CHARACTERISATION OF EXCHANGEABLE SEQUENCES THROUGH EMPIRICAL DISTRIBUTIONS

MARTIN BLADT

Abstract. The fact that the empirical distributions of an exchangeable sequence form a reverse-martingale is a well-know result. The converse statement is proved, under the additional assumption of stationarity. A similar reverse-martingale for separately exchangeable matrices is found and marginal characterisations are considered.

1. Introduction

The basic idea behind exchangeability is to remove the independence assumption from an independent and identically distributed (iid) sample, while at the same time keeping the same marginals and the symmetry properties that make these objects so easy to deal with. The symmetric property is easily seen to be equivalent to the notion of invariance under permutation of indices. These distributions have been extensively studied and were first shown in 1930 by Bruno de Finetti, in a special instance in [Fin30a], to be equivalent to the mixing of iid samples, and later in 1937 for a more general case in [Fin30b]. Refer to [Ald85a] for a good probabilistic introduction to the subject of exchangeability. This kind of symmetry plays a role in the philosophical foundation of the Bayesian paradigm.

A random sequence \( \xi = (\xi_1, \xi_2, \ldots) \) in a Borel space \((S, \mathcal{S})\) is said to be exchangeable if

\[
(\xi_{k_1}, \ldots, \xi_{k_m}) \overset{d}{=} (\xi_1, \ldots, \xi_m)
\]

for any collection \( k_1, \ldots, k_m \) of distinct elements of the index set in question. Generally, this set is \( \mathbb{N} \) or a finite subset of it. This notion is easily seen to be equivalent to the classic definition in terms of permutations.

If we further denote by \((\mathcal{F}_n)\) the tail \( \sigma \)-field generated by \( S_k = \sum_{i=1}^{k} \xi_i \), \( k = n, n+1, \ldots \), then by symmetry

\[
E(\xi_k | \mathcal{F}_{n+1})
\]

is almost surely the same random variable for all \( k \leq n + 1 \). In particular, we get

\[
E(n^{-1} S_n | \mathcal{F}_{n+1}) = E((n + 1)^{-1} S_{n+1} | \mathcal{F}_{n+1}) = (n + 1)^{-1} S_{n+1},
\]

such that

\[
\left( \frac{S_n}{n}, \mathcal{F}_n \right)
\]

is a reverse martingale. In fact, if we transform the variables with any bounded and measurable \( f \), we get, by the same arguments, that

\[
\frac{\sum_{i=1}^{n} f(\xi_i)}{n} = \int f \, d\eta_n =: \eta_n f
\]

is a reverse martingale, where \( \eta_n = n^{-1} \sum_{i=1}^{n} \delta_{\xi_i} \) are the empirical distributions of the underlying sequence \( \xi \). Since this is true for any such \( f \) we say that \( \eta_n \) itself is a reverse measure-valued martingale. This is the central argument of the following standard result (see for instance [Kal05]).
Lemma 2. Given a stationary sequence $\xi$ on $\mathbb{N}$, we may construct a stationary extension $\tilde{\xi}$ on $\mathbb{Z}$ such that $(\xi_{-n}, \xi_{-n+1}, \ldots)$ is defined recursively as a function of $\xi$ and some independent iid uniform variables $\vartheta_1, \ldots, \vartheta_n$.

Theorem 3. Let $\eta$ be the empirical distributions of the infinite, stationary sequence $\xi$. If $\eta$ is a reverse, measure-valued martingale, then $\xi$ is exchangeable.

Proof. Let $f_1, \ldots, f_n$ be measurable functions and set
\begin{equation}
\mathcal{T}_k = \sigma(\theta_{k-1} \eta) = \sigma(\eta_k, \theta_k \xi), \quad k \geq 1,
\end{equation}
where $\theta(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$ is the usual shift operator with $\theta_k = \theta^k$ and the second equality follows from the relation
\begin{equation}
k \eta_k = (k-1) \eta_{k-1} + \delta_{\xi_k}, \quad k \geq 1.
\end{equation}
Assume $\eta$ is a reverse, measure-valued martingale, with respect to $\mathcal{T}$. Then
\begin{align*}
E(f_1(\xi_1) \cdots f_n(\xi_n)) &= EE(f_1(\xi_1) \cdots f_n(\xi_n) | \mathcal{T}_2) \\
&= E[E(f_1(\xi_1) f_2(\xi_2) | \mathcal{T}_2) f_3(\xi_3) \cdots f_n(\xi_n)] \\
&= E[E(\eta_1 f_1(2\eta_2 - \eta_1) f_2 | \mathcal{T}_2) f_3(\xi_3) \cdots f_n(\xi_n)] \\
&= 2E[E(\eta_1 f_1 f_2 | \mathcal{T}_2) f_3(\xi_3) \cdots f_n(\xi_n)] \\
&= E[E(\eta_1 f_1 f_2 | \mathcal{T}_2) f_3(\xi_3) \cdots f_n(\xi_n)].
\end{align*}
Note that the second term on the right does not depend on the order of $f_1, f_2$. And the same is true for the first term on the right, since
\begin{equation}
E(\eta_1 f_1 f_2 | \mathcal{T}_2) = \eta_2 f_2 E(\eta_1 f_1 | \mathcal{T}_2) = \eta_2 f_1 f_2.
\end{equation}
We thus have by monotone-class extension that
\begin{equation}
(\xi_1, \xi_2, \ldots) \overset{d}{=} (\xi_2, \xi_1, \xi_3, \xi_4, \ldots).
\end{equation}
Now, extend \( \xi \) by Lemma 2 to a doubly infinite stationary sequence \( \hat{\xi} \), such that 
\[
(\hat{\xi}_{-n}, \hat{\xi}_{-n+1}, \ldots)
\]
is a function of \( \xi \) and some independent iid uniform variables \( \vartheta_1, \ldots, \vartheta_n \). By (5), applying the same function with the same uniform variables on both sides we obtain
\[
(\hat{\xi}_{-n}, \ldots, \hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \ldots) \overset{d}{=} (\xi_{-n}, \ldots, \xi_0, \xi_1, \xi_2, \ldots).
\]
But by stationarity, for any \( m, n \in \mathbb{N} \)
\[
(\hat{\xi}_{-n}, \ldots, \hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_m) \overset{d}{=} (\xi_{-n}, \ldots, \xi_0, \xi_1, \ldots, \xi_{n+m+1}),
\]
so that both relations now give
\[
(\hat{\xi}_{-n}, \ldots, \hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_m) \overset{d}{=} (\xi_{-n}, \ldots, \xi_0, \xi_1, \ldots, \xi_{n+m+1}),
\]
and permuting the \((n + 1)\)-th and \((n + 2)\)-th entries of the above vectors we get, again by (6),
\[
(\xi_1, \ldots, \xi_{n+m+1}) \overset{d}{=} (\hat{\xi}_{-n}, \ldots, \hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_m) \overset{d}{=} (\xi_1, \ldots, \xi_n, \xi_{n+2}, \xi_{n+1}, \xi_{n+3}, \ldots, \xi_{n+m+1}).
\]
Since \( n, m \) were arbitrary, and any general permutation can obtained through these switches in a finite amount of steps, we conclude that \( \xi \) is exchangeable. \( \square \)

**Remark 4.** As mentioned in the introduction, a proof of the above Theorem without the assumption of stationarity was given in [Kal05], using a different proof technique. However the proof appears to not be correct. To be more specific, following Theorem 2.4 of [Kal05], let \( \eta \) be a reverse martingale. Then by (4) for bounded, measurable \( f \) on \( S \),
\[
E(f(\xi_k)|T_k) = k\eta_kf - (k - 1)E(\eta_{k-1}f|T_k) = \eta_kf.
\]
Fix \( n \) and define, for \( k \leq n \),
\[
\zeta_k = \xi_{n-k+1}, \quad \beta_k = \sum_{j \leq k} \delta_{\zeta_j} = n\eta_n - (n - k)\eta_{n-k}, \quad \mathcal{F}_k = T_{n-k}.
\]
\( \zeta \) is \( \mathcal{F} \)-adapted and by (7),
\[
E(f(\zeta_{k+1})|\beta_n, \mathcal{F}_k) = E(f(\xi_{n-k})|T_{n-k}) = \eta_{n-k}f = (n - k)^{-1}(\beta_n - \beta_k)f,
\]
and this last expression is presumably enough to assert that \( \zeta \) is not only marginally a conditional urn-sequence, but actually even an urn sequence in the following sense
\[
P(\theta_k\zeta \in \cdot |\beta_n, \mathcal{F}_k) = \frac{(\beta_n - \beta_k)^{(n-k)}(n-k)!}{(n-k)!}, \quad k < n, \quad \beta_k = \sum_{j \leq k} \delta_{\zeta_j}, \quad k \leq n,
\]
yielding exchangeability of \( \zeta \) and consequently of \( \xi \). Here \( \mu^{(n)} \) is the factorial measure on \( S^n \) given by
\[
\mu^{(n)} = \sum_p \delta_{xop},
\]
the sum being over all distinct permutations of \( s \). However, since it is argued that equivalence between marginal and joint urn sequences follows from the following calculation.

Let \( f_i \geq 0 \) be measurable, and consider any sequence \( \xi \) with \( \beta_k = \sum_{j \leq k} \delta_{\xi_j}, k \leq n, \) and satisfying
\[
P(\xi_{k+1} \in \cdot |\beta_n, \mathcal{F}_k) = \frac{\beta_n - \beta_k}{n-k}, \quad k < n,
\]
for a generic filtration $\mathcal{F}$. Then

$$E(f_{k+1}(\xi_{k+1}) \cdots f_n(\xi_n)|\mathcal{F}_k, \beta_n) = E(f_{k+1}(\xi_{k+1}) \cdots f_{n-1}(\xi_{n-1})E(f_n(\xi_n)|\mathcal{F}_{n-1}, \beta_n)|\mathcal{F}_k, \beta_n)$$

$$= \frac{\beta_n - \beta_{n-1}}{n - n + 1} f_n E(f_{k+1}(\xi_{k+1}) \cdots f_{n-1}(\xi_{n-1})|\mathcal{F}_{k+1}, \beta_n)$$

$$= \cdots = \prod_{j=k+1}^n \frac{(\beta_n - \beta_{j-1})}{n - j + 1} f_j$$

$$= \frac{(\beta_n - \beta_k)^{n-k}}{(n-k)!} \prod_{j=k+1}^n f_j$$

and a monotone class extension supposedly yields the equivalence of definitions (10) and (9). But not only does the last equality above not hold in general (try with $k = 0$, $n = 2$), but even in the second equality, measurability of $\beta_{n-1}$ with respect to $\mathcal{F}_k, \beta_n$ was implied, which in general can not be assumed, so the proof of Theorem 2.4 in [Kal05] does not hold along those lines.

It is also worth mentioning that the previous urn-equivalence has repercussions on the proof of Theorem 2.3 in [Kal05] and also of Theorem 2.2 in [Kal05]. The proof of the latter theorem was very easily mendable, along the original lines and hence omitted in the present work.

We now explore a different assumption which leads to the same result, namely the Markovian property. First we show by a counterexample that, contrary to the stationarity assumption, the homogeneous Markov property is quite restrictive, in the sense that it does not have any hope to lead to a characterisation of the entire class of exchangeable sequences.

**Remark 5.** (A non-Markov, exchangeable sequence) The Markov property, for the sequence $\xi$ is defined classically, for an adapted process $X$ on a time scale $T \subseteq \mathbb{R}$, and a filtration $\mathcal{F}$ as the conditional independence relation

$$\mathcal{F}_t \perp \mathcal{X}_u, \ t \leq u \in T.$$ 

For a Markov sequence, if the transition kernels $\mu_n(\cdot|\xi_{n-1}) = P(\xi_n \in \cdot|\xi_{n-1})$ do not depend on $n$, we speak of a homogeneous Markov chain.

All iid sequences are clearly Markov, and so is the sequence $\alpha, \alpha, \ldots$ consisting of a single random variable. In fact, it is also quite easy to see that for any random variable $\alpha$ and an iid sequence of, say, uniform variables $\vartheta_1, \vartheta_2, \ldots$, the tuples

$$(\alpha, \vartheta_1), (\alpha, \vartheta_2), \ldots$$

also form a Markov sequence. Motivated by this fact, and by the well-known functional characterisation of exchangeable sequences which states that any exchangeable sequence can be seen as a function of such tuples, one may be inclined to believe in the following statement: the Markov property is satisfied by exchangeable sequences. However, this is not the case, which boils down to the fact that the transformation of a Markov sequence is not always Markov again. This is seen in the following example.

Consider the random variable $\alpha$ taking values $1, 0, -1$ with equal probability of $\frac{1}{3}$, and an independent sequence of iid $U(0, 1)$ variables $\vartheta_1, \vartheta_2, \ldots$, and define the function

$$f(a, b) = \begin{cases} 
1, & a = -1 \\
0, & a = 0 \\
1, & a = 1, \ b < \frac{1}{2} \\
2, & a = 1, \ b \geq \frac{1}{2}
\end{cases}$$
then, the sequence $\xi_n = f(\alpha, \vartheta_n)$, $n \geq 1$, is conditionally iid, given $\alpha$, such that by the easy implication of de Finetti’s theorem, it is exchangeable. On the other hand

$$P(\xi_1 = \xi_2 = \cdots = \xi_n = 1) = P(\alpha = -1) + P(\alpha = 1, \vartheta_1 \leq \frac{1}{2}, \ldots, \vartheta_n \leq \frac{1}{2})$$

$$= \frac{1}{3} + \frac{1}{3} \left(\frac{1}{2}\right)^n,$$

such that

$$P(\xi_1 = 1 | \xi_2 = \xi_3 = \cdots = \xi_n = 1) = \frac{2^n + 1}{2^{n+2}},$$

which is dependent of $n$, showing that in this case, $\xi$ is not a Markov sequence.

The problem of characterising when a Markov process is again Markov under transformations is not straightforward (see [BR58] in this connection), and will not be addressed here.

**Remark 6.** One can replace the stationarity assumption with the Markov one in Theorem 3 as follows.

**Claim:** Let $\eta$ be the empirical distributions of the infinite, homogeneous Markov sequence $\xi$ in Borel $S$. If $\eta$ is a reverse, measure-valued martingale, then $\xi$ is exchangeable.

**Proof of claim:** By an identical argument as in the first part of the proof of Theorem 3 we can obtain, purely by the reverse-martingale condition, that for $n \geq 2$,

$$(\xi_1, \xi_n, \xi_{n+1}, \ldots) \overset{d}{=} (\xi_n, \xi_1, \xi_{n+1}, \xi_{n+2}, \ldots).$$

In particular, $\xi_1 \overset{d}{=} \xi_n$ for any $n \in \mathbb{N}$. But by homogeneity, $P(\xi_{n+1} \in \cdot | \xi_n) = P(\xi_2 \in \cdot | \xi_1)$. Since the laws of Markov chains are determined uniquely by the initial distribution and transition probabilities, we conclude that

$$\xi = (\xi_1, \xi_2, \ldots) \overset{d}{=} (\xi_{n+1}, \xi_{n+2}, \ldots) = \theta_n \xi, \quad n \in \mathbb{N},$$

i.e., the sequence is stationary. Then we apply Theorem 3 to yield the desired exchangeability of $\xi$.

### 3. Random Matrices

Another way of extending the notion of exchangeability is to consider the multidimensional case, i.e., matrices, or, more generally, arrays. Separately exchangeable, also referred to as row-column exchangeable arrays (see below for the definition) were first considered by [Daw72], in the context of Bayesian statistics, and various initial characterisations were subsequently established independently through different methods by [Ald85b] and [Hoo82]. Some further extensions can be found in [Kal92]. A good compendium of the characterisation results can be found in Chapter 7 of [Kal05], where the author also includes his own contributions to the theory of exchangeable and even rotatable arrays.

Concerning the extension of martingale characterisations to symmetric arrays, work has been done in [IW96] [IW04] [IW05] with respect to the prediction sequences. A necessary reverse-martingale condition is presented within the present chapter, but the reverse implication, as in the one-dimensional case is quite more involved, since in this context stationarity does not have an obvious non-artificial definition other than marginally.

A $d$-dimensional random array is a random element $X : \mathbb{N}^d \to S$. For convenience let $S$ be Borel. We say that it is separately exchangeable if for any permutations $(p_i)$ of $\mathbb{N}$

$$X \circ (p_1, \ldots, p_d) \overset{d}{=} X.$$
From now on, we specialize to the case $d = 2$, and we speak of separately exchangeable matrices. Higher dimensions are treated in the same way with heavier notation. By a reverse, measure-valued martingale on $\mathbb{N}^2$ we mean a collection of random measures $\eta_{(n,m)\in\mathbb{N}^2}$ such that for any $n \geq k$ and $m \geq l$ and any measurable $f$ and we have

$$\eta_{n,m}f = E(\eta_{k,l}f|\theta_{n−1,m−1}\eta), \quad \theta_{s,t}\eta \equiv (\eta_{u,v} : u > s, v > t).$$

**Proposition 7.** (reverse martingale property of empirical distribution) Let $X$ be a finite or infinite random matrix taking values in Borel $S$, with empirical distributions

$$\eta_{n,m} = \sum_{i \leq m} \sum_{j \leq n} \delta_{X_{i,j}}.$$

If $X$ is separately exchangeable, then $\eta$ is a reverse measure-valued martingale.

**Proof.** Set

$$T_{k,l} = \sigma(\theta_{k−1,l−1}\eta) = \sigma(\eta_{k,l}, \theta_{k,l}^*X), \quad k, l \geq 1,$$

with $\theta_{k,l}^*X = X \backslash (X_{i,j})_{i \leq k, j \leq l}$, and where the second equality follows from the relation

$$kl\eta_{k,l} = (k−1)(l−1)\eta_{k−1,l−1} + \sum_{j=1}^{l} \delta_{X_{k,j}} + \sum_{i=1}^{k−1} \delta_{X_{i,j}}, \quad k, l \geq 1.$$  

Suppose $X$ is separately exchangeable. Then $(X_{i,j})_{i \leq n, j \leq m}$ is exchangeable over $T_{n,m}$, as is seen from (14), so for $k \leq n; l \leq m$ and measurable $f \geq 0$ on $S$ we have

$$E(f(X_{1,1})|T_{n,m}) = (kl)^{-1} \sum_{i \leq k; j \leq l} E(f(X_{i,j})|T_{n,m}) = E(\eta_{k,l}f|T_{n,m}).$$

Taking $k = n; l = n$ and $k = n − 1; l = m − 1$ gives the desired reverse martingale property:

$$E(\eta_{n−1,m−1}f|T_{m}) = E(\eta_{n,m}f|T_{m}) = \eta_{n,m}f, \ a.s.$$

\[\square\]

**Conjecture 8.** If (13) forms a reverse measure-valued martingale, then $X$ is a separately exchangeable random matrix.

**Proposition 9.** (Marginal families of reverse martingales) Let $X$ be a finite or infinite random matrix taking values in Borel $S$, with joint empirical distributions

$$\eta^{n}_{m} = \sum_{i \leq m} \frac{\delta_{Y_{i,n}^{m}}}{m}, \quad \eta^{m}_{n} = \sum_{i \leq n} \frac{\delta_{Z_{i,m}^{n}}}{n},$$

where

$$Y_{i}^{n} = (X_{1,j}, \ldots, X_{n,j}), \quad Z_{i}^{m} = (X_{i,1}, \ldots, X_{i,m}),$$

Then $X$ is separately exchangeable if and only if $(\eta^{n})$ and $(\eta^{m})$, $\forall n, m \geq 1$, are reverse measure-valued martingales and $(Y^{n})$, $(Z^{m})$, $\forall n, m \geq 1$ are stationary sequences.

**Proof.** A direct proof is available in the same spirit of Theorems [7] in conjunction with Theorem [3] but instead we give a short proof that builds on the latter two.

Assume $X$ is separately exchangeable. In particular $Y^{n}$ and $Z^{m}$ are exchangeable sequences on the respective Borel spaces $S^{n}$ and $S^{m}$. Then Theorem [4] yields that their empirical distributions $(\eta^{n})$ and $(\eta^{m})$ form reverse, exchangeable, measure-valued martingales. The stationarity is obvious.

Conversely, fixing the sub-array $A^{n,m} = (X_{i,j})_{i \leq n, j \leq m}$, Theorem [3] applied to the exchangeable, reverse, measure-valued martingale $\eta^{n}$ shows, in particular, that the columns
Y^n of An,m are exchangeable. An analogous argument then gives row-exchangeability of An,m, and hence of X.

\[ \Box \]

**Remark 10.** An analogous marginal characterisation result to Proposition 9 is easy to prove using the Markov property in place of stationarity of the sequences (Y^n), (Z^m), n, m ∈ N, and utilising Remark 6 in place of Theorem 8.

A unified, non-marginal characterisation without additional assumptions is an open problem. An interesting and related parametrisation problem for binary exchangeable matrices with additional symmetry (specifically, switch-symmetry) which may call for reverse-martingale methods is a conjecture stated in [Lau03].

**Acknowledgements.** During the author’s stay at Copenhagen University, several fruitful conversations with Steffen L. Lauritzen helped shape the present work.

**References**

[Ald85a] D J Aldous. Exchangeability and related topics. In: École d’Été de Probabilités de Saint-Flour XIII –1983. Lect. Notes in Math., 1117:1–198, 1985.

[Ald85b] D J Aldous. Representations for partially exchangeable arrays of random variables. J. Multivar. Anal., 11:581–598, 1985.

[BR58] C J Burke and M Rosenblatt. A Markovian function of a Markov chain. The Annals of Mathematical Statistics, 29(4):1112–1122, 12 1958.

[Daw72] A P Dawid. Contribution to discussion of Lindley and Smith. Roy. Statist. Soc. Ser. B, 34:29–30, 1972.

[Fin30a] B de Finetti. Fuzione caratteristica de un fenomeno aleatorio. Mem. R. Acc. Lincei, 4:86–133, 1930.

[Fin30b] B de Finetti. La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincaré, 7:1–68, 1930.

[Hoo82] D N Hoover. A normal form theorem for the probability logic L_{ω1p}, with applications. Symbolic Logic, 47, 1982.

[IW96] B G Ivanoff and N C Weber. Some characterizations of partial exchangeability. J. Austral. Math. Soc. (Series A), 61:345–359, 1996.

[IW04] B G Ivanoff and N C Weber. Predictable sampling for partially exchangeable arrays. Statistics and Probability Letters, 70:95–108, 2004.

[IW05] B G Ivanoff and N C Weber. Spreadable arrays and martingale structures. J. Austral. Math. Soc., 79:277–296, 2005.

[Kal92] Olav Kallenberg. Symmetries on random arrays and set-indexed processes. J. Theor. Probab., 5:727–765, 1992.

[Kal02] Olav Kallenberg. Foundations of Modern Probability. Springer, New York, 2nd edition, 2002.

[Kal05] Olav Kallenberg. Probabilistic Symmetries and Invariance Principles. Springer, New York, 2nd edition, 2005.

[Lau03] Steffen L. Lauritzen. Rasch models with exchangeable rows and columns. In J. M. Bernardo et al., editors, Bayesian statistics 7: Proceedings of the seventh valencia international meeting, pages 215–232. Oxford University Press, 2003.

[Loe78] M Loeve. Probability Theory, volume 2. Springer, New York, 4th edition, 1978.