Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems – MTNS 2010 • 5–9 July, 2010 • Budapest, Hungary

Stochastic Realization of Binary Exchangeable Processes

Lorenzo Finesso and Cecilia Prosdocimi

***Abstract*— A discrete time stochastic process is called exchangeable if its n-dimensional distributions are, for all n, invariant under permutation. By de Finetti theorem any exchangeable process is representable through a unique, generally infinite, mixture of i.i.d. processes. We formulate, as a stochastic realization problem, the question of characterizing the binary exchangeable processes which are finite mixtures of i.i.d. processes. The realizability conditions and an exact realization algorithm are given in terms of the Hankel matrix of the process. We establish a connection with the realization problem of deterministic positive linear systems of the relaxation type.**

I. AIM AND INTRODUCTION

The stochastic process (Y_n), $n = 1, 2\dots$ is said to be exchangeable if, for all n, the $n!$ permutations of the random variables $Y_1, \dots Y_n$ have the same joint distribution. Exchangeable processes are commonplace in Bayesian statistics. If for each $\theta \in \Theta$ we model the observations (Y_n) as an i.i.d. process with n-dimensional distribution $p(Y_1^n|\theta)$ then, by the law of total probability, the marginal distribution of the observations is

$$p(Y_1^n) := \int_\Theta p(Y_1^n|\theta)\mu(d\theta),$$

which clearly makes (Y_n) exchangeable. The remarkable de Finetti theorem asserts that all exchangeable processes can be represented as above i.e. as a, generally infinite, mixture of i.i.d. processes. If the support of the mixing measure μ is finite, (Y_n) is a *finite mixture of i.i.d. processes*,

$$p(Y_1^n) = \sum_{k=1}^{N} \mu(\{\theta_k\})p(Y_1^n|\theta_k). \quad (1)$$

Although these are very special exchangeable processes, they constitute a versatile class of models, usefully employed in many practical applications. Finite mixtures of i.i.d. processes are an appropriate statistical model when the population is naturally divided into a finite number of clusters, and the time evolution of a sample is i.i.d., but with distribution dependent on which cluster the sample belongs to. It is not only of theoretical, but also of practical interest to pose and solve the problem of characterizing those exchangeable processes which are finite mixtures and to devise algorithms for model construction from distributional data. Previous work on the theoretical problem of characterizing the finite mixtures among all exchangeable processes dates back to Dharmadhikari [5], where a hidden Markov model argument was used. We restrict attention to the binary case, i.e. the random variables Y_n take values onto $\{0, 1\}$ (in Section II we collect all the needed definitions and give some insight on excheangeability). The characterization of finite mixtures is rephrased as a stochastic realization problem in Section III. Thus formulated, the characterization can be completely solved resorting to known results on the classic moment problem on the unit interval. A fringe benefit of our approach compared to [5] is that both the realizability conditions as a finite mixture, and the actual computation of the parameters of a realization, presented in Section IV, can be based on the analysis of a set of Hankel matrices which are defined solely in terms of the given distributions $p_Y(y_1^n)$. The theory developed here can be applied to solve the realization problem of deterministic positive linear system of the relaxation type (as sketched in Section V). Possible ways of extending and generalizing these results are commented upon in the final Section of the paper.

L. Finesso is with the Institute of Biomedical Engineering, ISIB-CNR, Padova, Italy. lorenzo.finesso@isib.cnr.it

C. Prosdocimi is with the DIMAD Department, University of Florence, Italy. cecilia.prosdocimi@gmail.com

II. EXCHANGEABILITY AND MIXTURES

Let $\sigma = \sigma_1^n = \sigma_1 \dots \sigma_n \in \{0, 1\}^n$ be a binary string of length n. Denoting by $S(n)$ the group of permutations of $I_n := \{1, 2, \dots, n\}$ we say that the binary string τ of length n is a *permutation* of σ if there exists a permutation $\pi \in S(n)$ such that

$$\tau_1^n = \tau_1\tau_2 \dots \tau_n = \sigma_{\pi(1)}\sigma_{\pi(2)} \dots \sigma_{\pi(n)}.$$

Definition 1: The binary stochastic process (Y_n), $n = 1, 2, \dots$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values into $\{0, 1\}^\infty$, is *exchangeable* if for all n, for all $\sigma = \sigma_1^n$ and all permutations τ of σ

$$\mathbb{P}\{Y_1^n = \sigma_1^n\} = \mathbb{P}\{Y_1^n = \tau_1^n\}.$$

Example 1: Let (δ_n), $n = 1, 2, \dots$ be an i.i.d. binary process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\bar{Y}$ a binary random variable on the same probability space, not necessarily independent from (δ_n). The process $Y_n := \bar{Y}\,\delta_n$ is exchangeable.

Note that for most of this section the restriction to binary processes is not required. We phrase everything in terms of binary processes since the technique we use to prove the main results is tailored to the binary case. As an immediate consequence of the definition, exchangeable processes are stationary, moreover their random variables are identically distributed. The simplest example of exchangeable process is an i.i.d process. For a finite binary string y_1^n let $n_1 = \sum_{t=1}^n y_t$ be the number of 1-s in the string, and $p = \mathbb{P}\{Y_1 = 1\}$. Then for an i.i.d. process (Y_n)

$$\mathbb{P}\{Y_1^n = y_1^n\} = \prod_{t=1}^n \mathbb{P}\{Y_1 = y_t\} = p^{n_1}\,(1-p)^{n-n_1}, \quad (2)$$

ISBN 978-963-311-370-7

which, for a given length n, only depends on n_1, not on the locations of the 1's in the string y_1^n. The i.i.d. processes are the building blocks for the general exchangeable process which is characterized by the classic de Finetti theorem below. We need the following

Definition 2: A binary process (Y_n) is a *mixture of i.i.d. processes* if there exists a measure μ on the interval $[0, 1]$ such that

$$\mathbb{P}\{Y_1^n = y_1^n\} = \int_0^1 p^{n_1} (1-p)^{n-n_1} d\mu(p). \quad (3)$$

Theorem 1: (de Finetti [3]) (Y_n) is exchangeable if and only if (Y_n) is a mixture of i.i.d. processes. When this is the case the measure μ in (3) is unique.
”Exchangeable process” and ”mixture of i.i.d. processes” are therefore synonymous.

Remark 1: In engineering applications it is more natural to write models in terms of equations that describe the data generating mechanism. In this respect it can be shown that any exchangeable process (Y_n) can be represented as

$$Y_n \stackrel{d}{=} f(\bar{Y}, \delta_n), \quad (4)$$

where (δ_n) is an i.i.d. process, $\bar{Y}$ a random variable, not necessarily independent from (δ_n), and f a deterministic function. The symbol $\stackrel{d}{=}$ indicates that the two processes have the same distribution. Thus, for the process (Y_n) in Example 1, we can write $Y_n = f(\bar{Y}, \delta_n)$, where $f(x, y) = xy$.

From the operational point of view, it is easy to simulate a binary exchangeable process generating its random variables through a two-step stochastic mechanism. In the first step one selects randomly, according to the measure μ, the parameter p. The second step consists in using the selected p to generate i.i.d. random variables $Y_1, Y_2, \ldots$ according to (2).

Imposing constraints on the de Finetti measure μ one narrows the class of exchangeable processes. In this paper attention will be restricted to the subclasses defined below.

Definition 3: (Y_n) is a *finite* mixture of i.i.d. processes if the de Finetti measure μ in equation (3) is concentrated on a finite set, i.e. there exist $N < \infty$ points $p_1, p_2, \ldots, p_N$ in $[0, 1]$, with $\mu_k := \mu(\{p_k\})$ such that $\sum_{k=1}^N \mu_k = 1$. The measure μ is therefore

$$\mu(\cdot) = \sum_{k=1}^N \mu_k \delta_{p_k}(\cdot), \quad (5)$$

where δ_p indicates the Dirac measure concentrated in p.
(Y_n) is a *countable* mixture of i.i.d. processes if μ is concentrated on a countable set $K \subset [0, 1]$.

For a finite binary mixture of i.i.d. processes equation (3), which gives the probability of a trajectory, takes the explicit form

$$\begin{aligned}\mathbb{P}\{Y_1^n = y_1^n\} &= \sum_{k=1}^N \mu_k p_k^{n_1} (1-p_k)^{n-n_1} \\ &= \sum_{k=1}^N \mu_k \left(\prod_{t=1}^n p_k(y_t) \right).\end{aligned} \quad (6)$$

All the finite dimensional distributions of (Y_n) are determined by the parameters $\mathbf{p} := (p_1, \ldots, p_N)$ and $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_N)$, which completely specify the finite mixture of i.i.d processes from the probabilistic point of view.

Remark 2: Note that (6) is different from the classic definition of mixture encountered in applied statistics. In the classic definition one assumes that $\mathbb{P}\{Y_1 = y_1\} = \sum_{k=1}^N \mu_k p_k(y_1)$ and that the process (Y_n) is independent. The probability of a trajectory is given by

$$\mathbb{P}\{Y_1^n = y_1^n\} = \prod_{t=1}^n \left(\sum_{k=1}^N \mu_k p_k(y_t) \right),$$

to be compared with (6). A classic mixture is one i.i.d. process, not a mixture of i.i.d. processes.

Next to the widely known de Finetti characterization of general exchangeable processes there is a specialized result characterizing the subclasses of finite and countable mixtures of i.i.d. processes in terms of hidden Markov models (HMM).

Definition 4: The process (Y_n), taking values in a countable set I, is a finite (countable) HHM if there exists a stationary Markov chain (X_n) with values in a finite (countable) state space $\mathcal{X}$ and a function $f : \mathcal{X} \to I$ such that $Y_n = f(X_n)$ for all n.

Note that when defining a HMM, in general one does not require the underlying Markov chain to be stationary, but for the purpose at hand this is a very natural restriction. The characterization of finite and countable mixtures of i.i.d. processes, due to Dharmadhikari [5], is as follows.

Theorem 2: The binary exchangeable process (Y_n) is a finite (countable) mixture of i.i.d. processes if and only if (Y_n) is a finite (countable) HMM.

In [8] and [10], this result has been extended to the class of partially exchangeable processes and to the corresponding subclasses of finite (countable) mixtures of Markov chains.

III. Characterization of finite mixtures

In this section we propose an alternative to Dharmadhikari Theorem 2, stated as Theorem 3 below, to characterize exchangeable processes which are finite mixtures of i.i.d. processes. The fringe benefit of our approach is that it lends itself naturally to the construction of a representation similar to (6), when it exists. To put all this in system theoretic terms we pose the following problem.

Problem 1: (Stochastic realization of binary exchangeable processes) Given the joint distributions $p_Y(y_1^n) := \mathbb{P}\{Y_1^n = y_1^n\}$ of a binary exchangeable process (Y_n), find, when they exist, an integer N (possibly the smallest one) and parameters $(\mathbf{p}, \boldsymbol{\mu})$ such that

$$p_Y(y_1^n) = \sum_{k=1}^N \mu_k p_k^{n_1} (1-p_k)^{n-n_1}. \quad (7)$$

for any binary string y_1^n.
Note that if (Y_n) is a mixture of N i.i.d. processes the parameters $(\mathbf{p}, \boldsymbol{\mu})$ are unique, up to permutations of the indices. This is ensured by the uniqueness of the mixing measure μ in the de Finetti theorem.

In the following Theorem 3 we give a characterization of finite mixtures of i.i.d. processes looking at the ranks of a class of Hankel matrices constructed in terms of only the $p_Y(1^m)$. This is afforded by the following

Proposition 1: The probabilities $p_Y(1^m)$ of the strings 1^m for $m = 1, 2\ldots,$ completely specify the joint distributions of a binary exchangeable process (Y_n). In particular

$$p_Y(1^m 0^n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} p_Y(1^{m+k}). \tag{8}$$

The proof can be read in the second volume of Feller's book [7] (Chapter VII.4). Note that for an exchangeable process the probability of any binary string of given length n only depends on n_1, i.e. the number of 1 it contains. It follows that equation (8) permits to construct the probability of any finite string in terms of the probabilities $p_Y(1^m)$.

A. Hankel matrices

Given any binary process (Y_n), not necessarily exchangeable, one can associate to it a Hankel matrix whose elements are the probabilities of the strings 1^m. Formally, for any $n \in \mathbb{N}$, let $H_n = (h_{ij})_{0\le i,j\le n}$ be the $(n+1)\times(n+1)$ Hankel matrix, with entries $h_{ij} := h_{i+j} = p_Y(1^i 1^j) = p_Y(1^{i+j})$, with the convention that $p_Y(1^0) = 1$:

$$H_n := \begin{pmatrix} 1 & p_Y(1) & \ldots & p_Y(1^n) \\ p_Y(1) & p_Y(11) & \ldots & p_Y(1^{n+1}) \\ p_Y(11) & p_Y(111) & \ldots & p_Y(1^{n+2}) \\ \vdots & \vdots & \vdots & \vdots \\ p_Y(1^n) & p_Y(1^{n+1}) & \ldots & p_Y(1^{2n}) \end{pmatrix}. \tag{9}$$

The semi-infinite matrix H_∞ is defined in the obvious way.

Example 2: Let (Y_n) be a an i.i.d. binary process and let $p = p_Y(1)$. Then, since $p_Y(1^k) = p^k$, we have

$$H_n = \begin{pmatrix} 1 & p & \ldots & p^n \\ p & p^2 & \ldots & p^{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ p^n & p^{n+1} & \ldots & p^{2n} \end{pmatrix} = \begin{pmatrix} 1 \\ p \\ p^2 \\ \vdots \\ p^n \end{pmatrix} \begin{pmatrix} 1 & p & p^2 & \ldots & p^n \end{pmatrix}.$$

B. Characterization of finite mixtures of i.i.d. processes

The theorem below, styled as a theorem of the alternative, characterizes, in terms of the ranks of the Hankel matrices (H_n), the binary exchangeable processes which are finite mixtures of i.i.d. processes and therefore admit a realization as in equation (7).

Theorem 3: Let (H_n) be the Hankel matrices of the binary exchangeable process (Y_n) defined in equation (9), then exactly one of the following two statements holds.

- There exists a finite N such that

$$\operatorname{rank}(H_n) = \begin{cases} n+1 & \text{for } n = 0, \ldots, N-1 \\ N & \text{for } n \ge N \end{cases} \tag{10}$$

- $\operatorname{rank}(H_n) = n + 1$ for all $n \in \mathbf{N}$.

(Y_n) is a finite mixture of i.i.d. processes and admits a realization of the form (7) if and only if (10) holds, and N is the minimal size of the realization (order of p_Y).

Remark 3: It is a striking result that the *algebraic* rank of the Hankel matrices characterizes the existence and minimality of the realizations in a non Gaussian context.

In the remainder of this section we discuss the proof of the characterization part of the Theorem. The related realization algorithm will be presented in Section IV.

In order to characterize the exchangeable processes which are finite mixtures of i.i.d. processes one has to be able to characterize the de Finetti measures μ which are concentrated on a finite set of points, (see Theorem 1 and the ensuing Definition 3). As it will be explained below this is directly related to some properties of the moments of the measure μ.

Let μ be a probability measure on $[0, 1]$. The m-th moment α_m of μ is defined as

$$\alpha_m := \int_0^1 x^m \mu(dx). \tag{11}$$

Assume that the moments α_m are all finite and define the moments matrices and their determinants as

$$M_n := \begin{pmatrix} \alpha_0 & \alpha_1 & \ldots & \alpha_n \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_n & \ldots & \ldots & \alpha_{2n} \end{pmatrix} \text{ and } d_n := \det(M_n). \tag{12}$$

In particular, for the de Finetti measure μ of an exchangeable process, the moments are given by

$$\alpha_m = \int_0^1 p^m d\mu(p) = p_Y(1^m). \tag{13}$$

The moments matrices of μ and the Hankel matrices defined in equation (9) coincide for all $n \in \mathbb{N}$, i.e.

$$M_n = H_n.$$

Recall now the definition of point of increase of a real valued function.

Definition 5: Given a real valued function f, we say that x_0 is a *point of increase* of f if $f(x_0 + h) > f(x_0 - h)$ for all $h > 0$.

To determine whether the de Finetti measure μ is concentrated it is enough to study the set of points of increase of its distribution function,

$$F^\mu(x) := \mu([0, x]) \text{ for any } x \in [0, 1]. \tag{14}$$

It is easy to prove that F^μ has exactly N points of increase $p_1, \ldots, p_N$ if and only if μ is concentrated on $p_1, \ldots, p_N$.

To finish the characterization, adapting an argument in Cramer [2], Chapter 12.6, one can prove the following

Lemma 1: If F^μ has N points of increase, then $d_n \neq 0$ for $n = 0, \ldots, N-1$ and $d_n = 0$ for $n \ge N$. If F^μ has infinitely many points of increase, then $d_n \neq 0$ for all n.

Piecing together these considerations the proof of Theorem 3 is completed.

IV. The realization algorithm

Theorem 3 in the previous section gives a criterion to check whether an exchangeable binary process (Y_n) is a finite mixture of i.i.d. processes. In this section we propose an algorithm to compute the parameters $(\mathbf{p}, \boldsymbol{\mu})$ that identify a minimal realization of a finite mixture of i.i.d. processes. The algorithm is a by-product of Theorem 5 below, which follows from a well known result on Hankel matrices (quoted here in a simplified version, sufficient for our purposes)

Theorem 4 (see [9]): Let $H_\infty = (h_{i+j})_{i,j\geq 0}$ be an infinite Hankel matrix of rank N. Then the entries of H_∞ satisfy a recurrence equation of order N of the form:

$$h_m = a_{N-1}h_{m-1} + a_{N-2}h_{m-2} + \cdots + a_0 h_{m-N} \quad (15)$$

for suitable $a_0, a_1, \ldots, a_{N-1}$ and for all $m \geq N$. Defining the characteristic polynomial

$$q(x) := x^N - a_{N-1}x^{N-1} - a_{N-2}x^{N-2} - \cdots - a_0, \quad (16)$$

and assuming that its roots $p_1, \ldots, p_N$ are distinct, a basis $< v_1, \ldots, v_N >$ of the space of solutions of (15) is given by

$$\begin{aligned} v_1 &:= \begin{pmatrix} 1 & p_1 & p_1^2 & p_1^3 & \ldots\ldots \end{pmatrix}, \\ v_2 &:= \begin{pmatrix} 1 & p_2 & p_2^2 & p_2^3 & \ldots\ldots \end{pmatrix}, \\ &\vdots \\ v_N &:= \begin{pmatrix} 1 & p_N & p_N^2 & p_N^3 & \ldots\ldots \end{pmatrix}. \end{aligned}$$

Thus, for a set of N scalars $\mu_1, \ldots, \mu_N$, the entries h_m of (H_∞) can be written as

$$h_m = \sum_{k=1}^{N} \mu_k p_k^m. \quad (17)$$

Note that this is a general result on Hankel matrices and that (17) need not be a convex combination. When adapted to the Hankel matrix of a finite mixture of i.i.d. processes, the previous result becomes the basis for the construction of the realization. One can prove the following

Theorem 5: Let (Y_n) be a mixture of N i.i.d. binary sequences and μ the associated de Finetti measure, then the rank of the matrix H_∞ associated with (Y_n) is N, the roots of the polynomial $q(x)$ defined in equation (16) are distinct, and the measure μ is concentrated on the roots $p_1, \ldots, p_N$ of $q(x)$. Moreover the scalars $\mu_1, \ldots, \mu_N$ in equation (17) are $\mu_k := \mu(\{p_k\})$ for $k = 1, \ldots, N$.

When the process (Y_n) is a mixture of N i.i.d. processes, i.e. its Hankel matrix H_∞ is of rank N, Theorem 5 allows us to give an algorithm that uses H_N to identify the parameters $(\mathbf{p}, \boldsymbol{\mu})$ in equation (7). The algorithm is presented below.

A. The algorithm

We assume at the outset that (Y_n) is a mixture of N i.i.d. sequences, i.e. its Hankel matrix H_∞ has rank N. In order to find the parameters $(\mathbf{p}, \boldsymbol{\mu})$ we need to determine the polynomial $q(x)$, i.e. find its coefficients $a_0, \ldots, a_{N-1}$. To this end construct the matrix H_N as defined in equation (9), note that $h_1 = p_Y(1), \ldots, h_{2N} = p_Y(1^{2N})$ are given. The matrix H_N is a submatrix of H_∞, and its entries satisfy the recurrence equation (15) in Theorem 4. This gives

$$\begin{aligned} h_N &= a_{N-1}h_{N-1} + a_{N-2}h_{N-2} + \cdots + a_0 h_0 \\ h_{N+1} &= a_{N-1}h_N + a_{N-2}h_{N-1} + \cdots + a_0 h_1 \\ &\ldots\ldots \\ h_{2N-1} &= a_{N-1}h_{2N-2} + a_{N-2}h_{2N-3} + \cdots + a_0 h_{N-1} \end{aligned}$$

Denoting by $h^{(N)} := (h_N, h_{N+1}, \ldots, h_{2N-1})^\top$ (the components of $h^{(N)}$ are the first N elements of the $(N+1)$-th column of H_N) and by $\mathbf{a} := (a_0, a_1, \ldots, a_{N-1})^\top$ the unknown coefficients of $q(x)$, we can write the previous set of equations in matrix form

$$H_{N-1}\mathbf{a} = h^{(N)}. \quad (18)$$

The coefficients $\mathbf{a}$ of $q(x)$ are computed solving the linear system (18). Once $\mathbf{a}$ is determined, find the roots of $q(x)$, getting the points $p_1, \ldots, p_N$ where μ is concentrated.

To find the weights $\mu_1, \ldots, \mu_N$ recall that, by equations (13) and (7), for any m, we have

$$h_m = p_Y(1^m) = \sum_{k=1}^{N} \mu_k p_k^m. \quad (19)$$

Define the matrices

$$V := \begin{pmatrix} 1 & p_1 & p_1^2 & \ldots & p_1^{N-1} \\ 1 & p_2 & p_2^2 & \ldots & p_2^{N-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & p_N & p_N^2 & \ldots & p_N^{N-1} \end{pmatrix} \quad (20)$$

$$W := \begin{pmatrix} \mu_1 & 0 & 0 & \ldots & 0 \\ 0 & \mu_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \mu_N \end{pmatrix},$$

where V is by now known and, under the hypotheses, invertible. Equation (19) implies that

$$H_{N-1} = V^\top W V. \quad (21)$$

To find W, invert (21) to find

$$W = (V^\top)^{-1} H_{N-1} V^{-1}. \quad (22)$$

The weights $\mu_1, \ldots, \mu_N$ are the diagonal elements of W. This completes the construction of the realization.

Algorithm

- Find the coefficients $\mathbf{a}$ of $q(x)$ solving the linear system

$$H_{N-1}\mathbf{a} = h^{(N)}.$$

- Find $\mathbf{p}$ determining the roots of $q(x)$.
- Find $\boldsymbol{\mu}$ defining V as in (20) and computing

$$W = (V^{-1})^\top H_{N-1} V^{-1}.$$

V. Positive realization of linear systems

The results of this paper have a bearing on the theory of deterministic positive linear systems. In Theorem 7 below we identify a class of transfer functions for which the rank of the Hankel matrix corresponding to the impulse response of the system coincides with the order of positive realization.

The realization problem for single input single output deterministic *positive linear system* is as follows. Given a scalar rational function $G(z)$

$$G(z) = \frac{q_{n-1}z^{n-1} + \cdots + q_0}{z^n + p_{n-1}z^{n-1} + \cdots + p_0} = \sum_{k\geq 1} h_k z^{-k}, \quad (23)$$

where (h_k), $k = 1, 2 \ldots$, is the corresponding impulse response, find an N and a triple (A, b, c), where $A \in \mathbb{R}_+^{N\times N}$ and $b, c \in \mathbb{R}_+^N$ such that

$$G(z) = c^T(z\mathbb{I} - A)^{-1}b. \quad (24)$$

The smallest N for which a triple (A, b, c) as above can be found is called the *order* of the positive linear system. The reader is referred to [6] and the references therein for the conditions of existence of a positive realization and for the related realization algorithms. What concerns us here is the characterization of the order. In standard linear systems, where the positivity constraints on (A, b, c) are not enforced, the order of the system coincides with the rank of the Hankel matrix associated with the impulse response (h_k), i.e. the matrix

$$H_\infty := \begin{pmatrix} h_1 & h_2 & h_3 & \dots \\ h_2 & h_3 & h_4 & \dots \\ h_3 & h_4 & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (25)$$

but, as discussed in [1], for positive linear systems the rank of H_∞ is only a lower bound on the order.

Definition 6: The impulse response (h_k) is of the relaxation type if it is a completely monotone sequence (see the definition below), with $h_1 = 1$.

For the notion of linear systems of the relaxation type in continuous time (without positivity constraints on the realization) the reader is referred to e.g. [11]. We are not aware of previous work on discrete time *positive* systems of the relaxation type.

Recall the definition and characterization of completely monotone sequences, see [7] Chapter VII.1 for details. Given a sequence of real numbers (h_k), $k = 1, 2, \ldots$, one defines the differencing operator $\Delta h_k := h_{k+1} - h_k, k = 1, 2, \ldots$ which, if iterated r times, produces the sequence (for $r = 0$ it is the original sequence)

$$\Delta^r h_k = \sum_{\ell=0}^{r} (-1)^\ell \binom{r}{\ell} h_{k+r-\ell}.$$

The sequence (h_k) is *completely monotone* if for all $r = 0, 1, \ldots$ and all $k = 1, 2, \ldots$

$$(-1)^r \Delta^r h_k \geq 0$$

Completely monotone sequences with $h_1 = 1$ are characterized by the classic

Theorem 6 (Hausdorff): The sequence of real numbers $(h_k), k = 1, \ldots$ with $h_1 = 1$ is completely monotone if and only if it coincides with the sequence of moments $h_k := m_{k-1} = \int x^{k-1}\mu(dx)$ of some probability measure μ on $[0, 1]$.

Note that we have shifted the index of the moments only to take into account the strict causality of the system. For the proof of the theorem see e.g. [7], Chapter VII.3.

We are now ready to state the following

Theorem 7: Let $G(z)$ be a rational transfer function such that the corresponding impulse response (h_k) is of the relaxation type. Let H_n, $n = 0, 1, \ldots$, be the $(n+1)\times(n+1)$ principal submatrix of H_∞ defined in equation (25). Assume that

$$\text{rank}(H_n) = \begin{cases} n+1 & \text{for } n = 0, \ldots, N-1 \\ N & \text{for } n \geq N. \end{cases} \quad (26)$$

Then there exist $0 < p_1 < \ldots, < p_N < 1$ and $0 < \mu_1, \ldots, \mu_N < 1$, with $\sum_{i=1}^N \mu_i = 1$, such that, for $k = 1, \ldots$

$$h_k = \sum_{i=1}^{N} \mu_i p_i^{k-1}. \quad (27)$$

Moreover

$$A := \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & \dots & \dots & p_N \end{pmatrix}, \; b := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \; c := \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix}$$

is a positive realization of $G(z)$.

The proof of the Theorem follows directly from the results of the previous sections and by the same token the subclass of impulse responses of the relaxation type satisfying assumption (26) is not empty.

VI. Conclusions

The results of the paper can be extended and generalized in many directions. We are currently working on the generalization to partially exchangeable processes. A process is partially exchangeable if its n-dimensional distributions, for all n, are invariant under a smaller class of permutations, specifically those that leave invariant the transition counts between states. This is a much richer class of processes, the recursive equation generating them is of the form

$$Y_{n+1} = f(\bar{Y}, Y_n, \delta_n).$$

Markov chains correspond to the case of a degenerate random variable $\bar{Y}$. By a generalized version of de Finetti theorem, all *recurrent* partially exchangeable processes are mixtures, generally infinite, of Markov chains, [4]. The characterization of finite mixtures of Markov chains becomes the natural problem to pose in this context. Partial results are already available. More precisely, we have a realizability condition and we can give a bound on the order of the realization looking at the ranks of two families of Hankel matrices, but we still do not have an exact realization algorithm, for more details see [10]. Extension of the results to processes (Y_n) with more general state spaces, from binary to finite, is also under way. The same type of technical difficulties are encountered in both directions as the natural tool to attack both cases is the Hausdorff moment problem on the unit hypercube.

REFERENCES

[1] L. Benvenuti and L. Farina, A Note on Minimality of Positive Realizations, *IEEE Transactions of Circuits and Systems I*, vol. 45, 1998, pp. 676-677. Correction, *ibidem*, vol. 53, 2006, pp. 215.

[2] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ; 1946.

[3] B. de Finetti, La Prévision: ses Lois Logiques, ses Sources Subjectives, *Annales de l'Institut Henri Poincaré*, vol. 7, 1937, pp. 1-68.

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

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

[6] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications* Wiley, New York, NY; 2000.

[7] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2, Second edn., Wiley, New York, NY; 1971.

[8] L. Finesso and C. Prosdocimi, Partially Exchangeable Hidden Markov Models, *Proceedings of the European Control Conference, ECC'09*, Budapest, Hungary, 2009, pp. 3910-3914.

[9] F.R. Gantmacher, *The Theory of Matrices*, vol. 2, Chelsea, New York, NY; 1959.

[10] C. Prosdocimi, Partial Exchangeability and Change Detection for Hidden Markov Models, PhD thesis, University of Padova, 2010.

[11] J.C. Willems, Realization of Systems with Internal Passivity and Symmetry Constraints, *Journal of the Franklin Institute*, vol. 301, 1976, pp. 605-621.