P(X_N^{N-1} = x_N^{N-1}, Y_N^{N-1} \in dy_N^{N-1})}$$

$$= \frac{\prod_{n=1}^N P(Y_n \in dy_n \mid X_n = x_n) P(X_N = x_N)}{\prod_{n=1}^{N-1} P(Y_n \in dy_n \mid X_n = x_n) P(X_N^{N-1} = x_N^{N-1})}$$

$$= P(Y_N \in dy_N \mid X_N = x_N) P(X_N = x_N \mid X_{N-1} = x_{N-1})$$

$$= P(Y_N \in dy_N \mid X_N = x_N, X_{N-1} = x_{N-1}) P(X_N = x_N \mid X_{N-1} = x_{N-1})$$

$$= P(Y_N \in dy_N, X_N = x_N \mid X_{N-1} = x_{N-1}),$$

17
where the second and fourth equality follow by the conditional independence of the observations \((Y_n)\) in the definition of HMM.

**Lemma 2.** (Strong splitting property) Let \((Y_n)\) be a HMM with underlying Markov chain \((X_n)\), and \(\gamma\) be a stopping time for \((X_n, Y_n)\), then, for any \(x, \tilde{x}, \bar{x} \in \mathcal{X}\), and any \(y, \tilde{y}, \bar{y} \in S\) it holds that

\[
P(X_{\gamma+k} = x, Y_{\gamma+k} \in dy \mid X_{\gamma} = \tilde{x}, Y_{\gamma} \in d\tilde{y}, X_{\gamma\wedge n} = \bar{x}, Y_{\gamma\wedge n} \in d\bar{y}) = P(X_{\gamma+k} = x, Y_{\gamma+k} \in dy \mid X_{\gamma} = \tilde{x}).
\]

(16)

**Proof.** We manipulate separately the LHS and the RHS of equation (16). For readability denote 
\(C_r := \left( \gamma = r, X_r = \tilde{x}, Y_r \in d\tilde{y}, X_{r \wedge n} = \bar{x}, Y_{r \wedge n} \in d\bar{y} \right)\).

Applying Lemma 1, the numerator of the conditional probability on the LHS of equation (16) is

\[
P(X_{\gamma+k} = x, Y_{\gamma+k} \in dy \mid X_{\gamma} = \tilde{x}, Y_{\gamma} \in d\tilde{y}, X_{\gamma\wedge n} = \bar{x}, Y_{\gamma\wedge n} \in d\bar{y})
= \sum_{r \geq 1} P(X_{r+k} = x, Y_{r+k} \in dy \mid C_r) P(C_r)
= \sum_{r \geq 1} P(X_{r+k} = x, Y_{r+k} \in dy \mid X_r = \tilde{x}) P(C_r)
= \sum_{r \geq 1} P(Y_{r+k} \in dy \mid X_{r+k} = x) P(X_{r+k} = x, \mid X_r = \tilde{x}) P(C_r)
= f_x(y) P^{(k)}_{\tilde{x}, x} \sum_{r \geq 1} P(C_r)
= f_x(y) P^{(k)}_{\tilde{x}, x} P(X_\gamma = \tilde{x}, Y_\gamma \in d\tilde{y}, X_{\gamma\wedge n} = \bar{x}, Y_{\gamma\wedge n} \in d\bar{y}),
\]

where \(P^{(k)}_{\tilde{x}, x}\) is the \(\tilde{x}, x\)-entry of the \(k\)-step transition matrix of the Markov chain \((X_n)\). The numerator of the conditional probability on the RHS of equation (16), again applying Lemma 1, is
\[ P(X_{\gamma+k} = x, Y_{\gamma+k} \in dy, X_\gamma = \tilde{x}) = \sum_{r \geq 1} P(X_{r+k} = x, Y_{r+k} \in dy \mid \gamma = r, X_r = \tilde{x}) P(\gamma = r, X_r = \tilde{x}) = \sum_{r \geq 1} P(X_{r+k} = x, Y_{r+k} \in dy \mid X_r = \tilde{x}) P(\gamma = r, X_r = \tilde{x}) = f_x(y) P^k_{\tilde{x}, \tilde{x}} P(X_\gamma = \tilde{x}). \]

The lemma is proved comparing the expressions of the LHS and the RHS derived above. \(\square\)

**Definition 6.** Let \((Y_n)\) be a HMM with underlying Markov chain \((X_n)\), and let \(A \subset \mathcal{X} \times \mathcal{S}\). We say that the sequence of random times \((\gamma_n)_{n \geq 1}\) is a sequence of hitting times of \(A\) if

\[
\gamma_1 := \inf\{ t \geq 0 \mid (X_t, Y_t) \in A \}, \\
\gamma_n := \inf\{ t > \gamma_{n-1} \mid (X_t, Y_t) \in A \}.
\]

**Lemma 3.** (Generalized strong splitting property) Let \((Y_n)\) be a HMM with underlying Markov chain \((X_n)\). Let \((\gamma_n)\) be a sequence of hitting times of \(A\) for \((X_n, Y_n)\), where \(A \subset \mathcal{X} \times \mathcal{S}\). Then for any \(N\), and any \((x_1, y_1), \ldots, (x_N, y_N) \in A\)

\[
P(X_{\gamma_N} = x_N, Y_{\gamma_N} \in dy_N \mid X_{\gamma_{N-1}} = x_1^{N-1}, Y_{\gamma_{N-1}} \in dy_1^{N-1}) = P(X_{\gamma_N} = x_N, Y_{\gamma_N} \in dy_N \mid X_{\gamma_{N-1}} = x_{N-1}).
\]

**Proof.** Denote with \(A^c\) the complement of \(A\) in \(\mathcal{X} \times \mathcal{S}\), and with \((A^c)^r\) the \(r\)-th fold Cartesian product of \(A^c\). Let \(B := (X_{\gamma_{N-1}} = x_1^{N-1}, Y_{\gamma_{N-1}} \in dy_1^{N-1})\). Applying Lemma 2 in the third equality above, the numerator of the conditional probability on the LHS is
\[ P(X_{\gamma_1}^N = x_1^N, Y_{\gamma_1}^N \in dy_1^N) \]
\[ = \sum_{r \geq 1} P(\gamma_N = \gamma_{N-1} + r, X_{\gamma_{N-1}+r} = x_N, Y_{\gamma_{N-1}+r} \in dy_N \mid B) P(B) \]
\[ = \sum_{r \geq 1} \int_{(A_r)^{r-1}} P(X_{\gamma_{N-1}+r} = x_N, Y_{\gamma_{N-1}+r} \in dy_N, \]
\[ X_{\gamma_{N-1}+1}^{\gamma_{N-1}+r-1} = x_1^{r-1}, Y_{\gamma_{N-1}+1}^{\gamma_{N-1}+r-1} \in dy_1^{r-1} \mid B) P(B) d(\bar{x}_1^{r-1}, \bar{y}_1^{r-1}) \]
\[ = \sum_{r \geq 1} \int_{(A_r)^{r-1}} P(X_{\gamma_{N-1}+r} = x_N, Y_{\gamma_{N-1}+r} \in dy_N, \]
\[ X_{\gamma_{N-1}+1}^{\gamma_{N-1}+r-1} = x_1^{r-1}, Y_{\gamma_{N-1}+1}^{\gamma_{N-1}+r-1} \in dy_1^{r-1} \mid X_{\gamma_{N-1}} = x_{N-1}) P(B) \]
\[ = P(X_{\gamma_{N}} = x_N, Y_{\gamma_{N}} \in dy_N \mid X_{\gamma_{N-1}} = x_{N-1}) P(B), \]

and dividing by \( P(B) \) the lemma is proved. \( \square \)

**Remark 3.** By the same token, for any \((x_1, y_1), \ldots, (x_N, y_N) \in X \times S, \)
\[ P(X_{\gamma_{N+1}} = x_N, Y_{\gamma_{N+1}} \in dy_N \mid X_{\gamma_{N+1}}^{\gamma_{N-1}+1} = x_1^{N-1}, Y_{\gamma_{N+1}}^{\gamma_{N-1}+1} \in dy_1^{N-1}) \]
\[ = P(X_{\gamma_{N+1}} = x_N, Y_{\gamma_{N+1}} \in dy_N \mid X_{\gamma_{N-1}+1} = x_{N-1}). \]

**Lemma 4.** Let \((Y_n)\) be a HMM with underlying Markov chain \((X_n)\), and \((\gamma_n)\) be a sequence of hitting times of \(A\) for \((X_n, Y_n)\), where \(A = A_1 \times A_2 \subset X \times S\), then, for any \(x_1, \ldots, x_N \in A_1, \)
\[ P(X_{\gamma_{N}} = x_N \mid X_{\gamma_{N}-1}^{\gamma_{N}-1} = x_1^{N-1}) = P(X_{\gamma_{N}} = x_N \mid X_{\gamma_{N-1}} = x_{N-1}). \]

**Proof.** For readability let \( D := (X_{\gamma_{N}}^{\gamma_{N}-1}, Y_{\gamma_{N}-1}^{\gamma_{N}-1} \in dy_{1}^{N-1}) \). The numerator of the LHS is
\[ P(X_{\gamma_{1}}^{N} = x_1^{N}) = \int_{(A_2)^N} P(X_{\gamma_{N}} = x_N, Y_{\gamma_{N}} \in dy_N \mid D) P(D) \]
\[ = \int_{(A_2)^N} P(X_{\gamma_{N}} = x_N, Y_{\gamma_{N}} \in dy_N \mid X_{\gamma_{N-1}} = x_{N-1}) P(D) \]
\[ = P(X_{\gamma_{N}} = x_N \mid X_{\gamma_{N-1}} = x_{N-1}) P(X_{\gamma_{N}}^{\gamma_{N}-1} = x_1^{N-1}). \]
\( \square \)
Remark 4. By the same token, and using Remark 3 for any $x_1, \ldots, x_N \in \mathcal{X}$,
\[
P(X_{\gamma N+1} = x_N \mid X_{\gamma N-1}^N = x_{N-1}^N) = P(X_{\gamma N+1} = x_N \mid X_{\gamma N-1}^N = x_{N-1}^N).
\]

Lemma 5. (Strong conditional independence) Let $(Y_n)$ be a HMM with underlying Markov chain $(X_n)$, and $(\gamma_n)$ be a sequence of hitting times of $A$ for $(X_n,Y_n)$, where $A \subset \mathcal{X} \times \mathcal{S}$, then, for any $(x_1, y_1), \ldots, (x_N, y_N) \in A$,
\[
P(Y_{\gamma_n}^N \in dy_1^N \mid X_{\gamma_n}^N = x_1^N) = \prod_{k=1}^N P(Y_{\gamma_k} \in dy_k \mid X_{\gamma_k} = x_k).
\]

Proof. Let $E := (Y_{\gamma_n}^N \in dy_1^{N-1}, X_{\gamma_n}^{N-1} = x_{1}^{N-1})$ for readability.
\[
P(Y_{\gamma_n}^{N} \in dy_{1}^{N} 
\quad \mid X_{\gamma_n}^{N} = x_{1}^{N})
\quad = \prod_{k=1}^{N} P(Y_{\gamma_k} \in dy_k \mid X_{\gamma_k} = x_k)
\]
where the second equality follows by Lemma 3 and the last equality follows iterating the procedure. \hfill \Box

Remark 5. By the same token, using Remark 3 for any $x_1^N, y_1^N \in \mathcal{X} \times \mathcal{S}$,
\[
P(Y_{\gamma_n}^{N+1} \in dy_1^{N} \mid X_{\gamma_n}^{N+1} = x_{1}^{N}) = \prod_{k=1}^{N} P(Y_{\gamma_k+1} \in dy_k \mid X_{\gamma_k+1} = x_k).
\]

### 6.1.2 HMMs and countable mixtures of i.i.d. sequences

The following fact was used in the proof of Theorem 3. If the HMM $(Y_n)$ has an underlying Markov chain with block structured transition probability matrix, with identical rows within blocks, then $(Y_n)$ is a countable mixture of i.i.d. sequences.

Consider a Markov chain $(X_n)$ with values in $\mathcal{X}$ and transition matrix $P$ as follows
\[
P := \begin{pmatrix}
P_1 & 0 & \cdots & \cdots & 0 & 0 
0 & P_2 & 0 & \cdots & 0 & 0 
\vdots & \vdots & \ddots & \cdots & \vdots & \vdots 
0 & \cdots & \cdots & 0 & P_h & 0 
0 & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix},
P_h := \begin{pmatrix}
p_{11}^h & p_{12}^h & \cdots & p_{1h}^h 
p_{21}^h & p_{22}^h & \cdots & p_{2h}^h 
\vdots & \vdots & \ddots & \vdots 
p_{c1}^h & p_{c2}^h & \cdots & p_{ch}^h 
\end{pmatrix}.
\]

(17)
with \( h \in H \), a countable set. The block \( P_h \) has size \( l_h \). Some of the \( p_h \) can be null. The Markov chain \((X_n)\) has clearly \( H \) recurrence classes, one for each block, and no transient states. Let us indicate with \( C_h \) the \( h \)-th recurrence class, corresponding to the states of the \( h \)-th block, set \( C_h = \{c_h^1, \ldots, c_h^l_h\} \), where \( l_h \) can be infinite. Trivially \( \mathcal{X} = \bigcup_{h \in H} C_h \). An invariant distribution associated with the \( h \)-th block is \( \mu_h := (p_{c_h^1}^h, \ldots, p_{c_h^l_h}^h) \), and for any sequence \( \mu_h > 0 \) with \( \sum_{h \in H} \mu_h = 1 \), the vector

\[
\pi = (\mu_1 p^1, \ldots, \mu_h p^h, \ldots)
\]

is an invariant distribution for \( P \).

**Lemma 6.** Consider a HMM \((Y_n)\) where the underlying Markov chain \((X_n)\) has transition matrix \( P \) as in \((17)\), invariant measure \( \pi \) as in \((18)\), and assigned read-out distributions \( f_x(dy) \), then \((Y_n)\) is a countable mixtures of i.i.d. sequences where \( \tilde{p} \) takes values in the set \( \{P_h, \ h \in H\} \), with

\[
F_h(dy) := p_{c_h^1}^h f_{c_h^1}(dy) + \cdots + p_{c_h^l_h}^h f_{c_h^l_h}(dy),
\]

and \( \mathbb{P}(\tilde{p} = F_h) = \mu_h \).

**Proof.** Let us compute the finite distributions of \((Y_n)\). Let \( y_0, \ldots, y_n \in S^{n+1} \):

\[
\mathbb{P}(Y_0^N \in dy_0^N) = \sum_{x_0^N \in \mathcal{X}} \mathbb{P}(Y_0^N \in dy_0^N, X_0^N = x_0^N)
\]

\[
= \sum_{x_0^N \in \mathcal{X}} P(X_0 = x_0) \prod_{n=0}^{N} \mathbb{P}(Y_n \in dy_n \mid X_n = x_n) \prod_{n=1}^{N} \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1})
\]

\[
= \sum_{x_0^N \in \mathcal{X}} \pi_{x_0} \prod_{n=0}^{N} f_{x_n}(dy_n) \prod_{n=1}^{N} P_{x_{n-1}, x_n}
\]

\[
= \sum_{x_0^N \in \mathcal{X}} \pi_{x_0} f_{x_0}(dy_0) P_{x_0, x_1} f_{x_1}(dy_1) \cdots P_{x_{N-1}, x_N} f_{x_N}(dy_N)
\]

\[
= \sum_{h \in H} \sum_{x_0^N \in C_h} \mu_h \tilde{p}_h f_{x_0}(dy_0) p_{x_1}^h f_{x_1}(dy_1) \cdots p_{x_N}^h f_{x_N}(dy_N)
\]

\[
= \sum_{h \in H} \mu_h \tilde{F}_h(dy_0) F_h(dy_1) \cdots F_h(dy_N).
\]
where the second equality follows by the HMM properties, the fifth equality follows noting that $P_{x_n, x_{n+1}}$ is null for $x_n$ and $x_{n+1}$ in different recurrence classes, and it is equal to $p_{x_n, x_{n+1}}^h$ for $x_n$ and $x_{n+1}$ in the same recurrence class $C_h$. The expression above coincides with the representation of countable mixtures of i.i.d. sequences given in [2], thus completing the proof.

References

[1] D.J. Aldous, (1985), Exchangeability and Related Topics, in Ecole d’Été de Probabilités de Saint-Flour XIII, LNM-1117, Springer, Berlin, pp. 1-198

[2] B. de Finetti, (1938), Sur la condition d’equivalence partielle in Actualité Scientifiques et Industrielles, Hermann, Paris, vol. 739, pp. 5-18

[3] S.W. Dharmadhikari, (1964), Exchangeable Processes which are Functions of Stationary Markov Chains, in The Annals of Mathematical Statistics, vol. 35, pp. 429-430

[4] P. Diaconis and D. Freedman, (1980), de Finetti’s Theorem for Markov Chains in Annals of Probability, vol. 8, pp. 115-130

[5] P. Diaconis and D. Freedman, (2004), The Markov Moment Problem and de Finetti’s Theorem, Part I and Part II in Math. Z., vol. 247, pp. 183-212

[6] I. Epifani, and S. Fortini, and L. Ladelli, (2002), A characterization for mixtures of semi-Markov processes in Statist. and Prob. Letters , vol.60, pp. 445-457

[7] L. Finesso and C. Prosdocimi, (2009), Partially Exchangeable Hidden Markov Models, in Proceedings of the European Control Conference, ECC’09, Budapest, Hungary, pp. 3910-3914

[8] S. Fortini, L. Ladelli, G. Petris, and E. Regazzini, (2002), On mixtures of distributions of Markov chains in Stoch. Proc. Appl., vol. 100, pp. 147-165

[9] O. Kallenberg, (2005) Probabilistic Symmetries and Invariance Principles, Springer-Verlag
[10] C. Prosdocimi, (2010)  *Partial Exchangeability and Change Detection for Hidden Markov Models*, PhD thesis, cycle XXII, University of Padova