Exchangeable Processes: de Finetti Theorem Revisited

Ehud Lehrer* and Dimitry Shaiderman†

January 16, 2018

Abstract:

A sequence of random variables is exchangeable if the joint distribution of any finite subsequence is invariant to permutations. de Finetti Theorem states that every exchangeable infinite sequence is a convex combination of i.i.d. processes. In this paper we explore the relationship between exchangeability and frequency-dependent posteriors. We show that any real-valued stationary process is exchangeable if and only if its posteriors depend only on the empirical frequency of past events.

Classification numbers: 60A05, 60A10, 91A26

Keywords: Exchangeable process; de Finetti Theorem; Stationary process; Permutations-invariant posteriors; Frequency-dependent posteriors; Urn Schemes.

*Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Bd. de Constance, 77305 Fontainebleau Cedex, France. e-mail: lehrer@post.tau.ac.il. Lehrer acknowledges the support of the Israel Science Foundation, Grant 963/15.

†Tel Aviv University, Tel Aviv 69978, Israel. e-mail: dimitry@tau.ac.il.
1 Introduction

A sequence \( (X_n)_{n=1}^\infty \) of random variables is said to be \textbf{exchangeable} if the distribution of any finite cylinder is invariant to permutations, i.e.,

\[
(X_1, \ldots, X_m) \overset{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(m)})
\]

for every \( m \in \mathbb{N} \) and \( \pi \in S_m \) (the set of permutations over \( \{1, \ldots, m\} \)). The notion of exchangeability turned out to be an important cornerstone in the justification of Bayesian statistics. This fact is due to the remarkable representation theorem of de Finetti \[1\]. This theorem states that if a process is exchangeable, then and only then, it is a convex combination of i.i.d. processes. That is, an exchangeable process is an i.i.d. process with unknown parameters selected according to a distribution over the set of parameters. Kreps \[8\] summarizes the importance of this result as follows:

"...de Finetti’s theorem, which is, in my opinion, the fundamental theorem of statistical inference – the theorem that from a subjectivist point of view makes sense out of most statistical procedures."

Through the years there has been an effort to find other characterizations of exchangeability. A famous result is that of Kallenberg \[5\]. This result states that a stationary process is exchangeable if and only if its past samples affect the distribution of any future observation in the same manner, i.e.,

\[
X_{n+1} \mid X_1, \ldots, X_n \overset{d}{=} X_{n+k} \mid X_1, \ldots, X_n \quad \text{for every } n \geq 0, k \geq 1.
\]

Our interest in exchangeability arose in a recent work of Lehrer and Teper \[7\] who studied the relations between decision making theory and Bayesian updating. Consider a decision maker who updates her preferences. Suppose that after two histories that share the same empirical distributions she has the same preferences. The question arises as to whether a stationary process whose posteriors after any two histories that share the same frequency coincide is necessarily exchangeable.

In section \[3\] we deal with discrete valued stationary sequences. We introduce two properties, called permutation invariant posteriors (PIP) and positivity, and show that any discrete valued stationary process is exchangeable if and only if it satisfies these properties.

We then move on in section \[4\] to deal with real valued random processes. We give a natural generalization to the PIP property to the continuous case by defining the strong
PIP property. As it turns out, this property together with stationary are necessary and sufficient to the deduction of exchangeability.

2 Definitions and Existing Results

Let \((X_n)_{n=1}^{\infty}\) be a sequence of random variables defined over the probability space \((\Omega, \mathcal{F}, \Pr)\) taking values in \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\).

**Definition 1.** The random sequence \((X_n)_{n=1}^{\infty}\) is said to be exchangeable if the distribution of finite cylinders, is invariant under permutations of coordinates, i.e:

\[
(X_1, ..., X_m) \overset{d}{=} (X_{\pi(1)}, ..., X_{\pi(m)})
\]

for every \(m \in \mathbb{N}\) and \(\pi \in S_m\).

The seminal de Finetti Theorem [1] is the following:

**Theorem (de Finetti).** A binary sequence \((X_n)_{n=1}^{\infty}\) is exchangeable if and only if there exists a distribution function \(F\) on \([0,1]\) such that for all \(n \in \mathbb{N}\):

\[
\Pr(X_1 = x_1, ..., X_n = x_n) = \int_{0}^{1} \theta^{n\bar{x}_n} \cdot (1 - \theta)^{n-n\bar{x}_n} dF(\theta)
\]

where \(\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i\) is the sample mean. The sequence \((X_n)_{n=1}^{\infty}\) is said to have de-Finetti measure \(F\).


**Theorem (Hewitt and Savage).** Let \((X_n)_{n=1}^{\infty}\) be a sequence of exchangeable random variables taking values in some complete metric space \(\mathcal{X}\). Then there exists a unique probability measure \(\nu\) on the set of probability measures \(\mathcal{P}(\mathcal{X})\) on \(\mathcal{X}\), such that for any Borel sets \(\{A_i\}\) in \(\mathcal{X}\):

\[
\Pr(X_1 \in A_1, ..., X_n \in A_n) = \int Q(A_1) \cdots Q(A_n) \nu(dQ)
\]

When a random sequence satisfies the conclusion of the theorem, it is said to have de Finetti measure \(\nu\).
Definition 2. The random sequence \( (X_n)_{n=1}^{\infty} \) is said to be stationary if its distribution is invariant under time shifts, that is: \( \forall k > 0 \) and \( A \in \mathcal{B}(\mathbb{R}^\infty) \) we have:
\[
\Pr((X_1, X_2, \ldots) \in A) = \Pr((X_{k}, X_{k+1}, \ldots) \in A)
\]

The next definition was introduced by Berti et. al. [4], following the results of Kallenberg [5]:

Definition 3. The random sequence \( (X_n)_{n=1}^{\infty} \) is said to be conditionally identically distributed if:
\[
X_{n+1} \mid X_1, \ldots, X_n \overset{d}{=} X_{n+k} \mid X_1, \ldots, X_n \text{ for every } n \geq 0, k \geq 1
\]

That is, past events, effect the distribution of all the future steps in the sequence in the same manner.

The following characterization is due to Kallenberg [5].

Theorem (Kallenberg). A sequence \( X_1, X_2, \ldots \) of random variables is exchangeable iff it is stationary and conditionally identically distributed.

Recently, Lehrer and Teper [7] dealt with Bayesian updating in a dynamic decision making setup. The question regarding the relation between stationarity and exchangeability rose again. They posed the question whether a discrete valued stationary stochastic process is necessarily exchangeable whenever any two posteriors that follow histories sharing the same frequency coincide. With the help of an additional property, we provide an affirmative answer.

3 The Main Result - Discrete Case

Let \( (X_n)_{n=1}^{\infty} \) be a discrete valued random sequence defined on the probability space \( (\Omega, \mathcal{G}, \Pr) \) taking values in \( \Gamma^\infty := \Gamma \times \Gamma \times \ldots \), where \( \Gamma \subseteq \mathbb{R} \) is a countable set. Define \( \mathcal{H}_n = \{(h_1, \ldots, h_n); h_i \in \Gamma, \forall i \leq n\} \) be the set of all finite histories of length \( n \), and set \( \mathcal{H} = \bigcup_n \mathcal{H}_n \).

Let \( \mu \) be the probability measure induced on \( (\Gamma^\infty, \sigma(\mathcal{H})) \) by \( (X_n)_{n=1}^{\infty} \), that is:
\[
\mu((h_1, \ldots, h_n)) = \Pr(X_1 = h_1, \ldots, X_n = h_n), \forall n \in \mathbb{N}, (h_1, \ldots, h_n) \in \mathcal{H}_n.
\]

We shall now define the Permutation Invariant Posteriors (PIP) property that captures the notion that any two posteriors that follow histories sharing the same frequency coincide.

\( \sigma(\mathcal{H}) \) is the \( \sigma \)-algebra generated by \( \mathcal{H} \).
3.1 The PIP property

For each \( n \in \mathbb{N} \) define the set:

\[
\mathcal{H}_n^+ := \{(h_1, \ldots, h_n) \in \mathcal{H}_n; \mu((h_{\pi(1)}, \ldots, h_{\pi(n)})) > 0, \forall \pi \in S_n\}
\]

The set \( \mathcal{H}_n^+ \) contains all positive probability finite histories of length \( n \), with the property that all finite histories with the same empirical frequency are feasible as well, i.e., of positive probability. We next introduce the new property which plays a major role in our characterization.

**Definition 4.** A stationary discrete valued random sequence \((X_n)_{n=1}^{\infty}\) satisfies Permutation Invariant Posteriors (PIP) if \( \forall m \in \mathbb{N}, \forall (h_1, \ldots, h_m) \in \mathcal{H}_m^+ \) and \( \forall \pi \in S_m\):

\[
X_{m+1} \mid (h_1, \ldots, h_m) \overset{d}{=} X_{m+1} \mid (h_{\pi(1)}, \ldots, h_{\pi(m)}).
\]

The property PIP guarantees that the distribution of the random variable that follows a finite sequence of outcomes (i.e., the posterior) depends only on their empirical frequency, and not on their order.

The main interest of our work lies in the relation between exchangeable processes and ones that satisfy the stationarity and PIP conditions. Notice that by de Finetti Theorem each exchangeable sequence must be stationary and satisfy the PIP condition. As for the inverse, we first show that stationary - PIP sequences need not be exchangeable.

**Example 1.** Define the random sequence \((X_n)_{n=1}^{\infty}\) inductively by:

\[
\begin{cases}
X_1 = U(\{0, 1, 2\}) \\
X_n = (X_{n-1} + 1) \pmod{3}.
\end{cases}
\]

Notice that since \( \mathcal{H}_m^+ = \emptyset, \forall m \geq 2 \), the PIP condition follows. Stationarity follows too as the outcome of the sequence is determined solely based on \( X_1 \), having uniform distribution. Nevertheless, Exchangeability does not hold as,

\[
\Pr(X_1 = 0, X_2 = 1, X_3 = 2) = \frac{1}{6} \neq 0 = \Pr(X_1 = 1, X_2 = 0, X_3 = 2).
\]

From now on we will focus on stationary and PIP processes where positive probability is retained under all permutations. The latter property is explored in the following subsection.
3.2 The Positivity property

Definition 5. We say that a discrete valued random sequence \((X_n)_{n=1}^\infty\) has the positivity property if \(\forall m \in \mathbb{N}\) and for each \(h_1, \ldots, h_m) \in H_m:\)

\[
\mu((h_1, \ldots, h_m)) > 0 \implies \mu((h_{\pi(1)}, \ldots, h_{\pi(m)})) > 0, \forall \pi \in \mathbb{S}_m.
\]

I.e., the class of positive-probability finite length histories is invariant under all permutation of the indices. A discrete valued random process is said to be positive if it satisfies the positivity property.

The following example shows that the PIP and positivity properties alone are not sufficient for exchangeability.

Example 2. Our example can be found in [2]. Assume we have an urn with one red ball and two black balls. At each stage we draw a ball and replace it by two balls with the same color as the ball just drawn. The probability drawing a red ball in each stage is the proportion of black balls in the urn at that time. Let \(X_n\) be the color of the ball drawn at stage \(n\). By the construction the random sequence \((X_n)_{n=1}^\infty\) satisfies both PIP and positivity properties. A simple calculation shows that \(\Pr(X_1 = \text{red}, X_2 = \text{black}) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}\) which differs from \(\Pr(X_1 = \text{black}, X_2 = \text{red}) = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}\). Hence, this random sequence is not exchangeable.

To conclude the discussion regarding the insufficiency of a partial set of the discussed properties for exchangeability, let us now demonstrate that stationarity and positivity do not imply exchangeability.

Example 3. Consider a Markov Chain on a two - states space \(\{0, 1\}\) with the following transition matrix:

\[
T = \begin{pmatrix}
0 & 1 \\
\frac{1}{5} & \frac{4}{5} \\
\frac{1}{5} & \frac{4}{5}
\end{pmatrix}
\]

and initial state distribution \(\Pr(X_1 = 0) = 1 - \Pr(X_1 = 1) = \frac{4}{19}\). As the distribution \(v = (\frac{4}{19}, \frac{15}{19})\) satisfies the equation \(v = vT\), we get the \(v\) is a stationary distribution vector of the chain, and thus the resulting random sequence \((X_n)_{n=1}^\infty\) is stationary. By construction
it is also positive. However we have:

\[\Pr(X_1 = 1, X_2 = 0, X_3 = 0) = \frac{15}{19} \cdot \frac{1}{5} \cdot \frac{1}{4} = \frac{3}{19} \cdot \frac{1}{4} \neq \frac{3}{19} \cdot \frac{1}{5} = \frac{4}{19} \cdot \frac{3}{4} \cdot \frac{1}{5} = \Pr(X_1 = 0, X_2 = 1, X_3 = 0).\]

The sequence \((X_n)_{n=1}^{\infty}\) is therefore not exchangeable.

From now on we will deal only with stationary processes that satisfy both positivity and PIP assumptions. Such processes will be referred to as stationary and positive PIP.

### 3.3 Discrete Case - The main theorem

It is clear that any exchangeable process is stationary and positive PIP. Our main theorem establishes the inverse direction.

**Theorem 1.** A discrete valued random sequence is exchangeable iff it stationary and positive PIP.

### 3.4 The Proof

The first step in the proof is to show that a stationary and positive PIP sequence can be extended backwards in time. We then use Levy’s upwards theorem to understand the asymptotic behaviour of such processes. It turns out that the result of the discrete case is important in its own right, as it allows to give a straightforward and implicit criteria for exchangeability in the urn schemes introduced by Hill et. al. [2] (see Subsection 3.5 below). Section 4 extends our results to the continuous case.

**Theorem (Extension Theorem).** Assume \((X_n)_{n=1}^{\infty}\) is a stationary and positive PIP random sequence. There exist random variables \(X_0, X_{-1}, X_{-2}, \ldots\) such that \(\forall n \in \mathbb{N} \cup \{0\}:\)

\[(X_{-n}, X_{-n+1}, \ldots) \overset{d}{=} (X_1, X_2, \ldots).\]

In particular \(\forall n \in \mathbb{N} \cup \{0\}\) the random sequence \((X_{-n}, X_{-n+1}, \ldots)\) is stationary and positive PIP as well.
Proof. This result follows by Lemma 9.12 in [6], which states that a stationary sequence can be extended backwards to a two sided stationary sequence \((X_k)_{k \in \mathbb{Z}}\), with the property that \((X_{-n}, X_{-n+1}, ...) \overset{d}{=} (X_1, X_2, ...)\), \(\forall n \in \mathbb{N} \cup \{0\}\).

From this point up to the end of this section, let \((X_k)_{k \in \mathbb{Z}}\) be the extension, given by the extension theorem, of the stationary and positive PIP discrete valued sequence \((X_n)_{n=1}^\infty\).

We need a few notations. Let \(\Gamma^{-\infty} := \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{\infty} \{X_{-n} = h_i \mid h_i \in \Gamma\}\). \(\Gamma^{-\infty}\) can be thought of as the set of infinite histories occurring in the past. Denote by \(h_1, h_2, ... \) a history. For any \(h \in \Gamma^{-\infty}\) and \(n \in \mathbb{N}\) let \(h^n = \{X_{-n} = h_n, ..., X_{-2} = h_2, X_{-1} = h_1\}\). Finally, \(\mathcal{F}_{-n} := \sigma(X_{-n}, ..., X_{-1})\) is the \(\sigma\)-algebra generated by \(X_{-n}, ..., X_{-1}\) and \(\mathcal{F}_{-\infty} := \sigma\left(\bigcup_{n \geq 1} \mathcal{F}_{-n}\right)\) be the \(\sigma\)-algebra generated by \(\bigcup_{n \geq 1} \mathcal{F}_{-n}\).

**Levy upwards theorem** implies that for Pr - almost every \(h \in \Gamma^{-\infty}\) the sequence \(\Pr(X_0 = \gamma \mid h^n)\) converges as \(n \to \infty\). Denote the limit by \(\Pr(X_0 = \gamma \mid h)\).

The next lemma is a cornerstone of our proof. It supports the intuition that \(\Pr(X_0 = \gamma \mid h)\) is the asymptotic rate of occurrence of the letter \(\gamma\) in the history \(h\). For each \(\gamma \in \Gamma\) define the set \(A_\gamma = \left\{h \in \Gamma^{-\infty} \mid \Pr(X_0 = \gamma \mid h) > 0\right\}\).

**Lemma 1.** Pr - almost every history \(h \in A_\gamma\) contains the letter \(\gamma\) infinitely many times.

**Proof.** For each \(k \in \mathbb{Z}\) define \(C_k = \{X_k = \gamma, X_l \neq \gamma, \forall l \leq k\}\). The sets \((C_k)_{k \in \mathbb{Z}}\) are disjoint. By stationarity all \((C_k)_{k \in \mathbb{Z}}\) have the same probability. Thus,

\[
\Pr\left(\bigcup_{k \in \mathbb{Z}} C_k\right) = 0. \tag{1}
\]

By eq. (1), Pr - almost every \(h \in \Gamma^{-\infty}\) either contains no \(\gamma\) letters, or contains an infinite number of them. Define, \(B_{-1} = \{X_{-n} \neq \gamma, \forall n \in \mathbb{N}\}\). We obtain that Pr - almost every \(h \in \Gamma^{-\infty} \setminus B_{-1}\) contains infinitely many \(\gamma\) letters.

In case \(\Pr(B_{-1}) = 0\) the proof is complete. Otherwise, \(\Pr(B_{-1}) > 0\). It is sufficient to prove that for Pr - almost every \(h \in B_{-1}\),

\[
\Pr(X_0 = \gamma \mid h) = 0. \tag{2}
\]
Assume the contrary. It implies that there exists $D \subseteq B_{-1}$ with $\Pr(D) > 0$ such that for

$$\Pr \text{ -- almost every } h \in D, \Pr(X_0 = \gamma \mid h) > 0.$$  Thus, \(\exists n^* \text{ and } D_{n^*} \subseteq D \text{ with } \Pr(D_{n^*}) > 0\) such that $\Pr(X_0 = \gamma \mid h) > \frac{1}{n^*}$ for $\Pr \text{ -- almost every } h \in D_{n^*}$. By Pr-a.s convergence of $\Pr(X_0 = \gamma \mid h^n)$, there exists $M > 0$ such that the set $D_{n^*, M} := \left\{ h \in D_{n^*}; \Pr(X_0 = \gamma \mid h^n) > \frac{1}{2n^*}, \forall n > M \right\}$ has positive probability. Using stationarity, we get:

$$\Pr(X_0 = \gamma \mid D_{n^*, M}) \Pr(D_{n^*, M}) = \lim_{n \to \infty} \sum_{h^n; h \in D_{n^*, M}} \Pr(X_0 = \gamma \mid h^n) \Pr(h^n)$$

$$\geq \frac{1}{2n^*} \lim_{n \to \infty} \sum_{h^n; h \in D_{n^*, M}} \Pr(h^n)$$

$$= \frac{1}{2n^*} \Pr(D_{n^*, M}),$$

implying $\Pr(X_0 = \gamma \mid D_{n^*, M}) \geq \frac{1}{2n^*}$. We thus obtain,

$$\Pr(X_0 = \gamma \mid B_{-1}) \geq \frac{\Pr(X_0 = \gamma \mid D_{n^*, M}) \Pr(D_{n^*, M})}{\Pr(B_{-1})} \geq \frac{1}{2n^*} \frac{\Pr(D_{n^*, M})}{\Pr(B_{-1})} = 0.$$  (3)

Eq. (3) stands in contradiction with $0 = \Pr(C_0) = \Pr(X_0 = \gamma \mid B_{-1}) \Pr(B_{-1})$, which implies that $\Pr(X_0 = \gamma \mid B_{-1}) = 0$.

**Proposition 1.** For every $\gamma_1, \gamma_2 \in \Gamma$ we have\(^2\)

$$\mathbb{E}\left(\mathbb{1}(X_0 = \gamma_1, X_1 = \gamma_2) \mid \mathcal{F}_{-\infty}\right) = \mathbb{E}\left(\mathbb{1}(X_0 = \gamma_2, X_1 = \gamma_1) \mid \mathcal{F}_{-\infty}\right)$$  (4)

**Proof.** By Levy’s upwards theorem, eq. (4) is equivalent to

$$\lim_{n \to \infty} \mathbb{E}\left(\mathbb{1}(X_0 = \gamma_1, X_1 = \gamma_2) \mid \mathcal{F}_n\right) - \mathbb{E}\left(\mathbb{1}(X_0 = \gamma_2, X_1 = \gamma_1) \mid \mathcal{F}_n\right) = 0$$

which is true if and only if for $\Pr \text{ -- almost every history } h \in \Gamma^{-\infty}$ we have

$$\Pr(X_0 = \gamma_1, X_1 = \gamma_2 \mid h^n) - \Pr(X_0 = \gamma_2, X_1 = \gamma_1 \mid h^n) \xrightarrow{n \to \infty} 0,$$

By the stationarity assumption, the latter holds if and only if for $\Pr \text{ -- almost every history } h \in \Gamma^{-\infty}$ we have

$$\Pr(X_0 = \gamma_1 \mid h^n) \Pr(X_0 = \gamma_2 \mid (h^n, \gamma_1)) - \Pr(X_0 = \gamma_2 \mid h^n) \Pr(X_0 = \gamma_1 \mid (h^n, \gamma_2)) \xrightarrow{n \to \infty} 0.$$

\(^2\mathbb{1}(F)\) stands for the indicator function of the set $F$. 

8
Using the positivity and PIP and properties we get that eq. \((\text{1})\) holds if and only if for \(\Pr - \text{almost every}\) history \(h \in \Gamma^{-\infty}:\)

\[
\Pr(X_0 = \gamma_1 \mid h^n)\Pr(X_0 = \gamma_2 \mid (\gamma_1, h^n)) - \Pr(X_0 = \gamma_2 \mid h^n)\Pr(X_0 = \gamma_1 \mid (\gamma_2, h^n)) \xrightarrow{n \to \infty} 0.
\]

(5)

Consider now the sets \(A_{\gamma_1}\) and \(A_{\gamma_2}\) (recall the notation introduced just before Lemma 1). The above criteria obviously holds on \((A_{\gamma_1} \cup A_{\gamma_2})^c\). For \(A_{\gamma_1} \cap A_{\gamma_2}\), Lemma 1 implies that \(\Pr - \text{almost every}\) \(h \in A_{\gamma_1} \cap A_{\gamma_2}\) contains each of the \(\gamma_i\) letters \((i = 1, 2)\) infinitely many times. Let \(\{n_k\}_{k \geq 1}\) and \(\{n_l\}_{l \geq 1}\) be two subsequences such that \(h_{n_k} = \gamma_1\) and \(h_{n_l} = \gamma_2\). Thus,

\[
\Pr(X_0 = \gamma_2 \mid (\gamma_1, h^{n_k-1})) = \Pr(X_0 = \gamma_2 \mid h^{n_k}) \xrightarrow{k \to \infty} \Pr(X_0 = \gamma_2 \mid h),
\]

and

\[
\Pr(X_0 = \gamma_1 \mid (\gamma_2, h^{n_l-1})) = \Pr(X_0 = \gamma_1 \mid h^{n_l}) \xrightarrow{l \to \infty} \Pr(X_0 = \gamma_1 \mid h).
\]

Since Levy’s upward theorem ensures that the limit is \(\Pr\text{-a.s. unique},\) we get:

\[
\lim_{n \to \infty} \Pr(X_0 = \gamma_1 \mid h^n)\Pr(X_0 = \gamma_2 \mid (\gamma_1, h^n)) = \lim_{k \to \infty} \Pr(X_0 = \gamma_1 \mid h^{n_k-1})\Pr(X_0 = \gamma_2 \mid h^{n_k}) = \Pr(\gamma_1 \mid h)\Pr(\gamma_2 \mid h).
\]

Analogously, \(\lim_{n \to \infty} \Pr(X_0 = \gamma_2 \mid h^n)\Pr(X_0 = \gamma_1 \mid (\gamma_2, h^n)) = \Pr(\gamma_2 \mid h)\Pr(\gamma_1 \mid h),\) establishing eq. \((\text{5})\) for \(\Pr - \text{almost every}\) \(h \in A_{\gamma_1} \cap A_{\gamma_2}\). Finally let us show that the criteria holds for \(\Pr - \text{almost every}\) \(h \in A_{\gamma_1} \cap (A_{\gamma_2})^c\) (the case \(A_{\gamma_2} \cap (A_{\gamma_1})^c\) follows in a similar fashion). By Lemma 1 we have that

\[
\lim_{n \to \infty} \Pr(X_0 = \gamma_2 \mid h^n)\Pr(X_0 = \gamma_1 \mid (h^n, \gamma_2)) = 0
\]

Lemma 1 also implies the existence of a subsequence \(\{n_k\}_{k \geq 1}\) such that \(h_{n_k} = \gamma_1\). Thus by the \(\Pr\text{-a.s. uniqueness of the limit argument we get:}\)

\[
\lim_{n \to \infty} \Pr(X_0 = \gamma_1 \mid h^n)\Pr(X_0 = \gamma_2 \mid (\gamma_1, h^n)) = \lim_{k \to \infty} \Pr(X_0 = \gamma_1 \mid h^{n_k-1})\Pr(X_0 = \gamma_2 \mid h^{n_k}) = \Pr(\gamma_1 \mid h)\Pr(\gamma_2 \mid h) = 0,
\]

which concludes the proof of the proposition.  

The next corollary provides a property that guarantees that a discrete stationary and positive PIP sequence is exchangeable.
Corollary 1. Let \( w = (h_1, ..., h_n) \in \mathcal{H}_n \) with \( \mu(w) > 0 \). Then,
\[
\mu(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2 \mid w) = \mu(X_{n+1} = \gamma_2, X_{n+2} = \gamma_1 \mid w)
\] (6)
for every \( \gamma_1, \gamma_2 \in \Gamma \).

Proof. By stationarity \( \{X_n = h_1, ..., X_1 = h_n\} \in \mathcal{F}_- \subset \mathcal{F}_- \) is of positive probability. Thus, by eq. (4),
\[
\mu_p(w, \gamma_1, \gamma_2) = \mathbf{E}\left( \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2) \mid \mathcal{F}_- \right) \cdot \mathbbm{1}(X_n = h_1, ..., X_1 = h_n)
\]
This implies \( \mu((\gamma_1, \gamma_2) \mid w) = \mu((\gamma_2, \gamma_1) \mid w) \), as desired.

We are now in a position to prove the main result of this section, Theorem 1.

Proof of Theorem 1 A discrete valued exchangeable random sequence is clearly stationary and positive PIP. For the other direction let us first define for each word \( h = (h_1, ..., h_n) \in \mathcal{H}_n \) and permutation \( \pi \in S_n \), \( \pi(h) := (h_{\pi(1)}, ..., h_{\pi(n)}) \). The proof will follow by induction on the length of the words. Let \( n = 2 \). By taking the expected value on eq. (4), we get
\[
\mu_p((w, \gamma_1, \gamma_2)) = \mathbf{E}\left( \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2) \mid \mathcal{F}_- \right) \cdot \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2)
\]
This implies \( \mu((\gamma_1, \gamma_2) \mid w) = \mu((\gamma_2, \gamma_1) \mid w) \), as desired.

We are now in a position to prove the main result of this section, Theorem 1.

Proof of Theorem 1 A discrete valued exchangeable random sequence is clearly stationary and positive PIP. For the other direction let us first define for each word \( h = (h_1, ..., h_n) \in \mathcal{H}_n \) and permutation \( \pi \in S_n \), \( \pi(h) := (h_{\pi(1)}, ..., h_{\pi(n)}) \). The proof will follow by induction on the length of the words. Let \( n = 2 \). By taking the expected value on eq. (4), we get
\[
\mu_p((w, \gamma_1, \gamma_2)) = \mathbf{E}\left( \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2) \mid \mathcal{F}_- \right) \cdot \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2)
\]
This implies \( \mu((\gamma_1, \gamma_2) \mid w) = \mu((\gamma_2, \gamma_1) \mid w) \), as desired.

We are now in a position to prove the main result of this section, Theorem 1.

Proof of Theorem 1 A discrete valued exchangeable random sequence is clearly stationary and positive PIP. For the other direction let us first define for each word \( h = (h_1, ..., h_n) \in \mathcal{H}_n \) and permutation \( \pi \in S_n \), \( \pi(h) := (h_{\pi(1)}, ..., h_{\pi(n)}) \). The proof will follow by induction on the length of the words. Let \( n = 2 \). By taking the expected value on eq. (4), we get
\[
\mu_p((w, \gamma_1, \gamma_2)) = \mathbf{E}\left( \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2) \mid \mathcal{F}_- \right) \cdot \mathbbm{1}(X_{n+1} = \gamma_1, X_{n+2} = \gamma_2)
\]
This implies \( \mu((\gamma_1, \gamma_2) \mid w) = \mu((\gamma_2, \gamma_1) \mid w) \), as desired.

We are now in a position to prove the main result of this section, Theorem 1.
Corollary 1 we get,
\[
\mu(h^n, \gamma_1) = \mu(s^n, \gamma_2) = \mu(\gamma_2 | s^n) \mu(s^n) = \mu(\gamma_2 | (w^{n-1}, \gamma_1)) \mu((w^{n-1}, \gamma_1))
\]
\[
= \mu((\gamma_1, \gamma_2) | w^{n-1}) \mu(w^{n-1})
\]
\[
= \mu((\gamma_2, \gamma_1) | w^{n-1}) \mu(w^{n-1})
\]
\[
= \mu((w^{n-1}, \gamma_2, \gamma_1))
\]
\[
= \mu(h^n, \gamma_1),
\]
where the last equality is due to eq. (7). For events \((A_1 \times \ldots \times A_m) \subseteq \Gamma^m\) of positive probability and \(\pi \in \mathbb{S}_m\) we have:
\[
\mu((A_1 \times \ldots \times A_m)) = \sum_{s_1 \in A_1} \ldots \sum_{s_m \in A_m} \mu((s_1, \ldots, s_m))
\]
\[
= \sum_{s_{\pi(1)} \in A_{\pi(1)}} \ldots \sum_{s_{\pi(m)} \in A_{\pi(m)}} \mu((s_{\pi(1)}, \ldots, s_{\pi(m)}))
\]
\[
= \mu((A_{\pi(1)} \times \ldots \times A_{\pi(m)})),
\]
as desired.

3.5 Application to Urn Schemes

Bruce et. al. [2] tried to give a characterization to exchangeable urn schemes in terms of their de Finetti measures. They defined the notion of generalized t-color urn schemes as follows:

Definition 6. Suppose \(Y = \{Y_1, Y_2, \ldots\}\) is a sequence of random variables with possible values in \(\{1, \ldots, t\}\). The sequence \(Y\) is a generalized t-color urn scheme (each ball drawn from the urn is replaced by two balls of the same color) if for \(n = 1, 2, \ldots\) and \(1 \leq j \leq t\) the conditional probability \(\Pr(Y_{n+1} = j | Y_1 = c_1, \ldots, Y_n = c_n)\) that the next ball drawn is of color \(j\) depends only on the proportion of balls of each color at stage \(n\). That is,
\[
\Pr(Y_{n+1} = j | Y_1 = c_1, \ldots, Y_n = c_n) = f_j \left( \frac{r_1 + n_1}{m + n}, \frac{r_2 + n_2}{m + n}, \ldots, \frac{r_t + n_t}{m + n} \right),
\]
where \(r_i\) is the number of balls of color \(i\) in the urn initially, \(n_i\) the number of \(c_j\)'s equal to \(i\), and \(m = \sum_{i=1}^t r_i\) the total number of balls in the urn initially.
In their work Bruce et. al. [2] found a characterization for exchangeable generalized t-color urn schemes with $t = 2$ but were unable to generalize it for $t \geq 3$. Eq. (8) clearly ensures that generalized t-color urn schemes satisfy the PIP property. A straightforward implementation of Theorem 1 thus yields the following observation.

**Proposition 2.** A generalized stationary t-color urn scheme which satisfies the positivity property is exchangeable.

4 The Continuous Case

In this section we will deal with continuous real valued random sequences $(X_n)_{n=1}^\infty$ defined on a probability space $(\Omega, \mathcal{F}, \Pr)$. Let $\mu$ be the probability measure induced on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ by $(X_n)_{n=1}^\infty$. We will define in subsection 4.1 the strong PIP property for continuous processes. As it turns out the strong PIP property together with stationarity ensures exchangeability, see Theorem 2. Subsection 4.3 is devoted to the proof of the new criteria given in Theorem 2.

4.1 The strong PIP property

We wish to define a continuous variant of the PIP property in order to get a similar type characterization of exchangeable processes to that in Theorem 1. A natural way to translate the PIP property to the continuous case is the following:

Define first the sets:

$$O_m^+ := \{(A_1 \times \ldots \times A_m) \in \mathcal{B}(\mathbb{R}^m); \mu(A_{\pi(1)} \times \ldots \times A_{\pi(m)}) > 0, \forall \pi \in S_m\}.$$

The set $O_m^+$ contains all the events of the form $(A_1 \times \ldots \times A_m) \in \mathcal{B}(\mathbb{R}^m)$ that keep having positive probability under any permutation of the coordinates.

**Definition 7.** A stationary random sequence $(X_n)_{n=1}^\infty$ satisfies the Strong Permutation Invariant Posteriors property (strong PIP) if $\forall m \in \mathbb{N}$, $\forall (A_1 \times \ldots \times A_m) \in O_m^+$ and $\forall \pi \in S_m$:

$$\mu(X_{n+1} \in A | A_1 \times \ldots \times A_m) = \mu\left(X_{m+1} \in A | A_{\pi(1)} \times \ldots \times A_{\pi(m)}\right), \forall A \in \mathcal{B}(\mathbb{R}).$$

Notice that in the case of discrete-valued processes the strong PIP property implies the PIP property. As for the other direction notice that the stationary PIP process in Example
1 does not satisfy the strong PIP property. The reason is that the set \( \{0, 1, 2\}^{m-1} \times \{2\} \) is a member of \( O^+_m \), but

\[
\mu \left( X_{m+1} = 0 \mid \{0, 1, 2\}^{m-1} \times \{2\} \right) = 1 \neq 0 = \mu \left( X_{m+1} = 0 \mid \{0, 1, 2\}^{m-2} \times \{2\} \times \{0, 1, 2\} \right).
\]

Hence, the strong PIP property was given its name for a reason.

### 4.2 The Continuous Case - The main theorem

**Theorem 2.** A real valued random sequence is exchangeable iff it is stationary and satisfies the strong PIP property.

### 4.3 The Proof

We start the proof of Theorem 2 with the following technical lemma.

**Lemma 2.** Assume \( (X_n)_{n=1}^{\infty} \) is a discrete-valued stationary sequence. Let \( w = (h_1, ..., h_n) \in \mathcal{H}_n \) with \( \mu(w) > 0 \). Then, \( \forall i \in \{1, ..., n\} \) there exists \( k^* \in \mathbb{N} \) and a word \( s = (h_1, ..., h_n, ..., h_i) \in \mathcal{H}_{n+k^*} \) with \( \mu(s) > 0 \) (i.e. \( \mu(X_1 = h_1, X_2 = h_2, ..., X_n = h_n, X_{n+k^*} = h_i) > 0 \)).

**Proof.** For each \( m \in \mathbb{N} \) define \( C_m = \{X_m = h_i, X_{\ell} \not= h_i, \forall \ell \geq m\} \). The sets \( (C_m)_{m \in \mathbb{N}} \) are disjoint. By stationarity all \( (C_m)_{m \in \mathbb{N}} \) have the same probability. Thus,

\[
\Pr \left( \bigcup_{m \in \mathbb{N}} C_m \right) = 0. \tag{9}
\]

Eq. (9) implies that with probability 1 each word contains either infinite number of \( h_i \)'s or none. Define \(^3\)

\[
N_1 = w \cap \{X_{n+1} = h_i\} \\
N_k = w \cap \{X_{n+1} \neq h_i\} \cap ... \cap \{X_{n+k-1} \neq h_i\} \cap \{X_{n+k} = h_i\}, \quad k \geq 2
\]

Since \( w \) contains the letter \( h_i \) and the sets \( \{N_k\}_{k \in \mathbb{N}} \) are disjoint we have:

\[
0 < \mu(w) = \mu \left( \bigcup_{k \in \mathbb{N}} N_k \right) = \sum_{k=1}^{\infty} \mu(N_k).
\]

\footnote{We abuse notation: \( w \) stands also for the event \( \{X_1 = h_1, ..., X_n = h_n\} \).}
Hence there exists $k^*$ such that $\mu(N_{k^*}) > 0$. The latter implies the existence of $s \in \mathcal{H}_{n+k^*}$ of the form $s = (h_1, ..., h_n, ..., h_i)$ with $\mu(s) > 0$.

The next lemma gives an insight into the special structure of stationary and strong PIP sequences.

**Lemma 3.** Assume that $(X_n)_{n=1}^\infty$ is a discrete-valued stationary strong PIP sequence. Let $w = (w_1, ..., w_n) \in \mathcal{H}_n$ with $\mu(w_{\pi(1)}, ..., w_{\pi(n)}) > 0$, $\forall \pi \in \mathcal{S}_n$. Then $w \times \Omega^m \in O_{n+m}^+$ for all $m \in \mathbb{N}$.

**Proof.** The proof is by induction on $n$. Let $n = 1$. If $\mu(w_1) > 0$, stationarity implies $w_1 \times \Omega^m \in O_{1+m}^+$ for all $m \in \mathbb{N}$.

**The induction step:** Assume by contradiction that there exists a permutation $\{i_1, ..., i_n\}$ of $\{1, ..., n\}$ and non-negative integers $j_1, ..., j_{n+1}$ such that,

$$\mu \left( \Omega^{j_1} \times w_{i_1} \times \Omega^{j_2} \times \cdots \times \Omega^{j_n} \times w_{i_n} \times \Omega^{j_{n+1}} \right) = 0,$$

so that if $j_k = 0$ for some $k \leq n$, then $w_{i_{k-1}} \times \Omega^{j_k} \times w_{i_k} = w_{i_{k-1}} \times w_{i_k}$. Stationarity together with eq. (10) implies:

$$\mu \left( w_{i_1} \times \Omega^{j_2} \times \cdots \times \Omega^{j_n} \times w_{i_n} \right) = 0. \quad (11)$$

Since $\mu(w_{\pi(1)}, ..., w_{\pi(n)}) > 0$, $\forall \pi \in \mathcal{S}_n$, stationarity implies $\mu(w_{\rho(i_1)}, ..., w_{\rho(i_{n-1})}) > 0$, $\forall \rho \in \mathcal{S}_{n-1}$. Thus by the induction step $w_{i_1} \times \Omega^{j_2} \times \cdots \times w_{i_{n-1}} \times \Omega^{j_n} \in O_{n-1+j_2+\cdots+j_n}^+$. By eq. (11), the strong PIP property and stationarity we get:

$$0 = \mu \left( w_{i_n} \mid w_{i_1} \times \Omega^{j_2} \times \cdots \times w_{i_{n-1}} \times \Omega^{j_n} \right)$$

$$= \mu \left( w_{i_n} \mid \Omega^{j_2+\cdots+j_n} \times w_{i_1} \times \cdots \times w_{i_{n-1}} \right)$$

$$= \frac{\mu(w_{i_1}, ..., w_{i_{n-1}}, w_{i_n})}{\mu(w_{i_1}, ..., w_{i_{n-1}})},$$

implying $\mu(w_{i_1}, ..., w_{i_{n-1}}, w_{i_n}) = 0$, thus contradicting the assumption of the lemma.

**Lemma 4.** Assume that $(X_n)_{n=1}^\infty$ is a discrete-valued stationary strong PIP sequence. Let $h = (h_1, ..., h_n) \in \mathcal{H}_n$ with $\mu(h) > 0$. Then, for every permutation $\{i_1, ..., i_n\}$ we have:

$$\mu \left( h_{i_1}, ..., h_{i_n} \right) > 0. \quad (12)$$
Proof. The proof is by induction on \( n \). The case where \( n = 1 \) follows trivially.

The induction step: let \( \{i_1, \ldots, i_n\} \) be a permutation of \( \{1, \ldots, n\} \). An iterative use of Lemma 2 guarantees that there exists \( m \in \mathbb{N} \) and a word \( s \in \mathcal{H}_m \) of the form:

\[
s = (h_1, \ldots, h_n, \ldots, h_{i_1}, \ldots, h_{i_2}, \ldots, h_{i_{n-1}}, \ldots, h_{i_n}),
\]

such that \( \mu(s) > 0 \). Hence there exists \( w \in \mathcal{H}_{m-n} \) with \( w = (h_{i_1}, \ldots, h_{i_2}, \ldots, h_{i_{n-1}}, \ldots, h_{i_n}) \) and \( \mu(w) > 0 \). This implies the existence of numbers \( \{j_p\}_{p=1}^{n-1} \geq 0 \) with \( \sum_{p=1}^{n-1} j_p = m - 2n \) such that

\[
\mu(h_{i_1} \times \Omega^{j_1} \times \ldots \times \Omega^{j_{n-1}} \times h_{i_n}) > 0. \tag{13}
\]

By the induction step we have that \( \mu(h_{\rho(i_1)}, \ldots, h_{\rho(i_{n-1})}) > 0, \forall \rho \in \mathbb{S}_{n-1} \), which by Lemma 3 implies \( h_{i_1} \times \Omega^{j_1} \times \ldots \times h_{i_{n-1}} \times \Omega^{j_{n-1}} \in O_{m-n}^+ \). Thus, by eq. (13), the strong PIP property and stationarity we get:

\[
0 < \mu(h_{i_1} | h_{i_1} \times \Omega^{j_1} \times \ldots \times h_{i_{n-1}} \times \Omega^{j_{n-1}}) = \mu(h_{i_1} | \Omega^{m-2n} \times h_{i_1} \times \ldots \times h_{i_{n-1}}) = \frac{\mu(h_{i_1}, \ldots, h_{i_{n-1}}, h_{i_n})}{\mu(h_{i_1}, \ldots, h_{i_{n-1}})},
\]

which implies \( \mu(h_{i_1}, \ldots, h_{i_{n-1}}, h_{i_n}) > 0 \). This completes the proof.

Proposition 3. A discrete-valued stationary process \( (X_n)_{n \in \mathbb{N}} \) is exchangeable iff it satisfies the strong PIP property.

Proof. Assume \( (X_n)_{n \in \mathbb{N}} \) satisfies the strong PIP property. By Lemma 4 it satisfies the positivity property. Thus, Theorem 1 implies that it is exchangeable. The inverse direction follows easily from de Finetti Theorem.

We are now in a position to prove Theorem 2.

Proof of Theorem 2. As in the discrete case it follows from the generalized de Finetti representation theorem that real-valued exchangeable random sequences are stationary and strong PIP. For the other direction define \( Z^n_l = \frac{l}{2^n} \) on the event \( \left\{ X_i \in \left(\frac{l-1}{2^n}, \frac{l}{2^n}\right) \right\} \) for any \( l \in \mathbb{Z} \) and \( n \in \mathbb{N} \). The sequence \( (Z^n_l)_{l=1}^\infty \) is a discrete random sequence. Also as \( (X_i)_{i=1}^\infty \) is a stationary and strong PIP sequence, it follows that \( (Z^n_l)_{l=1}^\infty \) is one as well for
every \( n \in \mathbb{N} \). Thus, by Corollary 3 we get that \( (Z^n_i)_{i=1}^{\infty} \) is exchangeable for all \( n \in \mathbb{N} \). The latter yields that the sequence \( (X_i)_{i=1}^{\infty} \) is exchangeable on dyadic cubes. Indeed for every \( m, n \in \mathbb{N} \) and any permutation \( \pi \in \mathbb{S}_m \) we get:

\[
\begin{align*}
\mu\left(\left(\frac{l_1 - 1}{2^n}, \frac{l_1}{2^n}\right], \ldots, \left(\frac{l_m - 1}{2^n}, \frac{l_m}{2^n}\right]\right) &= \Pr\left(Z^n_1 = \frac{l_1}{2^n}, \ldots, Z^n_m = \frac{l_m}{2^n}\right) \\
&= \Pr\left(Z^n_1 = \frac{l_{\pi(1)}}{2^n}, \ldots, Z^n_m = \frac{l_{\pi(m)}}{2^n}\right) \\
&= \mu\left(\left(\frac{l_{\pi(1)} - 1}{2^n}, \frac{l_{\pi(1)}}{2^n}\right], \ldots, \left(\frac{l_{\pi(m)} - 1}{2^n}, \frac{l_{\pi(m)}}{2^n}\right]\right).
\end{align*}
\]  

As the restriction of \( \mu \) to \((\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))\) is a regular measure, and as any open set in \( \mathbb{R}^m \) can be decomposed to a disjoint union of dyadic cubes, we have the following identity:

\[
\mu(B) = \inf \left\{ \sum_{k=0}^{\infty} \mu(Q_k); Q_k \subset \mathbb{R}^m \text{ are disjoint dyadic cubes and } B \subset \bigcup_k Q_k \right\}. \tag{15}
\]

For each \( \pi \in \mathbb{S}_m \) define the map \( \pi(x_1, \ldots, x_m) := (x_{\pi(1)}, \ldots, x_{\pi(m)}) \). By eq. (14) we have that for each dyadic cube \( Q \subset \mathbb{R}^m \):

\[
\mu(Q) = \mu(\pi(Q)), \tag{16}
\]

where \( \pi(Q) = \{\pi(x_1, \ldots, x_m); (x_1, \ldots, x_m) \in Q\} \). Combining eqs. (15) and (16), we get that for every \( \pi \in \mathbb{S}_m \),

\[
\mu \left( (A_1 \times \ldots \times A_m) \right) = \mu \left( (A_{\pi(1)} \times \ldots \times A_{\pi(m)}) \right) = \mu \left( (A_{\pi(1)} \times \ldots \times A_{\pi(m)}) \right),
\]

proving that real-valued stationary and strong - PIP sequences are exchangeable. This completes the proof of Theorem 2. \( \blacksquare \)

References


---

4A dyadic cube \( Q \subset \mathbb{R}^m \) is a set of the form \( \frac{z}{2^n} + \left(0, \frac{1}{2^n}\right]^{m} \) for some \( n \in \mathbb{N}, z \in \mathbb{Z}^m \).


