ON MIXTURES OF DISTRIBUTIONS OF MARKOV CHAINS[1]

by

Sandra Fortini[a], Lucia Ladelli[b], Giovanni Petris[c]
and
Eugenio Regazzini[d*]

[a] *IMQ-Università "L. Bocconi", Milano, Italy.*
[b] *Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Italy.*
[c] *Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR, USA.*
[d] *Dipartimento di Matematica "F. Casorati", Università degli Studi di Pavia, Italy.*

Abstract

Let X be a chain with discrete state space S, and V be the matrix with entries $V_{i,n}$, where $V_{i,n}$ denotes the position of the process immediately after the nth visit to i. We prove that the law of X is a mixture of laws of Markov chains if and only if the distribution of V is invariant under finite permutations within rows (i.e., the $V_{i,n}$s are partially exchangeable in the sense of de Finetti). We also prove that an analogous statement holds true for mixtures of laws of Markov chains with atomic kernels. Going back to the discrete case, we analyze the relationships between partial exchangeability of V and Markov exchangeability in the sense of Diaconis and Freedman. The main statement is that the former is stronger than the latter, but the two are equivalent under the assumption of recurrence. Combination of this equivalence with the aforesaid representation theorem gives the Diaconis and Freedman basic result for mixtures of Markov chains.

MSC: Primary: 60G05. Secondary: 60J05, 62A15.

Keywords: Markov exchangeability (in Freedman's sense); Matrix of successor states; Mixtures of laws of Markov chains; Partial exchangeability (in de Finetti's sense).

[1]Research partially supported by MURST, Programma di Ricerca: Processi stocastici, calcolo stocastico e applicazioni. (1999)

* Corresponding author. Dipartimento di Matematica "F. Casorati", via Ferrata, 27100 Pavia, Italy.
E-mail adress: eugenio@dimat.unipv.it

1 Introduction

The problem of characterizing mixtures of distributions of Markov chains has been studied in depth by Freedman (1962a, 1962b, 1996), Diaconis and Freedman (1980a), (1980b) and, in the case of finite sequences, by Zaman (1984, 1986). For chains on a countable state space S, one of the major results attained is that the distribution of a recurrent chain is a mixture of distributions of Markov chains if and only if the same distribution satisfies a Markov invariance condition, i.e. is invariant with respect to all permutations which do not alter the number of transitions from i to j, whatever i and j in S may be. Kallenberg (1982) considers chains on more general state spaces and obtains characterizations for mixtures of distributions of Markov chains in terms of a condition of invariance under stopping times shifts.

The history of characterizations of mixtures of distributions of Markov chains dates back to de Finetti (1959), who formulated the following conjecture. See also Ch. V in de Finetti (1937). Given any S valued process, starting from a distinguished state, assume that all states are to be visited infinitely often. Let $V_{i,n}$ be the position of the process immediately after the nth visit to i. This random matrix is said to be *partially exchangeable* if, and only if, its distribution is invariant under finite permutations within rows. See de Finetti (1938). De Finetti hinted at the possibility of proving that partial exchangeability of the V-matrix is necessary and sufficient for a recurrent process to be represented in law as a mixture of laws of Markov chains. The subsequent characterizations – mentioned above – depart from the original formulation of de Finetti's conjecture, since they skip the relationship between mixture of distributions of Markov chains and partial exchangeability of V. In order to avoid terminological ambiguities, it is worth recalling that Diaconis and Freedman (1980a) use the term partial exchangeability to designate their Markov invariance condition. More recently, the matrix V has been considered in Zabell (1995). This author does focus on a few properties of the $V_{i,n}$s by resorting to the main result of Diaconis and Freedman, but he does not actually linger on the possible connection mentioned by de Finetti. For these reasons, we have thought it fit to take up again the characterization of mixtures of distributions of Markov chains according to de Finetti's original remarks. Thus we have been induced to analyze the relationships between partial exchangeability of V and Markov invariance in the sense of Diaconis and Freedman, on one hand, and, on the other hand, to extend the analysis to general state spaces.

The setup of the paper is as follows. Section 2 deals with the case of a discrete state space. We first provide a proof of de Finetti's conjecture, and then analyze the relationship between Freedman condition and partial exchangeability of the V-matrix. The results contained in this section can also be used to reobtain the main result in Diaconis and Freedman (1980a). In Section 3 we extend the results of Section 2 to any state space, obtaining a complete characterization for mixtures of laws of Markov chains having atomic kernels, i.e. kernels with countable range (in some space of probability measures). Cf., e.g., Nummelin (1984), Section 4.2. Section 3 includes also some hints at the possibility of approximating any mixture of laws of Markov chains with atomic kernels. Section 4 concludes the paper with a few comments focusing on some interpretative statistical issues suggested by the previous results.

2 Mixtures of Laws of Markov Chains With Discrete State Space

2.1 Main Result

Let $(\Omega, \mathcal{F}, P)$ be a probability space, S be a countable set endowed with its power set σ-algebra $\mathcal{S}$, and let $X = (X_n)_{n \geq 0}$ be a discrete time S-valued stochastic process defined on Ω. To any such process one can associate its *V-matrix* $V = (V_{s,j})_{s \in S, j \geq 1}$, where V_{sj} is the value of the process (successor state) immediately after the jth visit to state s. In order to avoid having rows of finite length, we introduce a new symbol $\partial \notin S$ and, if a state s is visited only a finite number of times k, we put $V_{sj} = \partial$ for $j > k$. All the X_n, on the other hand, are S-valued. The main theorem of the present section is that the law of X is a mixture of distributions of recurrent Markov chains if and only if the V-matrix is partially exchangeable. It turns out that to obtain uniqueness of the mixing measure, one has to enlarge the state space S to $S^* = S \cup \{\partial\}$ and consider mixtures of laws of Markov chains on state space S^*. In any case, it is assumed that $P(X \in S^\infty) = 1$. Let T^* be the subset of transition matrices on S^* for which ∂ is an absorbing state, equipped with the Borel σ-algebra which corresponds to the topology of elementwise convergence, and let us write $t(s, x)$ for the (s, x) element of the generic t in T^*. In order to avoid conditioning on the initial state of the process, we assume that X starts in a specific state x_0. In the common treatments of Markov chains on a countable state space, it is customary to drop those states which are not accessible from the initial state. In the present setting, this is no longer feasible, since the different transition matrices in the mixture may have different sets of accessible states. Yet, as far as the distribution of a Markov chain or a mixture of Markov chains is concerned, the transition probabilities from any inaccessible state are totally irrelevant (see also Example 2). Therefore, in order to achieve uniqueness of the mixing measure, one needs to enforce some restrictions on its support. Consider a measurable set $\mathcal{K}$ of T^* such that for every $t \in \mathcal{K}$ there is a set $A_t \subset S^*$ satisfying

KD1. $x_0 \in A_t$, $\partial \notin A_t$;

KD2. $t(x, y) = 0$ if $x \in A_t$ and $y \notin A_t$;

KD3. $t(y, \partial) = 1$ if $y \notin A_t$.

Within the setting of Theorem 1 below, A_t can be viewed as the set of the states that are accessible to a Markov chain starting at x_0 and whose evolution is regulated by the transition matrix t.

Theorem 1. *The V-matrix of X is partially exchangeable if and only if there is a random element $\tilde{t}$ of T^* such that*

(a) $P(X_1 = x_1, \dots, X_n = x_n \mid \tilde{t}\,) = \tilde{t}(x_0, x_1) \dots \tilde{t}(x_{n-1}, x_n)$ *a.s.-P;*

(b) $P(X_n = x_0 \text{ i.o.} \mid \tilde{t}\,) = 1$ *a.s.-P;*

(c) $P(\tilde{t} \in \mathcal{K}) = 1$ *for a measurable set $\mathcal{K}$ satisfying* KD1-KD3.

Moreover, $\tilde{t}$ is unique in distribution.

Roughly speaking, the theorem says that X is a recurrent Markov chain, conditionally on its random transition matrix, if and only if its V-matrix is partially exchangeable. The proof of Theorem 1 can be deduced from Theorem 7 in Diaconis and Freedman (1980a), applying the results of the next section, which states the equivalence between partial exchangeability of V and the assumptions made by Diaconis and Freedman. However, we think that a more elementary and direct proof, as the one given below, may be of some interest. Given any $(x, x_1, \dots, x_n)$ and $(x', x'_1, \dots, x'_n)$ in S^{n+1}, write $(x, x_1, \dots, x_n) \sim (x', x'_1, \dots, x'_n)$ if $x = x'$ and the two vectors contain the same number of transitions from u to v for any u and v in S. Lemma 1 below will be used in the proof of Theorem 1. Its meaning is clarified by the following example.

Example 1. Let $S = \{0, 1\}$ and $(s_0, \ldots, s_{10}) = (0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1)$. Consider the set of all sequences $(x_0, x_1, \ldots) \in S^\infty$ such that $(x_0, \ldots, x_{10}) = (s_0, \ldots, s_{10})$, and observe that a matrix V is consistent with this set if and only if $V_{0,1} = 0$, $V_{0,2} = V_{0,3} = 1$, $V_{0,4} = 0$, $V_{0,5} = V_{0,6} = 1$, $V_{1,1} = V_{1,2} = 0$, $V_{1,3} = 1$, $V_{1,4} = 0$. Interchange $V_{0,4}$ and $V_{0,6}$ and describe the set of all sequences $(0, x_1, \ldots) \in S^\infty$ which are consistent with the new matrix. This set consists of all sequences $(0, 0, 1, 0, 1, 0, 1, 1, 0, 1, x_{10}, x_{11}, \ldots)$, and the longest string $(s'_0, \ldots, s'_9)$ determined in a unique manner by the new matrix belongs to S^{10}: $(s'_0, \ldots, s'_9) \not\sim (s_0, \ldots, s_{10})$. On the other hand, all permutations of any row, which leave $V_{0,6}$ and $V_{1,4}$ unchanged, behave otherwise. For instance, the set of all sequences $(0, x_1, \ldots)$ which are consistent with a matrix V' such that $V'_{0,1} = 1$, $V'_{0,2} = V'_{0,3} = 0$, $V'_{0,4} = V'_{0,5} = V'_{0,6} = 1$, $V'_{1,1} = 0$, $V'_{1,2} = 1$, $V'_{1,3} = V'_{1,4} = 0$ is the set of all sequences satisfying $(0, x_1, \ldots, x_{10}) = (0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1)$, and this vector is apparently equivalent to $(s_0, \ldots, s_{10})$.

Lemma 1. *Let $s_1, \ldots, s_k$ be the distinct elements of $(x_0, \ldots, x_{m-1}) \in S^m$, with $m \geq 2$.*

(a) *There are unique n_i and $v_{s_i,1}, \ldots, v_{s_i,n_i}$ in S $(i = 1, \ldots, k)$ such that*

$$\{X_0 = x_0, \ldots, X_m = x_m\} = \{X_0 = x_0\} \cap \bigcap_{i=1}^{k} \bigcap_{j=1}^{n_i} \{V_{s_i,j} = v_{s_i,j}\}.$$

(b) *For each permutation σ_i of $(1, \ldots, n_i)$ with $\sigma_i(n_i) = n_i$ $(i = 1, \ldots, k)$, there is a unique vector $(x'_1, \ldots, x'_m)$ for which*

$$\{X_0 = x_0\} \cap \bigcap_{i=1}^{k} \bigcap_{j=1}^{n_i} \{V_{s_i,j} = v_{s_i,\sigma_i(j)}\} = \{X_0 = x_0, X_1 = x'_1, \ldots, X_m = x'_m\}$$

and $(x_0, x_1, \ldots, x_m) \sim (x_0, x'_1, \ldots, x'_m)$.

Proof. It suffices to prove (b), since (a) is obvious. No generality is lost by assuming that $x_m = b$, where b could be different from each s_i. Define $\zeta = (x_0, x'_1, \ldots, x'_{r+1})$ to be the longest string determined by the array $\{(v_{s_i,\sigma_i(1)}, \ldots, v_{s_i,\sigma_i(n_i)}) : i = 1, \ldots, k\}$. See Example 1.

First we prove that $x'_{r+1} = b$. In fact, if $x'_{r+1} = s_j \neq b$, we would have $\zeta = (x_0, \ldots, s_j, v_{s_j,\sigma_j(1)}, \ldots, v_{s_j,\sigma_j(n_j)}, \ldots, s_j)$ with $v_{s_j,\sigma_j(n_j)} = v_{s_j,n_j} \neq s_j$ since $x_m = b \neq s_j$. Thus, ζ would contain (n_j+1) terms s_j, a flagrant contradiction when $x_m = b \neq s_j$. Then $x'_{r+1} = b$, and vectors ζ, $(x_0, \ldots, x_m)$ have the same number of b's.

Now we have to extend the last assertion to all $s_i \neq b$. To this aim define A to be the set of all $s_i \neq b$ satisfying $\sum_{\nu=0}^{r} \mathbf{I}_{\{s_i\}}(x'_\nu) = n_i$ with $x'_0 = x_0$. Thus, if some $s_i \neq b$ does not belong to A, there is q in $\{0, \ldots, m-1\}$ such that $q = \max\{j : 0 \leq j \leq m-1, x_j \notin A\}$, which entails $x_{q+1} = v_{x_q,n_{x_q}} \in A$. Since there is j such that $x_q = s_j \notin A$, then $x_{q+1} = v_{s_j,n_j} \notin A$, a contradiction. Therefore, $s_i \in A$ for any $s_i \neq b$, too. Thus, $r+1 = m$ and $(x_0, x'_1, \ldots, x'_m) \sim (x_0, x_1, \ldots, x_m)$. $\square$

Proof of Theorem 1.
Let us prove the "if" part of the theorem first. Consider the events

$$E = \bigcap_{i=1}^{k} \bigcap_{j=1}^{n_i} \{V_{s_i,j} = v_{s_i,j}\},$$

$$E' = \bigcap_{i=1}^{k} \bigcap_{j=1}^{n_i} \{V_{s_i,j} = v_{s_i,\sigma_i(j)}\},$$

with $k \in \mathbb{N}$, $s_1, \ldots, s_k \in S$, $n_1, \ldots, n_k \in \mathbb{N}$, $v_{s_i,j}$ in S^*, σ_i a permutation of $(1, \ldots, n_i)$, $(i = 1, \ldots, k)$. Since (a) and (b) imply that no state can be visited by X a finite number of times only, for every $s \in S$

$$\bigcup_{n \geq 1} \{V_{s,n} = \partial\} = \bigcap_{n \geq 1} \{V_{s,n} = \partial\} \qquad \text{a.s.-}P,$$

and we can limit ourselves to consider $v_{s_i,j}$ in S for all i and j. Moreover, there is no loss of generality in assuming that $\sigma_i(n_i) = n_i$ $(i = 1, \ldots, k)$. In this case there is a countable set of pairwise disjoint events e_ℓ of the type $\{X_0 = x_0, X_1 = x_{1,\ell}, \ldots, X_{\nu_\ell} = x_{\nu_\ell,\ell}\}$ such that $E = \cup e_\ell$ and, by (a) of Lemma 1, each e_ℓ admits an expression $\tilde{e}_\ell$ in terms of elements of the V-matrix. Let $\sigma(\tilde{e}_\ell)$ be the result of the application to $\tilde{e}_\ell$ of the permutations

σ_i which map E into E'. Then $E' = \cup\sigma(\tilde{e}_\ell)$, with $\sigma(\tilde{e}_\ell)\cap\sigma(\tilde{e}_m) = \emptyset$ if $\ell \neq m$. From (b) of Lemma 1, each $\sigma(\tilde{e}_\ell)$ admits a unique representation of the type $e'_\ell = \{X_0 = x_0, X_1 = x'_{1,\ell}, \dots, X_{\nu_\ell} = x'_{\nu_\ell,\ell}\}$ with $(x_0, x_{1,\ell}, \dots, x_{\nu_\ell,\ell}) \sim (x_0, x'_{1,\ell}, \dots, x'_{\nu_\ell,\ell})$. Thus

$$\begin{aligned} P(E) &= \sum_\ell P(e_\ell) \\ &= \sum_\ell P(e'_\ell) \\ &= \sum_\ell P(\sigma(\tilde{e}_\ell)) = P(E'). \end{aligned}$$

To prove the "only if" part, let δ_x be the degenerate probability measure defined by $\delta_x(\{x\}) = 1$, $x \in S^*$. By de Finetti's representation theorem for partially exchangeable random variables we have that, for every $s \in S$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{V_{s,i}} = \theta_s \qquad \text{a.s.-}P,$$

where the limit appearing in the left-hand side is in the topology of weak convergence. Moreover, conditionally on $\theta = (\theta_s)_{s\in S}$, the random variables $V_{s,j}$ are independent with

$$P\Big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{n_i}\{V_{s_i,j} = v_{s_i,j}\} \mid \theta\Big) = \prod_{i=1}^{k}\prod_{j=1}^{n_i}\theta_{s_i}(v_{s_i,j}),$$

for every $k \in \mathbb{N}$, $s_1, \dots, s_k \in S$, $n_1, \dots, n_k \in \mathbb{N}$, and $v_{s_i,j}$ in S^*. Define $\tilde{t}$ by setting $\tilde{t}(s,\cdot) = \theta_s$ for every s in S. Then, a.s.-P,

$$\begin{aligned} &P(X_1 = x_1, \dots, X_n = x_n \mid \tilde{t}\,) \\ &\quad = P\Big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{n_i}\{V_{s_i,j} = v_{s_i,j}\} \mid \tilde{t}\,\Big) \\ &\quad = \prod_{i=1}^{k}\prod_{j=1}^{n_i}\tilde{t}(s_i, v_{s_i,j}) \\ &\quad = \tilde{t}(x_0, x_1)\dots\tilde{t}(x_{n-1}, x_n), \end{aligned}$$

where $s_1, \dots, s_k$ are the distinct elements of $(x_0, \dots, x_n)$. To prove (b), it is enough to show that if V is partially exchangeable, then x_0 is a recurrent state. We have the inclusions $\{X_n = x_0 \text{ i.o.}\}^c \subset \{\theta_{x_0} = \delta_\partial\} \subset \{\theta_{x_0}(\partial) > 0\}$. Therefore

$$\begin{aligned} P(X_1 \in S) &= E(P(X_1 \in S \mid \theta_{x_0})) \\ &= E(\theta_{x_0}(S)) \\ &= 1 - E(\theta_{x_0}(\partial)) < 1 \end{aligned}$$

unless $E(\theta_{x_0}(\partial)) = 0$. Hence, $\theta_{x_0}(\partial) = 0$ almost surely. In order to prove (c), define $\mathcal{K}$ to be the intersection of the three sets

$$\Big\{t : \textstyle\sum_n t^n(x_0, x_0) = \infty\Big\},$$

$$\bigcap_{x \in S, y \in S^*} \Big\{t : \text{if } \textstyle\sum_n t^n(x_0, x) = \infty \text{ and } \sum_n t^n(x_0, y) < \infty, \text{ then } t(x, y) = 0\Big\},$$

$$\bigcap_{y \in S} \Big\{t : \text{if } \textstyle\sum_n t^n(x_0, y) < \infty, \text{ then } t(y, \partial) = 1\Big\},$$

where t^n is the n-step transition. The set $\mathcal{K}$ is measurable and, with $A_t = \{x \in S : \sum_n t^n(x_0, x) = \infty\}$, it satisfies KD1-KD3. From the definition of $\tilde{t}$ and part (b), one can easily see that $\tilde{t}$ belongs to $\mathcal{K}$ almost surely. □

Since in Theorem 1 the distribution of X is determined by the values of $\tilde{t}(u, v)$ for $u, v \in S$ only, one may wonder if it would be possible to obtain an analogous result considering a random transition matrix from S to S (instead of S^*). The example below shows that uniqueness is not guaranteed in that case.

Example 2. Let $S = \{1, 2\}$, $P(X_0 = 1) = 1$ and let $\tilde{r}$ be a random transition matrix on S^*, whose distribution is defined by

$$P(\{\tilde{r}(1,1) \in dx,\ \tilde{r}(2,1) \in dy\}) = \alpha\delta_1(dx)\nu_1(dy) + (1 - \alpha)\nu_2(dx\, dy),$$

where $\alpha \in (0, 1)$ and the support of ν_2 is included in $[a, b] \times [0, 1]$ with $0 < a < b < 1$. Suppose that

$$P(X_1 = x_1, \dots, X_n = x_n \mid \tilde{r}\,) = \tilde{r}(1, x_1)\tilde{r}(x_1, x_2) \dots \tilde{r}(x_{n-1}, x_n) \qquad \text{a.s.-}P$$

for every n and $x_1, \dots, x_n$ in S. One can define another random transition matrix $\tilde{r}_1$ as follows:

$$\tilde{r}_1(u,v) = \begin{cases} Z & \text{if } u=2,\ v=1 \text{ and } \tilde{r}(1,1)=1 \\ \tilde{r}(u,v) & \text{otherwise} \end{cases}$$

where Z is any random variable defined on $(\Omega, \mathcal{F}, P)$ with support included in $[0,1]$. The distribution of $\tilde{r}_1$ is

$$P(\{\tilde{r}_1(1,1) \in dx,\ \tilde{r}_1(2,1) \in dy\}) = \alpha\delta_1(dx)\nu_1^Z(dy) + (1-\alpha)\nu_2(dx\,dy),$$

where ν_1^Z is the distribution of Z. Consider two Borel sets B_1 and B_2 of $\Re$ and the events $A_1 = \{\tilde{r}_1(1,1) \in B_1\}$ and $A_2 = \{\tilde{r}_1(2,1) \in B_2\}$. Denoting with n_{ij} the number of transitions from i to j in the string $(1, x_1, \dots, x_n)$, one has

$$\begin{aligned}
\int_{A_1\cap A_2} & \tilde{r}_1(1,x_1)\tilde{r}_1(x_1,x_2)\dots\tilde{r}_1(x_{n-1},x_n)\,\mathrm{d}P \\
&= \int_{A_1\cap A_2\cap\{\tilde{r}(1,1)=1\}} \prod_{i,j}\big(\tilde{r}_1(i,j)\big)^{n_{ij}}\,\mathrm{d}P \\
&\quad + \int_{A_1\cap A_2\cap\{\tilde{r}(1,1)\neq 1\}} \prod_{i,j}\big(\tilde{r}_1(i,j)\big)^{n_{ij}}\,\mathrm{d}P \\
&= P(A_1\cap A_2\cap\{\tilde{r}(1,1)=1\})\mathbf{I}_{\{0\}}(n_{12}) \\
&\quad + \int_{\{\tilde{r}(1,1)\in B_1\}\cap\{\tilde{r}(2,1)\in B_2\}\cap\{\tilde{r}(1,1)\neq 1\}} \prod_{i,j}\big(\tilde{r}(i,j)\big)^{n_{ij}}\,\mathrm{d}P \\
&= P(\{X_1=x_1,\dots,X_n=x_n\}\cap A_1\cap A_2\cap\{\tilde{r}(1,1)=1\})\mathbf{I}_{\{0\}}(n_{12}) \\
&\quad + P(\{X_1=x_1,\dots,X_n=x_n\}\cap A_1\cap A_2\cap\{\tilde{r}(1,1)\neq 1\}) \\
&= P(\{X_1=x_1,\dots,X_n=x_n\}\cap A_1\cap A_2).
\end{aligned}$$

This proves that

$$P(X_1=x_1,\dots,X_n=x_n \mid \tilde{r}_1) = \tilde{r}_1(1,x_1)\tilde{r}_1(x_1,x_2)\dots\tilde{r}_1(x_{n-1},x_n) \qquad \text{a.s.-}P$$

Note that almost surely $P(X_n = x_0 \text{ i.o.} \mid \tilde{r}) = P(X_n = x_0 \text{ i.o.} \mid \tilde{r}_1) = 1$. Hence the representation of X as a recurrent Markov chain conditionally on its transition matrix is not unique. The unique random transition matrix $\tilde{t}$

on S^* of Theorem 1 is in this case

$$\left[\tilde{t}(u,v)\right]_{\substack{u\in S\\ v\in S^*}} = \begin{cases} \begin{bmatrix} \tilde{r}(1,1) & \tilde{r}(1,2) & 0 \\ \tilde{r}(2,1) & \tilde{r}(2,2) & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{if } \tilde{r}(1,1) \neq 1 \\ \begin{bmatrix} \tilde{r}(1,1) & \tilde{r}(1,2) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} & \text{if } \tilde{r}(1,1) = 1 \end{cases}$$

□

2.2 On the Relationship Between Partial Exchangeability and Freedman Condition

The S-valued process X, defined on $(\Omega, \mathcal{F}, P)$, is said to obey *Freedman condition* if

$$P(X_0 = x_0, \dots, X_n = x_n) = P(X_0 = x_0', \dots, X_n = x_n')$$

whenever $(x_0, \dots, x_n) \sim (x_0', \dots, x_n')$, $n = 1, 2, \dots$. In the present subsection we give a detailed account of the relationships between Freedman condition and partial exchangeability of the V-matrix of X. The main result is that the latter is a stronger condition than the former, but the two are equivalent under the assumption of recurrence. Before proceeding we need another definition. For any $x \in S$, let $\Lambda_x = \{X_n = x \text{ i.o.}\}$. We say that X is *strongly recurrent* if

$$P\Big(\bigcap_{j=0}^{n} \{X_j = x_j\}\Big) = P\Big(\bigcap_{j=0}^{n} (\{X_j = x_j\} \cap \Lambda_{x_j})\Big) \qquad \text{a.s.-}P$$

for every n and $x_0, \dots, x_n$ in S. Any state is visited by a strongly recurrent process either infinitely many times, or never. The term recurrent will denote a process satisfying the weaker condition $P(\Lambda_{X_0}) = 1$, which under our assumptions can also be written as $P(\Lambda_{x_0}) = 1$.

Remark 1. If either one of the conditions

(a) V is partially exchangeable,

(b) X is strongly recurrent,

is satisfyed, then for every $s \in S$

$$\bigcup_{n\geq 1}\{V_{s,n} = \partial\} = \bigcap_{n\geq 1}\{V_{s,n} = \partial\} \qquad \text{a.s.-}P.$$

Theorem 2. *The V-matrix of the process X is partially exchangeable if and only if X is recurrent and satisfies Freedman condition.*

The proof of the theorem is set out as a sequence of lemmas: the first two prove the "only if" part, while the remaining two prove the "if" part.

Lemma 2. *If V is partially exchangeable, then X is strongly recurrent.*

Proof. Consider the event $\{X_0 = x_0, X_1 = x_1, \dots, X_{n+1} = x_{n+1}\}$. On this event, almost surely, the rows of V corresponding to $x_0, \dots, x_n$ contain no ∂, hence those states are visited infinitely many times. Marginalization over the value of X_{n+1} gives the result. □

Lemma 3. *If V is partially exchangeable, then X satisfies Freedman condition.*

Proof. Consider $(x_0, x_1, \dots, x_n) \sim (x_0, x_1', \dots x_n')$ and let $s_1, \dots, s_k$ be the distinct elements of $(x_0, x_1, \dots, x_{n-1})$. Then there exist $n_1, \dots, n_k$, $v_{s_i,j}$ in S and permutations $\sigma_1, \dots, \sigma_k$ such that

$$\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = \bigcap_{i=1}^{k}\bigcap_{j=1}^{n_i}\{V_{s_i,j} = v_{s_i,j}\},$$

$$\{X_0 = x_0, X_1 = x_1', \dots, X_n = x_n'\} = \bigcap_{i=1}^{k}\bigcap_{j=1}^{n_i}\{V_{s_i,j} = v_{s_i,\sigma_i(j)}\}.$$

Since the probability of the events on the right-hand sides are equal, the claim follows. □

Lemma 4. *If X is strongly recurrent and satisfies Freedman condition, then V is partially exchangeable.*

Proof. Using Remark 1, the main argument used in the proof of the "if" part of Theorem 1 carries over to the present setting. □

Lemma 5. *If X is recurrent and satisfies Freedman condition, then it is strongly recurrent.*

Proof. Since $P(X_0 = x_0) = 1$, the proof of the "if" part of Theorem 1 can be paraphrased to show that the random elements $(V_{x_0,n})_{n\geq 1}$ are exchangeable. Thus, letting $A^+ = \{\tilde{\theta}_{x_0}(x_1) > 0\}$ and using Cantelli lemma, we can write

$$\begin{aligned} P(X_1 = x_1) &= P(V_{x_0,1} = x_1) \\ &= E(\tilde{\theta}_{x_0}(x_1)\mathbf{I}_{A^+}) \\ &= E\big(\mathbf{I}_{A^+}P(\{V_{x_0,1} = x_1\} \cap \{V_{x_0,n} = x_1 \text{ i.o.}\} \mid \tilde{\theta}_{x_0})\big) \\ &= E\big(P(\{V_{x_0,1} = x_1\} \cap \{V_{x_0,n} = x_1 \text{ i.o.}\} \mid \tilde{\theta}_{x_0})\big) \\ &= P(\{V_{x_0,1} = x_1\} \cap \{V_{x_0,n} = x_1 \text{ i.o.}\}), \end{aligned}$$

which implies $\{X_1 = x_1\} \subset \{V_{x_0,n} = x_1 \text{ i.o.}\} \subset \Lambda_{x_0}$, a.s.-$P$. Therefore, the definition of strong recurrence is satisfied with $n = 1$. Suppose, inductively, that it holds for a fixed but arbitrary positive integer ν. Consider $x_1, \dots, x_{\nu+1}$ such that $P(X_1 = x_1, \dots, X_\nu = x_\nu) > 0$. Let $P^{(\nu)}(\cdot) = P(\cdot \cap \{X_1 = x_1, \dots, X_\nu = x_\nu\})/P(X_1 = x_1, \dots, X_\nu = x_\nu)$ and $X_n^{(\nu)} = X_{\nu+n}$. Then under $P^{(\nu)}$ the process $X^{(\nu)}$ starts at x_ν at time zero and

$$P^{(\nu)}(X_n^{(\nu)} = x_\nu \text{ i.o.}) = \frac{P(\{X_1 = x_1, \dots, X_\nu = x_\nu\} \cap \Lambda_{x_\nu})}{P(X_1 = x_1, \dots, X_\nu = x_\nu)} = 1$$

by the inductive hypothesis. Moreover, if $(x_\nu, \dots, x_{\nu+n}) \sim (x'_\nu, \dots, x'_{\nu+n})$, then $(x_0, \dots, x_\nu, \dots, x_{\nu+n}) \sim (x_0, \dots, x'_\nu, \dots, x'_{\nu+n})$ and $P^{(\nu)}(X_0^{(\nu)} = x_\nu, \dots, X_n^{(\nu)} = x_{\nu+n}) = P^{(\nu)}(X_0^{(\nu)} = x'_\nu, \dots, X_n^{(\nu)} = x'_{\nu+n})$. Hence, under $P^{(\nu)}$, $X^{(\nu)}$ is recurrent and satisfies Freedman condition. From the first part of the proof, one has that $\{X_1^{(\nu)} = x_{\nu+1}\} \subset \{X_1^{(\nu)} = x_{\nu+1}\} \cap \Lambda_{x_{\nu+1}}$, a.s.-$P^{(\nu)}$. This implies that $P(X_1 = x_1, \dots, X_{\nu+1} = x_{\nu+1}) = P(\{X_1 = x_1, \dots, X_{\nu+1} = x_{\nu+1}\} \cap \Lambda_{x_{\nu+1}})$. ☐

It is not difficult to show that without recurrence, partial exchangeability of V is stronger than Freedman condition. For example, let $S = \{0, 1, 2\}$, $x_0 = 0$, and let P satisfy $P(X_n = 2) = 1$ for $n \geq 4$ and $P(X_1 = 1, X_2 = 0, X_3 = 0) = P(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{1}{2}$. Then Freedman condition is satisfied, but the V-matrix of X is not partially exchangeable, since

$$\frac{1}{2} = P(V_{0,1} = 1, V_{0,2} = 0, V_{0,3} = 2) \neq P(V_{0,1} = 1, V_{0,2} = 2, V_{0,3} = 0) = 0.$$

¿From Theorem 1 and the equivalence result of the present subsection, one can easily reobtain the following result of Diaconis and Freedman (1985).

Theorem 3 (Diaconis and Freedman). *The process X satisfies Freedman condition and $P(X_0 = x_0 \text{ i.o.}) = 1$ if and only if there is a probability measure μ on the set of all transition matrices on S such that:*

(a) $\mu(\{r : x_0 \text{ is recurrent w.r.t. } r\}) = 1$,

(b) *for every $n \geq 1$ and $x_1, \dots, x_n$ in S,*

$$P(X_1 = x_1, \dots, X_n = x_n) = \int r(x_0, x_1) \dots r(x_{n-1}, x_n)\, \mu(\mathrm{d}r).$$

3 Mixtures of Laws of Markov Chains With Uncountable State Space

3.1 Mixtures of Laws of Markov Chains with Atomic Kernels

A representation theorem analogous to the one presented in Section 2 can be given also in the case of a process with continuous state space. We start by reviewing the definitions contained in the beginning of that section, modifying them as needed in the more general setting. Let S be a Polish space with Borel σ-algebra $\mathcal{S}$. Moreover, let $S^* = S \cup \{\partial\}$, with $\partial \notin S$, and let $\mathcal{S}^*$ be the σ-algebra generated by $\mathcal{S}$ and $\{\partial\}$. $X = (X_n)_{n \geq 0}$ is an S-valued discrete-time stochastic process defined on $(\Omega, \mathcal{F}, P)$, with $X_0 = x_0$ for some $x_0 \in S$. In order to exploit de Finetti's theorem for exchangeable random variables, we consider a discretization of the state space. Accordingly, the definition of the V-matrix of the process has to be slightly generalized. Consider a countable measurable partition of S^*, $\mathcal{E} = \{E_i\}_{i \geq 0}$, with $E_0 = \{\partial\}$, and assume, without loss of generality, that $x_0 \in E_1$. For any positive integer j, consider the sequence $(\nu_{j,n})_{n \geq 1}$ of visiting times of X to E_j and, for j and n in $\mathbb{N}$, define the (j, n) element of the V-matrix of the process X, relative to $\mathcal{E}$, to be the value of X immediately after the nth visit to E_j:

$$V_{j,n} = X_{\nu_{j,n}+1}.$$

As we did in Section 2, we put $V_{j,n} = \partial$ whenever $\nu_{j,n} = \infty$. Let T^* be the set of all kernels from $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ to S^*, endowed with the smallest σ-algebra that makes all the maps $t \mapsto t(n, A)$, $n \in \mathbb{N}_0$, $A \in \mathcal{S}^*$, measurable. One can associate to any element t of T^* a transition kernel on S^* by setting

$$k_t(x, A) = \sum_{j \geq 0} \mathbf{I}_{E_j}(x) t(j, A).$$

In order to achieve uniqueness of the mixing measure in Theorem 4, we restrict its support to a measurable set $\mathcal{K} \subset T^*$ of transition kernels t for which there is a set $A_t \subset \mathbb{N}_0$ satisfying

KC1. $1 \in A_t$, $0 \notin A_t$;

KC2. $t(k, E_j) = 0$ if $k \in A_t$ and $j \notin A_t$;

KC3. $t(j, \{\partial\}) = 1$ if $j \notin A_t$.

Theorem 4. *The V-matrix of X is partially exchangeable if and only if there is a random element $\tilde{t}$ of T^* such that*

(a) $P(X_1 \in B_1, \ldots, X_n \in B_n \mid \tilde{t}\,)$
$= \int_{B_1} \cdots \int_{B_n} k_{\tilde{t}}(x_0, \mathrm{d}x_1) k_{\tilde{t}}(x_1, \mathrm{d}x_2) \ldots k_{\tilde{t}}(x_{n-1}, \mathrm{d}x_n) \qquad$ *a.s.-P;*

(b) $P(X_n \in E_1 \text{ i.o.} \mid \tilde{t}\,) = 1 \qquad$ *a.s.-P;*

(c) $P(\tilde{t} \in \mathcal{K}) = 1$ *for a measurable set $\mathcal{K}$ satisfying* KC1-KC3.

Moreover, $\tilde{t}$ is unique in distribution.

The proof of the theorem is based on three lemmas, essentially dealing with various properties of the discretized version of the process X defined by

$$Y_n = \sum_{j \geq 0} j \, \mathbf{I}_{E_j}(X_n), \qquad n = 0, 1, \ldots .$$

Note that the results of Section 2 apply to Y, and the sequences of visiting times of X to E_j and of Y to state j coincide for every j.

Lemma 6. *Let $i, j \in \mathbb{N}_0$, $B \in \mathcal{S}$ and $A \in \sigma\{X_1, \ldots, X_{n-1}, Y_n, X_{n+1}, \ldots\}$. If* (a) *of* Theorem 4 *holds then, on the set $\{\omega : \tilde{t}(i, E_j) > 0\}$,*

$$\begin{aligned} &P(A \cap \{Y_{n-1} = i\} \cap \{Y_n = j\} \cap \{X_n \in B\} \mid \tilde{t}) \\ &= \tilde{t}(i, B \mid E_j) P(A \cap \{Y_{n-1} = i\} \cap \{Y_n = j\} \mid \tilde{t}) \end{aligned}$$

almost surely, where $\tilde{t}(i, B|E_j) = \tilde{t}(i, B \cap E_j)/\tilde{t}(i, E_j)$.

Proof. It is enough to consider $A = \{(X_1, \dots, X_{n-1}) \in C\} \cap \{(X_{n+1}, \dots, X_{n+k}) \in D\}$. We have

$$
\begin{aligned}
&P((X_1, \dots, X_{n-1}) \in C, Y_{n-1} = i, Y_n = j, X_n \in B, (X_{n+1}, \dots, X_{n+k}) \in D \mid \tilde{t}) \\
&\quad = \int_{C \times D} k_{\tilde{t}}(x_0, \mathrm{d}x_1) \dots k_{\tilde{t}}(x_{n-2}, \mathrm{d}x_{n-1}) \mathbf{I}_{E_i}(x_{n-1}) \tilde{t}(i, B \cap E_j) \tilde{t}(j, \mathrm{d}x_{n+1}) \\
&\qquad\qquad k_{\tilde{t}}(x_{n+1}, \mathrm{d}x_{n+2}) \dots k_{\tilde{t}}(x_{n+k-1}, \mathrm{d}x_{n+k}) \\
&\quad = \tilde{t}(i, B \mid E_j) \int_{C \times D} k_{\tilde{t}}(x_0, \mathrm{d}x_1) \dots k_{\tilde{t}}(x_{n-2}, \mathrm{d}x_{n-1}) \mathbf{I}_{E_i}(x_{n-1}) \tilde{t}(i, E_j) \\
&\qquad\qquad \tilde{t}(j, \mathrm{d}x_{n+1}) k_{\tilde{t}}(x_{n+1}, \mathrm{d}x_{n+2}) \dots k_{\tilde{t}}(x_{n+k-1}, \mathrm{d}x_{n+k}) \\
&\quad = \tilde{t}(i, B \mid E_j) P(A \cap \{Y_{n-1} = i\} \cap \{Y_n = j\} \mid \tilde{t})
\end{aligned}
$$

□

Lemma 7. *Let* $j, n, s \in \mathbb{N}$ *and* $B \in \mathcal{S}^*$. *Then, if* (a) *of* Theorem 4 *holds,*

$$\{V_{j,n} \in B\} \cap \{\nu_{j,n} \neq s\} \in \sigma\{X_1, \dots, X_s, Y_{s+1}, X_{s+2}, \dots\}.$$

Proof. Write the event above as

$$\bigcup_{t \neq s} \{X_{t+1} \in B, X_t \in E_j, \textstyle\sum_{0 \le m \le t} \mathbf{I}_{E_j}(X_m) = n\}.$$

□

Lemma 8. *Let* $j, n_1, \dots, n_j \in \mathbb{N}$, $A_{i,k} \in \mathcal{S}$ $(i = 1, \dots, j, k = 1, \dots, n_i)$, $\ell : \mathbb{N}^2 \to \mathbb{N}$. *If* (a) *of* Theorem 4 *holds then, on the set* $\{\omega : \tilde{t}(i, E_{\ell(i,k)}) > 0, i = 1, \dots, j, k = 1, \dots, n_i\}$,

$$
\begin{aligned}
&P(V_{1,1} \in A_{1,1} \cap E_{\ell(1,1)}, \dots, V_{j,n_j} \in A_{j,n_j} \cap E_{\ell(j,n_j)} \mid \tilde{t}) \\
&= P(V_{1,1} \in E_{\ell(1,1)}, \dots, V_{j,n_j} \in E_{\ell(j,n_j)} \mid \tilde{t}) \prod_{i=1}^{j} \prod_{k=1}^{n_i} \tilde{t}(i, A_{i,k} \mid E_{\ell(i,k)})
\end{aligned}
$$

almost surely.

Proof. Using Lemma 6 and Lemma 7, we have that

$$
\begin{aligned}
&P(V_{1,1} \in A_{1,1} \cap E_{\ell(1,1)}, \dots, V_{j,n_j} \in A_{j,n_j} \cap E_{\ell(j,n_j)} \mid \tilde{t}) \\
&= \sum_{s=1}^{\infty} P(V_{1,1} \in A_{1,1} \cap E_{\ell(1,1)}, \dots, X_s \in E_j, \\
&\qquad\qquad X_{s+1} \in A_{j,n_j} \cap E_{\ell(j,n_j)}, \nu_{j,n_j} = s \mid \tilde{t}) \\
&= \tilde{t}(j, A_{j,n_j} \mid E_{\ell(j,n_j)}) P(V_{1,1} \in A_{1,1} \cap E_{\ell(1,1)}, \dots, V_{j,n_j} \in E_{\ell(j,n_j)} \mid \tilde{t})
\end{aligned}
$$

The result is obtained by iterating the procedure. □

We are now ready to prove Theorem 4.

Proof of Theorem 4. We prove the "only if" part first. By de Finetti's representation theorem for partially exchangeable random variables we have that, for every $r \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{V_{r,i}} = \theta_r \qquad \text{a.s.-}P.$$

Moreover, conditionally on $\theta = (\theta_r)_{r\in\mathbb{N}}$, the random variables $V_{i,j}$ are independent with

$$P\big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{n_i}\{V_{r_i,j} \in A_j^{(r_i)}\} \mid \theta\big) = \prod_{i=1}^{k}\prod_{j=1}^{n_i} \theta_{r_i}(A_j^{(r_i)}),$$

for every $k, r_1, \dots, r_k, n_1, \dots, n_k$ in $\mathbb{N}$, and $A_j^{(r_i)}$ in $\mathcal{S}^*$. Define $\tilde{t}$ by setting $\tilde{t}(r,\cdot) = \theta_r$ for every r in $\mathbb{N}$. To prove (a) assume, without real loss of generality, that $B_i \subset E_{r_i}$ for some integers $r_1, \dots, r_n$. If $a_1, \dots, a_k$ are the distinct elements of $(1, r_1, \dots, r_n)$, the event $\{X_1 \in B_1, \dots, X_n \in B_n\}$ can be expressed as $\bigcap_{i=1,\dots,k}\bigcap_{j=1,\dots,n_j}\{V_{a_i,j} \in B_{s(i,j)}\}$ for some integers $n_1, \dots, n_k$ and $s : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$. Therefore,

$$\begin{aligned} &P(X_1 \in B_1, \dots, X_n \in B_n \mid \tilde{t}) \\ &\quad = \prod_{i=1,\dots,k}\ \prod_{j=1,\dots,n_j} \tilde{t}(a_i, B_{s(i,j)}) \\ &\quad = \int_{B_1} \dots \int_{B_n} k_{\tilde{t}}(x_0, \mathrm{d}x_1) k_{\tilde{t}}(x_1, \mathrm{d}x_2) \dots k_{\tilde{t}}(x_{n-1}, \mathrm{d}x_n) \end{aligned}$$

almost surely. This proves (a). Part (b) and (c) can be obtained from (b) and (c) of Theorem 1 applied to the discretized process Y.

To prove the "if" part, observe that, given $\tilde{t}$, Y is a recurrent Markov chain with transition probabilities $\tilde{t}_Y(i,j) = \tilde{t}(i, E_j)$, $i,j \in \mathbb{N}$. Theorem 1 can be invoked to show that the V-matrix of Y is partially exchangeable. This, together with Lemma 8, concludes the proof of the "if" part. □

3.2 Mixtures of laws of Markov chains with atomic kernels. A few additional results

The results proved in the previous subsection can be extended to the more general setting of a stochastic sequence $(X, Y) = ((X_n, Y_n))_{n\geq 0}$ defined on $(\Omega, \mathcal{F}, P)$, where X_n, Y_n take values in S_∂ and $\mathbb{N}_0$, respectively. Here, $(\nu_{j,n})_{n\geq 1}$ is meant as the sequence of visiting times of Y to j, and the (j, n) element of the V-matrix of (X, Y) is the value of X immediately after the nth visit of Y to j: $V_{j,n} = X_{\nu_{j,n}+1}$, with the usual proviso that $X_\infty = \partial$. For the sake of simplicity, we keep the hypothesis that $X_0 = x_0$, and we will focus on probabilities P satisfying the following conditions:

(i) $P(X \in S^\infty) = 1$;

(ii) Y_n and $(X_0, Y_0, \dots, X_{n-1}, Y_{n-1}, V)$ are conditionally independent given X_n, for every $n \geq 1$, and Y_0 and V are conditionally independent given X_0;

(iii) there is a kernel s from S_∂ to $\mathbb{N}_0$ such that $P(Y_n = y|X_n) = s(X_n, y)$ holds a.s.-P for each $n \geq 0$, with $s(x, \mathbb{N}) = 1$ for every $x \in S$ and $s(\partial, \cdot) = \delta_0(\cdot)$.

Preserve the definition of T^* given in Subsection 3.1 and, for each $t \in T^*$ introduce the following transition kernel on S^*

$$k_t^*(x, A) = \sum_{y\in\mathbb{N}_0} s(x, y)t(y, A).$$

To pave the way for an extension of Theorem 4, assume that there is a T^*-valued random element $\tilde{t}$ such that

$$P\{Y_0 = y_0, \dots, Y_n = y_n, X_1 \in A_1, \dots, X_n \in A_n \mid \tilde{t} = t\}$$

$$= s(x_0, y_0) \prod_{r=1}^{n} \int_{A_r} s(x_r, y_r)t(y_{r-1}, \mathrm{d}x_r)$$

holds for every t in T^*, $A_1, \dots, A_n$ in $\mathcal{S}_\partial$ and $y_0, \dots, y_n$ in $\mathbb{N}_0$. It can be shown that there is a countable class of measurable, pairwise disjoint subsets of S: $E_{1,t}, E_{2,t}, \dots$ such that: $t(y, E_{i,t}) = 1$ for each y in $R_{i,t}$ and $t(y, \partial) = 0$ for all $y \in \cup_{k\geq 1} E_{k,t}$, $s(x, R_{i,t}) = 1$ for every $x \in E_{i,t}$.

In view of these remarks and in order to achieve uniqueness of the mixing measure also in the present setting, consider for each $y_0 \in \mathbb{N}$ a measurable set $\mathcal{K}^*_{y_0} \subset T^*$ of transition kernels t for which there are measurable set E_t and a set R_t satisfying

KG1. $x_0 \in E_t$, $\partial \notin E_t$, $y_0 \in R_t$, $0 \notin R_t$;

KG2. $t(k, E_t) = 1$ for every k in R_t;

KG3. $t(k, \{\partial\}) = 1$ if $k \notin R_t$.

By paraphrasing the arguments employed in the previous subsection, we obtain the following proposition.

Theorem 5. *Let P obey* (i).

If the V-matrix of (X, Y) is partially exchangeable and P satisfies (ii) *and* (iii)*, then there is a T^*-valued random element $\tilde{t}$ such that*

(a) $P(X_1 \in B_1, \dots, X_n \in B_n \mid \tilde{t}\,)$
$= \int_{B_1} \dots \int_{B_n} k^*_{\tilde{t}}(x_0, \mathrm{d}x_1) k^*_{\tilde{t}}(x_1, \mathrm{d}x_2) \dots k^*_{\tilde{t}}(x_{n-1}, \mathrm{d}x_n) \qquad$ *a.s.-P;*

(b) $P(\lambda_{Y_0} = +\infty) = 1$*;*

(c) $P(\tilde{t} \in \mathcal{K}^*_{y_0} \mid Y_0 = y_0) = 1$ *for a measurable set $\mathcal{K}^*_{y_0}$ satisfying* KG1-KG3, *and for any $y_0 \in \mathbb{N}$.*

Moreover, $\tilde{t}$ is unique in distribution.

Conversely, assume that (a) *holds for $k^*_{\tilde{t}}$ with s such that $s(x, \mathbb{N}) = 1$ if $x \in S$, and $s(\partial, \cdot) = \delta_0(\cdot)$. Then, there is a stochastic sequence $(Y_n)_{n \geq 0}$ of $\mathbb{N}_0$-valued random elements, for which*

$$P(Y_0 = y_0, \dots, Y_n = y_n, X_1 \in B_1, \dots, X_n \in B_n \mid \tilde{t})$$

$$s(x_0, y_0) \prod_{r=1}^{n} \int_{B_r} s(x_r, y_r) \tilde{t}(y_{r-1}, \mathrm{d}x_r).$$

Moreover, if (b) *and* (c) *are valid for these Y_n, then the resulting V-matrix is partially exchangeable and s satisfies* (ii)-(iii).

We omit the proof of this result. It is provided in the unpublished first draft of the present paper. See Fortini *et al.* (1999).

Suppose that the assumptions of the second part of Theorem 5 are in force. It is easy to verify that a possible choice of the $\mathbb{N}_0$-valued random sequence (Y_n) is the following. Let $Z_0, Z_1, \ldots$ be a sequence of independent random variables on $(\Omega, \mathcal{F}, P)$ with common uniform distribution on $[0,1]$. For every n define

$$\begin{array}{ll} Y_n(\omega) = j & \text{if } Z_n(\omega) \in I_j(X_n(\omega), X_{n+1}(\omega), \tilde{t}(\omega)) \\ Y_n(\omega) = 0 & \text{otherwise,} \end{array}$$

where

$$\begin{array}{ll} I_j(x,y,t) = [\sum_{i=1}^{j-1} s(x,i) f_{i,x}(y), \sum_{i=1}^{j} s(x,i) f_{i,x}(y)] & j \in \{2,3,\ldots\} \\ I_1(x,y,t) = [0, s(x,1) f_{1,x}(y)] & \end{array}$$

and $f_{i,x}$ is a version of the Radon-Nikodym derivative $\frac{t(i,dy)}{k_t^*(x,dy)}$ taking values in $[0, s(x,i)]$ whenever $s(x,i) \neq 0$.

3.3 Approximations of Arbitrary Mixtures of Laws of Markov Chains. Partial Results.

Let $\mathcal{A} \subset \mathcal{S}_\partial$ denote a *countable* convergence-determining class [see Section 2 in Billingsley (1999)] and let $\mathcal{P}$ denote a sequence of countable partitions of S_∂: $\pi_n = \{E_{0,n}, E_{1,n}, \ldots\}, n \geq 1$, with $E_{j,n} \in \mathcal{A}$ for every $j \geq 0$ and $n \geq 1$, such that π_{n+1} is a refinement of π_n for every n. Choose $x_{j,n} \in E_{j,n}$ for every $j \geq 0$ and $n \geq 1$. Now, suppose that M is any mixture of distributions of S_∂-valued Markov chains which have initial state x_0, i.e. : $M(\cdot) = \int_I M_y(\cdot)\lambda(\mathrm{d}y)$ where: $(I, \mathcal{I}, \lambda)$ is some probability space and, for any y in I, M_y is the distribution of a Markov chain with initial state x_0 and transition probability K_y such that $K_y(x, S) = 1$ for every $x \in S$. In the sequel, we will assume that $x_0 \in E_{0,n}$ and set $x_{0,n} = x_0$ for every n. Again, given any $n \geq 1$ and $y \in I$, we will define $K_y^{(n)}(x, \cdot) = K_y(x_{j,n}, \cdot)$ whenever $x \in E_{j,n}$ $(j \geq 0, n \geq 1)$ and indicate by $M_y^{(n)}$ the Markov distribution on $(S_\partial^\infty, \mathcal{S}_\partial^\infty)$ with initial state x_0 and transition probability $K_y^{(n)}$.

Theorem 6 proves that, in the presence of some additional conditions for $\{K_y : y \in I\}$, M can be approximated by $M^{(n)} := \int_I M_y^{(n)}\lambda(\mathrm{d}y)$ in the topology of weak convergence.

Theorem 6. *If, for λ-almost all y, we suppose that*

(a) $x \to K_y(x, A)$ *is continuous for each* $A \in \mathcal{A}$*;*

(b) $\sup_j \sup_{x \in E_{j,n}} |K_y(x, A) - K_y(x_{j,n}, A)| \to 0$ *as* $n \to \infty$*, for every* $A \in \mathcal{A}$*,*

then $M^{(n)} \Rightarrow M$ *as* $n \to \infty$*. Moreover, if, besides* (a)-(b) *we have*

(c) $\sum_{m \geq 1} \sum_{j_1, \ldots, j_{m-1}} K_y(x_0, E_{j_1,n}) \ldots K_y(x_0, E_{j_{m-1},n}) = +\infty$ *for λ-almost all y,*

then the elements of $V^{(n)}$ *are partially exchangeable.*

Thus, any mixture of distributions of Markov chains, which satisfy (a)-(c), can be thought of as the limiting law of a sequence of processes $X^{(n)}$ which yield partially exchangeable matrices $V^{(n)}$ (with respect to suitable partitions of S_∂).

Before proving Theorem 6, we recall a simple fact about condition (c). If $(\xi_\nu)_{\nu \geq 0}$ is a Markov chain with initial state x_0 and transition probability $K_y^{(n)}$, then the sequence $(\xi'_\nu)_{\nu \geq 0}$, defined by

$$\xi'_\nu = \sum_{j \geq 0} x_{j,n} \mathbf{I}_{E_{j,n}}(\xi_\nu)$$

for any $\nu \geq 0$, is a Markov chain with state space $\{x_0, x_{1,n}, x_{2,n}, \ldots\}$, initial state x_0 and one-step transition probabilities $k_y^{(n)}(x_{i,n}, x_{j,n}) = K_y(x_{i,n}, E_{j,n})$. Hence, condition (c) is equivalent to recurrence of $(\xi'_\nu)_{\nu \geq 0}$.

Proof of Theorem 6. In view of well known results on weak convergence on product spaces, it suffices to prove that $M_y^{(n)}(A_1 \times \cdots \times A_\nu \times S_\partial^\infty) \to M_y(A_1 \times \cdots \times A_\nu \times S_\partial^\infty)$ as $n \to \infty$, for every $A_1, A_2, \ldots$ in $\mathcal{A}$, for every ν and for λ-almost all y [recall that $\mathcal{A}$ is countable]. In fact, for $\nu = 1$, the statement holds true. Now, suppose that it is valid for some $\nu > 1$. Then,

$$|M_y^{(n)}(A_1 \times \cdots \times A_{\nu+1} \times S_\partial^\infty) - M_y(A_1 \times \cdots \times A_{\nu+1} \times S_\partial^\infty)|$$

$$\leq \left| \int_{A_1 \times \cdots \times A_\nu} K_y^{(n)}(x_0, \mathrm{d}x_1) \ldots K_y^{(n)}(x_{\nu-1}, \mathrm{d}x_\nu) \{K_y^{(n)}(x_\nu, A_{\nu+1}) - K_y(x_\nu, A_{\nu+1})\} \right|$$

$$+ \left| \int_{A_1 \times \cdots \times A_\nu} K_y^{(n)}(x_0, \mathrm{d}x_1) \ldots K_y^{(n)}(x_{\nu-1}, \mathrm{d}x_\nu) K_y(x_\nu, A_{\nu+1}) \right.$$

$$- \int_{A_1 \times \cdots \times A_\nu} K_y(x_0, \mathrm{d}x_1) \dots K_y(x_{\nu-1}, \mathrm{d}x_\nu) K_y(x_\nu, A_{\nu+1}) \Bigg| .$$

The former addend is majorized by $\sup_{x \in S} |K_y^{(n)}(x, A_{\nu+1}) - K_y(x, A_{\nu+1})|$, which, by (b), converges to 0 as $n \to \infty$, and the latter one converges to 0 from the inductive hypothesis and condition (a). Thus, $M_y^{(n)} \Rightarrow M_y$ by the induction principle.

To conclude the proof, it is enough to recall that (c) implies recurrence of $E_{0,n}$, for each n, and, then, invoke Theorem 4. □

It is not difficult to conceive different versions of the previous proposition. For example, suppose that there is a σ-finite measure μ on $(S, \mathcal{S})$ such that $K_y(x, \cdot) \ll \mu(\cdot)$ for λ-almost all y and for μ-almost all x, and indicate by $g_y(x, \cdot)$, $g_y^{(n)}(x, \cdot)$ versions of the derivatives of $K_y(x, \cdot)$, $K_y^{(n)}(x, \cdot)$, with respect to μ, respectively. Then, assumptions (a) and (b) in Theorem 6 could be replaced by the sole condition:

(a)' $\int_S |g_y(x, s) - g_y^{(n)}(x, s)| \mu(\mathrm{d}s) \to 0$ *as* $n \to \infty$ *for* λ*-almost all* y *and* μ*-almost all* x.

In fact, suppose that the inductive hypothesis

$$\int_{S^\nu} |g_y(x_0, x_1) \dots g_y(x_{\nu-1}, x_\nu) - g_y^{(n)}(x_0, x_1) \dots g_y^{(n)}(x_{\nu-1}, x_\nu)| \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_\nu) \to 0$$

is valid. Then,

$$\int_{S^{\nu+1}} |g_y(x_0, x_1) \dots g_y(x_\nu, x_{\nu+1}) - g_y^{(n)}(x_0, x_1) \dots g_y^{(n)}(x_\nu, x_{\nu+1})| \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_{\nu+1})$$

$$\leq \int_{S^\nu} g_y(x_0, x_1) \dots g_y(x_{\nu-1}, x_\nu) \int_S |g_y(x_\nu, x) - g_y^{(n)}(x_\nu, x)| \mu(\mathrm{d}x) \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_\nu)$$

$$+ \int_{S^\nu} |g_y(x_0, x_1) \dots g_y(x_{\nu-1}, x_\nu) - g_y^{(n)}(x_0, x_1) \dots g_y^{(n)}(x_{\nu-1}, x_\nu)| \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_\nu).$$

The former addend converges to 0 in view of (a)' and the dominated convergence theorem, while the latter goes to zero by virtue of the inductive hypothesis.

Example 3. As an application of the previous statements, consider the case of a mixture of Markov laws with state space $\Re$, initial state $x_0 = 0$ and transition probabilities: $K_\sigma(x, A) = \int_A \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(s-x)^2}{2\sigma^2}\}\mathrm{d}s$ whenever $x \in \Re$ and $K_\sigma(\partial, \cdot) = \delta_\partial(\cdot)$ for every $\sigma > 0$. It is easy to verify conditions (a)' and (c) when: $E_{0,n} = (-1/2^n, 0]$, $E_{1,n} = (0, 1/2^n]$, $E_{2,n} = (-2/2^n, -1/2^n], \ldots,$ $E_{2(n2^n-1),n} = (-n, -n+1/2^n]$, $E_{n2^{n+1}-1,n} = (n-1/2^n, n]$, $E_{n2^{n+1},n} = (-n, n]^c$; $x_{1,n} = 1/2^n$, $x_{2,n} = -1/2^n$, $\ldots$, $x_{n2^{n+1},n} = -n$. As a matter of fact, for any fixed x and sufficiently large n, there exists $k < n2^{n+1}$ such that $x \in E_{k,n}$. Therefore,

$$\int_{\mathbf{R}} |g_\sigma(x,s) - g_\sigma^{(n)}(x,s)|\mathrm{d}s = (\sigma\sqrt{2\pi})^{-1} \int_{\Re} |\exp(-\frac{s^2}{2\sigma^2}) - \exp(-\frac{(s-(x_{k,n}-x))^2}{2\sigma^2})|\mathrm{d}s$$

$$= 2\{\Phi_\sigma(\frac{x_{k,n}-x}{2}) - \Phi_\sigma(-\frac{x_{k,n}-x}{2})\}$$

$$\leq 2(\sigma\sqrt{2\pi})^{-1}(x_{k,n}-x) \leq (\sigma\sqrt{2\pi})^{-1}/2^{n-1} \to 0$$

as $n \to \infty$, where Φ_σ is the cumulative distribution function of the Gaussian law with mean zero and variance σ^2.

Finally, to prove that (c) holds, define $\xi'_\nu = \sum_{j\geq 0} x_{j,n}\mathbf{I}_{E_{j,n}}(\xi_\nu)$ and observe that, in our present case, $(\xi'_\nu)_{\nu\geq 0}$ has a finite state space and strictly positive one-step transition probabilities.

4 Concluding Remarks

According to de Finetti, exchangeability and partial exchangeability play an important role in the reconstruction of the Bayes-Laplace approach to induction and statistics from the subjective standpoint. In de Finetti's view, statistical inference must confine itself to considering objective hypotheses on something that can actually be observed. This condition is necessary for the phrase "to learn from experience" to have a real meaning when applied to statistical methods. Therefore he suggests to replace the notion of "conditionally independent and identically distributed observations, given the unknown common marginal distribution" with that of exchangeability. In the same spirit, the notion of "observations from a Markov chain, given the unknown transition kernel" needs to be replaced by another involving observable events only, without any reference to metaphysical entities such as unknown transition kernels or unknown probability laws. Partial exchangeability of the V-matrix is such a condition. In fact the elements of the V-matrix are observable and Theorems 1, 4 and 5 establish the equivalence with the usual Bayesian formulation. Moreover, the three theorems show that the prior distribution of the transition kernel is determined by the limiting distribution of the empirical processes associated to the rows of V. This suggests that to elicit a prior distribution on the unobservable transition kernel one can think of the distribution of the observable empirical processes when the number of observations is large.

¿From a practical point of view, partial exchangeability of the V-matrix may be assumed whenever one considers the last outcome before any observation as a relevant attribute of the observation itself and, in addition, once observations are classified according to this attribute, time order becomes irrelevant. In particular, in the situation discussed in Section 3, the influential attribute of each observation is not the actual value of the preceding outcome, but: (a) its belonging to one of the elements of a distinguished partition of the space state (see Subsection 3.1) or, alternatively, (b) the determination of some (random) element Y_n associated with each observation X_n, for every n, provided that each Y_n has countable range (see Subsection 3.2).

AKNOWLEDGEMENTS: The authors wish to thank an anonymous referee for comments and suggestions that greatly improved the presentation of the paper.

REFERENCES

Billingsley, P., 1999. Convergence of probability measures (2nd Edition). Wiley, New York.

Diaconis, P., Freedman, D., 1980a. De Finetti's theorem for Markov chains. Ann. Probab. 8, 115-130.

Diaconis, P., Freedman, D. 1980b. De Finetti's generalizations of chains. Ann. Probab. 8, 115-130.

Diaconis, P. and Freedman, D. ,1980b. De Finetti's generalizations of exchangeability. In: Jeffrey, R.C. (Ed.), Studies in Inductive Logic and Probability, vol.II, Un. of California Press, Berkeley, pp. 233-249.

de Finetti, B., 1937. La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincaré 7, 1-68.

de Finetti, B. 1938. Sur la condition d'"equivalence partielle". Actualités Scientifiques et Industrielles 739, 5-18. Hermann, Paris.

de Finetti, B., 1959. La probabilità e la statistica nei rapporti con l'induzione secondo i diversi punti di vista. In: Atti Corso CIME su Induzione e Statistica, Cremonese, Roma, pp. 1-115. [English traslation in B. de Finetti, 1972. Probability , Induction and Statistics, Wiley, pp. 147-227. New York.]

de Finetti, B., 1972. Probability, Induction and Statistics. Wiley, New York.

Fortini, S., Ladelli, L., Petris, G., and Regazzini, E., 1999. On mixtures of distributions of Markov chains. Preprint.

Freedman, D., 1962a. Mixture of Markov processes. Ann. Math. Statist. 33, 114-118.

Freedman, D., 1962b. Invariants under mixing which generalize de Finetti's theorem. Ann. Math. Statist. 33, 916-923.

Freedman, D., 1996. De Finetti's theorem in continuous time. In: Ferguson, T.S., Sharpley, L.S., MacQueen, J.B. (Eds.) Statistics, Probability and Game Theory. IMS Lecture Notes - Monograph series, Vol.30.

Kallenberg, O., 1982. Characterizations and embedding properties in exchangeability. Z. Wahrscheinlichkeitstheorie 60, 249-281.

Nummelin, E., 1984. General irreducible Markov chains and non-negative operators. Cambridge Univerity Press, Cambridge.

Zabell, S., 1995. Characterizing Markov exchangeable sequences. Journal of Theoretical Probab. 8, 175-178.

Zaman, A., 1984. Urn models for Markov exchangeability. Ann, Probab. 11, 223-229.

Zaman, A., 1986. A finite form of de Finetti's theorem for stationary Markov exchangeability. Ann. Probab. 14, 1418-1427.