Journal of Theoretical Probability, Vol. 12, No. 3, 1999

Multivariate Sampling and the Estimation Problem for Exchangeable Arrays

Olav Kallenberg[1]

Received March 18, 1998; revised July 30, 1998

We consider random arrays and the associated empirical distributions obtained by multivariate sampling from a stationary process. Under suitable conditions, one gets convergence toward a separately exchangeable array and its ergodic distribution. The result is related to the statistical problem of estimating the representing function of an exchangeable array. The latter problem is well-posed only for shell-measurable arrays, where the grid processes based on finite subarrays form consistent estimates with respect to a suitable norm. In general, the required consistency holds only in the distributional sense for the generated arrays.

KEY WORDS: Stationary processes; exchangeable arrays; functional representations; empirical distributions; limit theorems; consistent estimation.

1. INTRODUCTION

Fix a stationary process X on $\mathbb{R}^d$ and consider an independent collection of random variables τ_k^i, $i = 1,..., d$, $k \in \mathbb{N}$. Writing $\tau = (\tau_k^i)$, we may define a d-dimensional random array $\xi = X \circ \tau$ by

$$\xi_k = \xi_{k_1,..., k_d} = X(\tau_{k_1}^1,..., \tau_{k_d}^d), \qquad k = (k_1,..., k_d) \in \mathbb{N}^d.$$

We say that ξ is obtained from X by *multivariate sampling* and refer to the distribution μ of the array $\tau = (\tau_k^i)$ as the associated *sampling distribution*.

The one-dimensional case was studied in Kallenberg.[12] There we showed that, if the sampling distributions $\mu_1, \mu_2,...$ are *asymptotically*

[1] Department of Mathematics, Auburn University, 218 Parker Hall, Auburn, Alabama 36849-5310. E-mail: olavk@mail.auburn.edu.

0894-9840/99/0700-0859$16.00/0

invariant, in a sense that $\|\mu_n * \delta_t - \mu_n\| \to 0$ as $n \to \infty$ for each t, then the associated sequences $\xi_1, \xi_2,...$ tend in distribution toward an exchangeable sequence ξ, and the corresponding empirical distributions $\eta_n = P[\xi_n \in \cdot \mid X]$ converge in probability to the associated ergodic distribution $\eta = \rho^\infty$, where ρ denotes the directing random measure of ξ. The quoted proposition may be regarded as a simultaneous extension of the mean ergodic theorem and the classical results of de Finetti–Ryll-Nardzewski and Glivenko–Cantelli.

Our original motivation was to look for extensions of the mentioned result to higher dimensions. Despite some efforts, the *general* convergence problem for the arrays ξ_n and the associated empirical distributions η_n remains open. The problem is interesting, because a possible limiting array ξ would have to be *separately exchangeable*, i.e., invariant in distribution under permutations in each index separately, and the measures η_n would converge to the corresponding *ergodic distribution* $\eta = P[\xi \in \cdot \mid \mathscr{I}]$, where $\mathscr{I}$ denotes the shift-invariant σ-field of ξ.

In this paper, the sequences $\tau^i = (\tau_1^i, \tau_2^i,...)$, $i = 1,..., d$, are assumed to be mutually independent. Taking limits in one variable at a time, we may conclude from the one-dimensional result that ξ_n converges toward a separately exchangeable array ξ. (It is not clear, however, whether the distribution of ξ may depend on the order of limits.) In this setting, we may also establish the much deeper convergence of the empirical distributions η_n. Here we rely on a construction by successive randomizations and a rather subtle argument involving conditional independence relations and equivalences between various conditioning σ-fields. Our demonstration occupies the entire Section 2.

The mentioned results are related to a basic statistical problem for exchangeable arrays, first posed as an open problem in (15.6) of Aldous.[2] To explain the background in a simple case, recall from Aldous[1,2] that any two-dimensional, separately exchangeable array $X = (X_{ij})$ admits a representation

$$X_{ij} = f(\alpha, \xi_i, \eta_j, \zeta_{ij}), \qquad i, j \in \mathbb{N} \tag{1.1}$$

in terms of some measurable function f on $[0, 1]^4$ and some random variables α, ξ_i, η_j, and ζ_{ij}, $i, j \in \mathbb{N}$, that are independent and $U(0, 1)$ (uniformly distributed on $[0, 1]$). The general problem is to construct consistent estimates of the representing function f, based on a single realization of X. More precisely, we are asked to find some processes φ_n, measurably dependent on the restrictions $\chi_n X = (X_{ij}; i, j \leqslant n)$, such that $\varphi_n \to f$ in an appropriate sense. Since the entire realization is based on a single value of α, we need to restrict our attention to *ergodic* arrays, where α is a constant and the representation (1) can be simplified to

$$X_{ij} = f(\xi_i, \eta_j, \zeta_{ij}), \qquad i, j \in \mathbb{N} \tag{1.2}$$

for some measurable function f on $[0, 1]^3$.

Still the estimation problem for f turns out not to be well-posed. This is because, as shown in Proposition 3 and Theorem 2, every separately exchangeable array can be approximated arbitrarily closely by so-called *shell-measurable* arrays, which allow representations as in (1.1) or (1.2) without the variables ζ_{ij}. Hence, on the basis of a finite sample, the general arrays in (1.1) or (1.2) are indistinguishable, and we may rather confine our attention to arrays of the form

$$X_{ij} = f(\xi_i, \eta_j), \qquad i, j \in \mathbb{N} \tag{1.3}$$

where f is now a measurable function on $[0, 1]^2$. An array of this type will be said to be *simple*.

There is yet another difficulty, caused by the obvious nonuniqueness of the representing function in (1.1), (1.2), or (1.3). For example, if f is a representing function of the array in (1.3), then so is $\tilde{f}(x, y) \equiv f(T_1 x, T_2 y)$ for any measure-preserving transformations T_1 and T_2 on $[0, 1]$ (though with a possibly different set of random variables ξ_i and η_j). In general, we say that two functions f and g are *equivalent* (written as $f \sim g$) if they can be used to represent the same array X. The best we can hope for is to find some $\chi_n X$-measurable processes φ_n such that $\inf\{\|f - g\|; g \sim \varphi_n\} \to 0$ for a suitable norm $\|\cdot\|$. In this form, the problem for simple arrays is solved by Theorem 4.

A different approach to the estimation problem, applying even to the general arrays in (1.1), is to measure the closeness of two representing functions in terms of the distributions of the generated arrays. Thus, for any measurable functions f and $f_1, f_2,...$ on $[0, 1]^4$, we may construct the associated arrays X and $X_1, X_2,...$ as in (1.1), using an arbitrary collection of i.i.d. $U(0, 1)$ random variables α, ξ_i, η_j, and ζ_{ij}. We can now *define* the convergence $f_n \to f$ by the condition $X_n \xrightarrow{d} X$. If the proposed approximating functions φ_n are random, we need to replace the distribution of each X_n by its conditional version $\eta_n = P[X_n \in \cdot \mid \varphi_n]$. Similarly, if X is nonergodic, the distribution of X should be replaced by its ergodic version $\eta = P[X \in \cdot \mid \mathscr{I}]$. The problem is then to prove that $\eta_n \xrightarrow{w} \eta$ a.s. for a suitable choice of processes φ_n. Here a solution is given by Corollary 2.

The close connection between our two main topics should now be clear. In fact, both Theorem 1 and Corollary 2 deal with empirical distributions of arrays obtained by multivariate sampling from suitable processes, and we get convergence toward the ergodic distribution of some exchangeable array. But there are also essential differences. Thus, in Theorem 1 the underlying process is fixed while the sampling distribution varies with n; the situation in Corollary 2 is the opposite. Now the latter result turns out to be an easy consequence of a general law of large numbers for exchangeable arrays, given in Theorem 3—a result that is precisely of the type of Theorem 1.

In Section 4, the estimation problem will be discussed in the context of so-called *π-exchangeable* arrays of arbitrary dimension d, where π denotes a partition of the set $\{1,..., d\}$. This requires some further discussion of general exchangeable arrays, provided by Section 3. In particular, the basic representation theorem of Aldous[1] and Hoover[5] is extended in Proposition 1 to the general π-exchangeable case.

Readers unacquainted with exchangeability theory may consult Aldous[2] for some general background and motivation. In particular, his Sections 14 and 15 contain an elementary discussion of exchangeable arrays. The latter have been studied extensively by many authors, and additional references include Diaconis and Freedman[4]; Hoover[6]; and Kallenberg.[8–10] Some further information about stationary and exchangeable sequences and processes may be found in Kallenberg,[11] [especially in Chap. 9].

We conclude this section with some remarks on notation. The distribution or *law* of a random element ξ with respect to the basic probability measure P is denoted by $\mathscr{L}(\xi)$. Independence between two random elements ξ and η is expressed as $\xi \perp\!\!\!\perp \eta$, and the corresponding notion of conditional independence, given a σ-field $\mathscr{F}$, is written as $\xi \perp\!\!\!\perp_{\mathscr{F}} \eta$. Conditional probabilities and expectations are often written as $P^{\mathscr{F}} = P[\,\cdot \mid \mathscr{F}\,]$ and $E^{\mathscr{F}} = E[\,\cdot \mid \mathscr{F}\,]$, respectively. Unless otherwise specified, the double bars $\|\cdot\|$ denote the supremum norm or the total variation when applied to functions or signed measures, respectively. A relation $f \lesssim g$ between two positive functions f and g means that $f \leqslant cg$ for some constant $c < \infty$. Finally, we define a *U-array* as an indexed collection of independent $U(0, 1)$ random variables.

2. MULTIVARIATE SAMPLING

The purpose of this section is to prove the following limit theorem for d-dimensional random arrays or *d-arrays* $\xi^1, \xi^2,...,$ formed by multivariate sampling from a stationary process X in a Polish space S. We assume the parameter space of X to be of the form $T_1 \times \cdots \times T_d$, where each T_k is a closed additive subgroup of $\mathbb{R}$.

In the present context, we define the *empirical distribution* η_n of ξ^n as the random probability measure $P[\xi^n \in \cdot \mid X]$ on $S^{\mathbb{N}^d}$. The limit of ξ^n is a separately exchangeable d-array ξ, and we obtain convergence of η_n toward the associated *ergodic distribution* $P[\xi \in \cdot \mid \mathscr{I}]$, where $\mathscr{I}$ denotes the *invariant σ-field of ξ*—the set of all ξ-measurable events that are invariant under arbitrary shifts. The asserted convergence $\eta_n \xrightarrow{wP} \eta$ is defined most conveniently through the subsequence criterion. Thus for every subsequence $N' \subset \mathbb{N}$, we claim the existence of a further subsequence $N'' \subset N'$ such that $\eta_n \xrightarrow{w} \eta$ a.s. along N''.

Theorem 1. Let X be a stationary and measurable process on $T_1 \times \cdots \times T_d$ with values in a Polish space S. For each $i \leqslant d$, let $\mu_1^i, \mu_2^i, \ldots$ be asymptotically invariant distributions on T_i. For every $n \in \mathbb{N}^d$, use multivariate sampling from X to form a d-array ξ^n with sampling distribution $\mu_n = \bigotimes_{i \leqslant d} \mu_{n_i}^i$ and empirical distribution $\eta_n = P[\xi^n \in \cdot \mid X]$. Then $\xi^n \xrightarrow{d} \xi$ and $\eta_n \xrightarrow{wP} \eta$ as $n_1, \ldots, n_d \to \infty$ in this order, for some separately exchangeable d-array ξ with ergodic distribution η.

Several lemmas are needed for the proof. The following result will be used to construct the limiting array ξ by successive randomizations.

Lemma 1. Fix an S-valued process X on T, where both S and T are Borel, and let $\mathscr{F}$ be a sub-σ-field on Ω. Then there exist, on a possibly extended probability space, some processes $X_1, X_2, \ldots$ that along with X are conditionally i.i.d. given $\mathscr{F}$. If X is measurable, we may choose the X_n to be measurable as well.

Proof. Let the basic probability space be $(\Omega, \mathscr{A}, P)$ and write $\mathscr{S}$ for the σ-field in S. Put $\hat{\Omega} = \Omega \times (S^T)^\infty$ and $\hat{\mathscr{A}} = \mathscr{A} \otimes (\mathscr{S}^T)^\infty$. For any $A \in \hat{\mathscr{A}}$ we may choose a countable set $I \subset T$ and some $B \in \mathscr{A} \otimes (\mathscr{S}^I)^\infty$ such that $A = B \times (S^{T \setminus I})^\infty$. Since S^I is again Borel, there exists a regular conditional distribution $\mu_I = P^{\mathscr{F}}\{X_I \in \cdot\}$. The infinite product measure μ_I^∞ is again $\mathscr{F}$-measurable, and we may define

$$\hat{P}A = E \int 1_B(\cdot, x)\, \mu_I^\infty(dx), \qquad A = B \times (S^{T \setminus I})^\infty, \quad B \in \mathscr{A} \otimes (\mathscr{S}^I)^\infty \tag{2.1}$$

By the a.s. uniqueness of regular conditional distributions, we have $\mu_I = \mu_J(\cdot \times S^{J \setminus I})$ a.s. for all $I \subset J$, which shows that $P\hat{A}$ is independent of the choice of I. For any disjoint sets $A_1, A_2, \ldots \in \hat{\mathscr{A}}$, we may choose a countable set $I \subset T$ and some $B_1, B_2, \ldots \in \mathscr{A} \otimes (\mathscr{S}^I)^\infty$ such that $A_n = B_n \times (S^{T \setminus I})^\infty$ for all n. By (2.1) and monotone convergence,

$$\hat{P} \bigcup_n A_n = E \int \sum_n 1_{B_n}(\cdot, x)\, \mu_I^\infty(dx)$$

$$= \sum_n E \int 1_{B_n}(\cdot, x)\, \mu_I^\infty(dx) = \sum_n \hat{P} A_n$$

which shows that $\hat{P}$ is countably additive. This completes the construction of the extended probability space $(\hat{\Omega}, \hat{\mathscr{A}}, \hat{P})$.

Writing the generic elements of $\hat{\Omega}$ as $\hat{\omega} = (\omega; x_{n,t}, n \in \mathbb{N}, t \in T)$, we define

$$\hat{X}_{0,t}(\hat{\omega}) = X_t(\omega), \qquad \hat{X}_{n,t}(\hat{\omega}) = x_{n,t}, \qquad n \in \mathbb{N}, \quad t \in T.$$

For any $n \in \mathbb{N}$ and $A_0, \ldots, A_n \in \mathscr{S}^T$, we may choose a countable set $I \subset T$ and some $B_0, \ldots, B_n \in \mathscr{S}^I$ such that $A_k = B_k \times S^{T \setminus I}$ for all k. Using (2.1) and the definition and $\mathscr{F}$-measurability of μ_I, we get for any $F \in \mathscr{F}$

$$\hat{P}\left[\bigcap_{k \geq 0} \{\hat{X}_k \in A_k\}; F \times S^T\right] = E\left[\prod_{k \geq 1} \mu_I B_k; X \in A_0; F\right]$$

$$= E\left[\prod_{k \geq 0} \mu_I B_k; F\right]$$

$$= \hat{E}\left[\prod_{k \geq 0} \mu_I B_k; F \times S^T\right]$$

Since F was arbitrary, we obtain a.s.

$$\hat{P}^{\hat{\mathscr{F}}} \bigcap_k \{\hat{X}_k \in A_k\} = \prod_k \mu_I B_k = \prod_k P^{\mathscr{F}}\{X \in A_k\} = \prod_k \hat{P}^{\hat{\mathscr{F}}}\{\hat{X}_k \in A_k\}$$

which shows that $\hat{X}, \hat{X}_1, \hat{X}_2, \ldots$ are conditionally i.i.d. given $\hat{\mathscr{F}}$.

Now assume that X is product-measurable. Then X satisfies the condition of Dellacherie and Meyer,[3] [Thm. IV.30]. Since this criterion depends only on the finite-dimensional distributions which are the same for $\hat{X}_1, \hat{X}_2, \ldots$, we may apply the quoted result in the opposite direction to see that the latter processes have measurable versions $X_1, X_2, \ldots$. □

The next result shows that the original invariance properties are preserved by the randomization in Lemma 1.

Lemma 2. Fix two commuting groups $\mathscr{T}$ and $\mathscr{U}$ of measurable transformations on some measurable space S, and write $\mathscr{I}$ for the $\mathscr{U}$-invariant σ-field in S. Let the random elements $\xi_1, \xi_2,...$ in S be $\mathscr{T}$-stationary and conditionally i.i.d. given $\xi_1^{-1}\mathscr{I}$. Then the whole sequence $(\xi_1, \xi_2,...)$ is $\mathscr{T}$-stationary.

Proof. Let $I \in \mathscr{I}$, $T \in \mathscr{T}$, and $U \in \mathscr{U}$ be arbitrary. Since T and U commute, we have

$$U^{-1}T^{-1}I = T^{-1}U^{-1}I = T^{-1}I$$

which shows that $T^{-1}\mathscr{I} \subset \mathscr{I}$. By the group property of $\mathscr{T}$ we have even $T\mathscr{I} \subset \mathscr{I}$, and so $\mathscr{I} = T^{-1}T\mathscr{I} \subset T^{-1}\mathscr{I}$. Thus in fact $T^{-1}\mathscr{I} = \mathscr{I}$.

Now fix any measurable sets $B_1,..., B_n \subset S$ and a mapping $T \in \mathscr{T}$. Using the conditional i.i.d. property twice together with the $\mathscr{T}$-stationarity of ξ_1 and the $\mathscr{T}$-invariance of $\mathscr{I}$, we get

$$\begin{aligned} P\bigcap_k \{T\xi_k \in B_k\} &= E\prod_k P[T\xi_1 \in B_k \mid \xi_1^{-1}\mathscr{I}] \\ &= E\prod_k P[T\xi_1 \in B_k \mid (T\xi_1)^{-1}\mathscr{I}] \\ &= E\prod_k P[\xi_1 \in B_k \mid \xi_1^{-1}\mathscr{I}] = P\bigcap_k \{\xi_k \in B_k\} \end{aligned}$$

By a monotone class argument, this extends to $(T\xi_k) =^d (\xi_k)$. $\square$

We also need the following conditional version of Fubini's theorem.

Lemma 3. Fix a σ-finite measure space $(T, \mathscr{T}, \mu)$, a measurable process $X \geqslant 0$ on T, and a sub-σ-field $\mathscr{F}$ on Ω. Then the process $E^{\mathscr{F}} X_t$ has a measurable version, and moreover

$$E^{\mathscr{F}} \int X_t \mu(dt) = \int E^{\mathscr{F}} X_t \mu(dt) \qquad \text{a.s.} \tag{2.2}$$

Proof. We may clearly assume that μ is bounded. If the process $E^{\mathscr{F}} X_t$ has two measurable versions Y and Y', then $\int Y_t \mu(dt) = \int Y'_t \mu(dt)$ a.s. by Fubini's theorem. Hence, if (2.2) holds for one measurable version, it must be true for all.

Next consider some measurable processes $X^1, X^2,...$ with $0 \leqslant X^n \uparrow X$ and such that every process $E^{\mathscr{F}} X^n_t$ has a measurable version Y^n. Since each process Y^n can be replaced by $0 \vee Y^1 \vee \cdots \vee Y^n$, we may assume that

$0 \leqslant Y^1 \leqslant Y^2 \leqslant \cdots$. By monotone convergence for conditional expectations, the process $Y_t = \sup_n Y_t^n$ is a version of $E^{\mathscr{F}} X_t$, and (2.2) extends by ordinary monotone convergence to the processes X and Y.

Now assume that $X = 1_{A \times B}$ for some $A \in \mathscr{A}$ and $B \in \mathscr{T}$. Then for any version $\alpha = P^{\mathscr{F}} A$, the process $E^{\mathscr{F}} X_t$ has the measurable version $\alpha 1_B(t)$, and both sides of (2.2) equal $\alpha\, \mu B$ a.s. The statement extends by a monotone class argument to the indicator function X of an arbitrary set in $\mathscr{A} \otimes \mathscr{T}$. By additivity, the assertion is then true for any simple measurable function $X \geqslant 0$, and the general result follows as before by monotone convergence. □

The convergence $\eta_n \xrightarrow{wP} \eta$ in Theorem 1 will be proved by a recursive argument, based on the following one-dimensional proposition related to results in Kallenberg.[12]

Lemma 4. Fix a measurable space S, a probability space (U, ν), and some asymptotically invariant distributions $\mu_1, \mu_2, \ldots$ on T^d. Let X be an S-valued, stationary, and measurable process on $T \times U$, put $X_0 = X(0, \cdot)$, and let $X_0, \ldots, X_d$ be measurable and conditionally i.i.d., given the invariant σ-field $\mathscr{I}$ of X. Then for any bounded, measurable functions $f_1, \ldots, f_d$: $S \to \mathbb{R}$,

$$\int \mu_n(dt) \int \nu(du) \prod_{k \leqslant d} f_k \circ X(t_k, u) \xrightarrow{P} E^{\mathscr{I}} \int \nu(du) \prod_{k \leqslant d} f_k \circ X_k(u)$$

In the following proof and throughout the remainder of this section, we use λ_r to denote the uniform distribution on $[0, r] \cap T$.

Proof. Using Fubini's theorem and Lemma 3 together with the conditional independence of the X_k, we get

$$\begin{aligned}
E&\left| \int \mu_n(dt) \int \nu(du) \prod_k f_k \circ X(t_k, u) - E^{\mathscr{I}} \int \nu(du) \prod_k f_k \circ X_k(u) \right| \\
&\leqslant \int \nu(du)\, E \left| \int \mu_n(dt) \prod_k f_k \circ X(t_k, u) - E^{\mathscr{I}} \prod_k f_k \circ X_k(u) \right| \\
&\leqslant \int \nu(du)\, E \left| \int (\mu_n - \mu_n * \lambda_r^d)(dt) \prod_k f_k \circ X(t_k, u) \right| \\
&\quad + \int \nu(du)\, E \left| \int (\mu_n * \lambda_r^d)(dt) \prod_k f_k \circ X(t_k, u) - \prod_k E^{\mathscr{I}} f_k \circ X_k(u) \right| \qquad (2.3)
\end{aligned}$$

Here the first term on the right is bounded by

$$\|\mu_n - \mu_n * \lambda_r^d\| \prod_k \|f_k\| \leqslant \int \|\mu_n - \mu_n * \delta_t\| \, \lambda_r^d(dt) \prod_k \|f_k\|$$

which tends to 0 as $n \to \infty$ for fixed r, by the asymptotic invariance of the μ_n and dominated convergence. Furthermore, it is clear from the stationarity of X and the invariance of $E^{\mathscr{I}} f_k \circ X_0(u)$ that the expected value in the second term on the right of (2.3) is bounded by

$$\begin{aligned}
&\int \mu_n(ds)\, E \left| \int \lambda_r^d(dt) \prod_k f_k \circ X(s_k + t_k, u) - \prod_k E^{\mathscr{I}} f_k \circ X_k(u) \right| \\
&\quad \leqslant \sum_k \int \mu_n(ds)\, E \left| \int \lambda_r(dt)\, f_k \circ X(s_k + t, u) - E^{\mathscr{I}} f_k \circ X_0(u) \right| \prod_{j \neq k} \|f_j\| \\
&\quad = \sum_k E \left| \int \lambda_r(dt)\, f_k \circ X(t, u) - E^{\mathscr{I}} f_k \circ X(0, u) \right| \prod_{j \neq k} \|f_j\|
\end{aligned}$$

Here the last sum is independent of n and tends to 0 as $r \to \infty$, by the mean ergodic theorem. Hence, by dominated convergence, the right-hand side of (2.3) tends to 0 as $n \to \infty$ and then $r \to \infty$, and the assertion follows. □

Finally, we need some general identities involving conditional expectations with respect to invariant σ-fields. Here and below, the operators θ^t denote shifts on T. All processes X are tacitly assumed to be product-measurable. In particular, this guarantees the measurability of $\theta^t X$ as a function on $T^2 \times \Omega$.

Lemma 5. Let X be an S-valued, stationary process on T, fix any measurable function $f: S^T \to U$, and put $Y_t \equiv f(\theta^t X)$. Write $\mathscr{I}$ and $\mathscr{J}$ for the invariant σ-fields of X and Y. Then $E^{\mathscr{I}} = E^{\mathscr{J}}$ on $L^1(Y)$.

Proof. Define a function $F: S^T \to U^T$ by

$$(F(x))_t = f(\theta^t x), \qquad x \in S^T, \quad t \in T$$

and note that $Y = F(X)$. For any $x \in S^T$ and $s, t \in T$ we have

$$(F(\theta^t x))_s = f(\theta^{s+t} x) = (F(x))_{s+t} = (\theta^t F(x))_s$$

which shows that F commutes with the shifts θ^t. Thus, for any function $g: U^T \to \mathbb{R}$,

$$g(\theta^t Y) = g \circ \theta^t F(X) = g \circ F(\theta^t X), \qquad t \in T$$

Assuming g to be measurable with $g(Y) \in L^1$ and using the ergodic theorem in T, we get a.s.

$$\begin{aligned} E^{\mathscr{I}} g(Y) &= \lim_{r \to \infty} \int \lambda_r(dt)\, g(\theta^t Y) \\ &= \lim_{r \to \infty} \int \lambda_r(dt)\, g \circ F(\theta^t X) \\ &= E^{\mathscr{I}} g \circ F(X) = E^{\mathscr{J}} g(Y) \qquad \square \end{aligned}$$

The last result extends as follows to limits of convergent sequences.

Corollary 1. The last lemma remains true when U is a metric space and $f_n(\theta^t X) \xrightarrow{P} Y_t$ as $n \to \infty$ for some measurable functions $f_1, f_2,...$: $S^T \to U$.

Proof. For each $t \in T$ we have $f_n(\theta^t X) \to Y_t$ a.s. along some subsequence $N_t \subset \mathbb{N}$. Moreover, $Y_t = g_t(\theta^t X)$ a.s. for some measurable function g_t: $S^T \to U$, by Lemma 1.10 in Kallenberg.[11] Since X is stationary, we may choose both $N_t = N$ and $g_t = g$ to be independent of t. Thus, $Y_t = g(\theta^t X)$ a.s. for all $t \in T$. The assertion now follows by Lemma 5. $\square$

Here is a related result for exchangeable sequences.

Lemma 6. Let the random elements $\xi_1, \xi_2,...$ be conditionally i.i.d. given some σ-field $\mathscr{F}$, and write $\mathscr{I}$ for the invariant σ-field of $\xi = (\xi_n)$. Then $E^{\mathscr{F}} = E^{\mathscr{I}}$ on $L^1(\xi)$.

Proof. By a monotone class argument together with the linearity and monotone convergence properties of conditional expectations, it suffices to prove, for any finite collection of measurable sets $B_1,..., B_m$ in the state space S, that a.s.

$$P^{\mathscr{F}} \bigcap_{k \leqslant m} \{\xi_k \in B_k\} = P^{\mathscr{I}} \bigcap_{k \leqslant m} \{\xi_k \in B_k\}$$

By Lemma 5 we may then replace $\mathscr{I}$ by the invariant σ-field of the sequence $f(\xi_1), f(\xi_2),...$, where f denotes the function $(1_{B_1},..., 1_{B_m})$ from S to $\{0, 1\}^m$. This reduces the proof to the case when S is finite. For such an S we may introduce a regular conditional distribution $\mu = P^{\mathscr{F}}\{\xi_1 \in \cdot\}$, and we note that $E^{\mathscr{F}} = E^{\mu} = E^{\mathscr{I}}$ on $L^1(\xi)$, by Aldous[2] [Lemma 2.12 and Cor. 3.12]. $\square$

We conclude with a multivariate result of the same type.

Lemma 7. Let X be a stationary process on $T = T_1 \times \cdots \times T_d$ with one-dimensional invariant σ-fields $\mathcal{I}_1, \ldots, \mathcal{I}_d$, and put $\mathcal{I} = \bigcap_k \mathcal{I}_k$. Then $E^{\mathcal{I}} = E^{\mathcal{I}_1} \cdots E^{\mathcal{I}_d}$ on $L^1(X)$.

Proof. Fix any measurable function $f: S^T \to \mathbb{R}$ with $E\,|f(X)| < \infty$. By the ergodic theorems in $T_1, \ldots, T_d$ and T [cf. Krengel[13]], we have a.s.

$$\begin{aligned} E^{\mathcal{I}} f(X) &= \lim_{r_1, \ldots, r_d \to \infty} \int \lambda_{r_1}(dt_1) \cdots \int \lambda_{r_d}(dt_d)\, f(\theta_1^{t_1} \cdots \theta_d^{t_d} X) \\ &= \lim_{r_1 \to \infty} \int \lambda_{r_1}(dt_1) \cdots \lim_{r_d \to \infty} \int \lambda_{r_d}(dt_d)\, f(\theta_1^{t_1} \cdots \theta_d^{t_d} X) \\ &= E^{\mathcal{I}_1} \cdots E^{\mathcal{I}_d} f(X) \end{aligned}$$

□

We are now ready to prove the main result of this section.

Proof of Theorem 1. Arguing by induction on d, it is convenient to add the assertion that $\eta_n \otimes_k f_k \xrightarrow{P} \eta \otimes_k f_k$ for every finite tensor product of bounded measurable functions $f_k: S \to \mathbb{R}$. Thus we assume this condition together with the statement of the theorem for all dimensions $\leqslant d-1$. Note that the hypothesis is vacuously fulfilled for $d = 1$.

Turning to the d-dimensional case, let $\mathcal{I}_1$ denote the invariant σ-field of X with respect to shifts in T_1. By Lemma 1 there exist some measurable processes $Y_1, Y_2, \ldots$ on $T_2 \times \cdots \times T_d$ such that $X(0, \cdots)$, $Y_1, Y_2, \ldots$ are conditionally i.i.d. given $\mathcal{I}_1$. Writing $\mu_n' = \mu_{n_2}^2 \otimes \cdots \otimes \mu_{n_d}^d$ for all $n \in \mathbb{N}^d$, we get by Lemma 4

$$\begin{aligned} \eta_n \bigotimes_k f_k &= \int \mu_n(dt) \prod_k f_k \circ X(t_k) \\ &\xrightarrow{P} E^{\mathcal{I}_1} \int \mu_n'(dt) \prod_k f_k \circ Y_{k_1}(t_{k'}) \end{aligned} \tag{2.4}$$

where $t = (t_k^i) \in T_1^{\mathbb{N}} \times \cdots \times T_d^{\mathbb{N}}$, $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$, and $k' = (k_2, \ldots, k_d)$. When $d = 1$ we may take $\xi_k = Y_k$ for all k, and (2.4) becomes

$$\eta_n \bigotimes_k f_k \xrightarrow{P} E^{\mathcal{I}_1} \prod_k f_k(\xi_k) = \eta \bigotimes_k f_k$$

By Kallenberg,[12] [Lemma 2.2] we obtain $\eta_n \xrightarrow{wP} \eta$, and so by dominated convergence $\xi^n \xrightarrow{d} \xi$. This completes the proof for $d = 1$.

Next assume that $d \geqslant 2$. By Lemma 2 the process $Y = (Y_1, Y_2, \ldots)$ is again stationary under shifts in $T_2, \ldots, T_d$. Regarding Y as a process on

$T_2 \times \cdots \times T_d$ with values in S^∞, we may introduce for each $n \in \mathbb{N}^d$ the associated empirical measure η'_n based on μ'_n and let ζ_n be a $(d-1)$-array with distribution $E\eta'_n$. By the induction hypothesis, there exists a separately exchangeable $(d-1)$-array ξ in S^∞ with ergodic distribution η', such that as $n_2,..., n_d \to \infty$ in this order, we have $\zeta_n \xrightarrow{d} \xi$ and $\eta'_n \xrightarrow{wP} \eta'$ along with the corresponding condition for finite tensor products. The convergence $\zeta_n \xrightarrow{d} \xi$ remains true if ξ and $\zeta_1, \zeta_2,...$ are regarded as d-arrays in S. Since each ζ_n is exchangeable in the first index, the same property holds for ξ, which is then separately exchangeable in all d indices.

Now let $\mathscr{I}_i$ denote the invariant σ-field of ξ with respect to shifts in index i, and put $\mathscr{I} = \bigcap_{i \geq 1} \mathscr{I}_i$ and $\mathscr{I}' = \bigcap_{i \geq 2} \mathscr{I}_i$. By the formula for tensor products, we get as $n_2,..., n_d \to \infty$ in this order

$$\int \mu'_n(dt) \prod_k f_k \circ Y_{k_1}(t_{k'}) = \eta'_n \bigotimes_k f_k \xrightarrow{P} \eta' \bigotimes_k f_k = E^{\mathscr{I}'} \prod_k f_k(\xi_k) \tag{2.5}$$

Combining this with (2.4) and using the L^1-contraction property of conditional expectations, we get as $n_1 \to \infty$ and then $n_2,..., n_d \to \infty$ in this order

$$\eta_n \bigotimes_k f_k \xrightarrow{P} E^{\mathscr{I}_1} \int \mu'_n(dt) \prod_k f_k \circ Y_{k_1}(t_{k'}) \xrightarrow{P} E^{\mathscr{I}_1} E^{\mathscr{I}'} \prod_k f_k(\xi_k) \tag{2.6}$$

By Lemma 6 we may replace $\mathscr{I}_1$ in (2.6) by the corresponding invariant σ-field $\mathscr{I}_1^Y$ for the process Y. Extending (2.5) to the form

$$\int \mu'_n(dt) \prod_k f_k \circ Y_{k_1+h}(t_{k'}) \xrightarrow{P} E^{\mathscr{I}'} \prod_k f_k(\xi_{k_1+h,k'}) \equiv \gamma_h, \qquad h \in \mathbb{Z}_+$$

it is next clear from Corollary 1 that $\mathscr{I}_1^Y$ can be replaced by the invariant σ-field $\mathscr{I}_\gamma$ for the limiting sequence γ. By Lemma 5 we may finally replace $\mathscr{I}_\gamma$ by $\mathscr{I}_1$. Writing $\eta = P^{\mathscr{I}} \circ \xi^{-1}$, we get by Lemma 7

$$\eta_n \bigotimes_k f_k \xrightarrow{P} E^{\mathscr{I}} \prod_k f_k(\xi_k) = \eta \bigotimes_k f_k$$

As before, we may use Kallenberg,[12] [Lemma 2.2] to conclude that $\eta_n \xrightarrow{wP} \eta$. This completes the induction. □

3. EXCHANGEABLE ARRAYS

In this section we provide some more detailed information about exchangeable arrays and their functional representations, as required for our discussion in Section 4.

We consider random arrays X indexed by $\mathbb{N}^d = \{1, 2,...\}^d$—*random d-arrays* for short—which are invariant in distribution under certain permutations of the indices. A permutation $p = (p_1, p_2,...)$ on $\mathbb{N}$ is said to be *finite* if $p_i = i$ for all but finitely many $i \in \mathbb{N}$. Given any partition π of the set $\{1,..., d\}$ and a collection $p = (p^I; I \in \pi)$ of permutations $p^I = (p_1^I, p_2^I,...)$ on $\mathbb{N}$, we may form the array

$$(X \circ p)_k = X(p_{k_1}^{\pi_1},..., p_{k_d}^{\pi_d}), \qquad k = (k_1,..., k_d) \in \mathbb{N}^d$$

where π_i denotes the subset $I \in \pi$ that contains i. Thus, starting from X, we construct the new array $X \circ p$ by applying permutation p^I to all indices $i \in I$. We say that X is *π-exchangeable* if $X \circ p =^d X$ for every set $p = (p^I)$ of finite permutations on $\mathbb{N}$. This generalizes the previously considered notions of *separately* or *jointly exchangeable* arrays, where $\pi = \{\{1\},..., \{d\}\}$ or $\pi = \{\{1,..., d\}\}$, respectively.

For each π, the class of π-exchangeable arrays may be characterized by a general representation formula, generalizing the representations described in Section 1. The version for separately or jointly exchangeable arrays of arbitrary dimension was obtained by Hoover,[5] in a formidable unpublished manuscript based on mathematical logic and nonstandard analysis. An elementary discussion of the underlying ideas is given in Hoover.[6] The representation in the two-dimensional, separately exchangeable case was also obtained—independently and by different methods—by Aldous.[1] A standard approach to the general representation theorem appears, along with various extensions and ramifications, in Kallenberg.[9]

To describe the basic results, we define for each $k \in \mathbb{N}^d$ and $I \in 2^d$ (i.e., $I \subset \{1,..., d\}$) a vector $k_I \in \mathbb{Z}_+^d = \{0, 1,...\}^d$ and a set $\tilde{k}_I \subset \mathbb{N}$ by

$$k_I = (k_1 1_I(1),..., k_d 1_I(d)), \quad \tilde{k}_I = \{k_i; i \in I\}, \qquad k \in \mathbb{N}^d$$

The representation in the separately exchangeable case may then be written as

$$X_k = f(\xi \circ k_I; I \in 2^d) \quad \text{a.s.}, \qquad k \in \mathbb{N}^d$$

where f is a measurable function on $[0, 1]^{2^d}$ and ξ is a U-array with index set $\mathbb{Z}_+^d$. In the jointly exchangeable case, we have instead a representation

$$X_k = f(\xi \circ \tilde{k}_I; I \in 2^d) \quad \text{a.s.}, \qquad k \in \mathbb{N}^d \tag{3.1}$$

where ξ is now a U-array indexed by the collection $\tilde{\mathbb{N}}_d$ of all subsets $K \subset \mathbb{N}$ of cardinality $|K| \leqslant d$.

To state the more general representation for π-exchangeable arrays, we define for each $k \in \mathbb{N}^d$ and $I \in 2^d$ a function $\tilde{k}_{\pi I}: \pi \to \tilde{\mathbb{N}}_d$ by

$$\tilde{k}_{\pi I}(J) = \tilde{k}_{I \cap J} = \{k_i;\, i \in I \cap J\}, \qquad J \in \pi$$

As index set for the U-array ξ we choose $\times_{J \in \pi} \tilde{\mathbb{N}}_{|J|}$—the set of all functions $K_J \in \tilde{\mathbb{N}}_{|J|}$, $J \in \pi$. We say that ξ *exists* if it can be constructed on a suitable extension of the original probability space.

Proposition 1. Let X be a random d-array with values in a Borel space S. Then X is π-exchangeable iff there exist a measurable function $f: [0, 1]^{2^d} \to S$ and a U-array ξ on $\times_{I \in \pi} \tilde{\mathbb{N}}_{|I|}$ such that

$$X_k = f(\xi \circ \tilde{k}_{\pi I};\, I \in 2^d) \quad \text{a.s.}, \qquad k \in \mathbb{N}^d \tag{3.2}$$

Proof. Since X is also jointly exchangeable, it can be represented as in (3.1). Restricting X to the index set $\times_{I \in \pi} N_I^{|I|}$ for some disjoint, infinite subsets $N_I \subset \mathbb{N}$, $I \in \pi$, we note that (3.1) reduces to the form (3.2). Now introduce for each $I \in \pi$ a bijection $p^I: \mathbb{N} \to N_I$, and put $p = (p^I)$. Then $X \circ p \stackrel{d}{=} X$ by the π-exchangeability of X, and so X has the same distribution as the right-hand side of (3.2). The result now follows by Kallenberg,[11] [Thm. 5.10]. □

Say that a π-exchangeable array X is *simple* if it admits a representation

$$X_k = f(\alpha, \xi_{k_1}^{\pi_1}, \ldots, \xi_{k_d}^{\pi_d}), \qquad k = (k_1, \ldots, k_d) \in \mathbb{N}^d \tag{3.3}$$

for some measurable function f on $[0, 1]^{d+1}$ and some i.i.d. $U(0, 1)$ random variables α and ξ_k^I, $I \in \pi$, $k \in \mathbb{N}$. In particular, we get in the separately exchangeable case

$$X_k = f(\alpha, \xi_{k_1}^1, \ldots, \xi_{k_d}^d), \qquad k = (k_1, \ldots, k_d) \in \mathbb{N}^d$$

and in the jointly exchangeable case

$$X_k = f(\alpha, \xi_{k_1}, \ldots, \xi_{k_d}), \qquad k = (k_1, \ldots, k_d) \in \mathbb{N}^d$$

The property of simplicity is independent of representation, in the sense that if X has one representation as in (3.3), then all representation are essentially of this type. This is a consequence of the following result, which extends Props. 4.1 and 4.6 of Kallenberg[10] to general π-exchangeable arrays. For an array ξ indexed by $\times_{I \in \pi} \tilde{\mathbb{N}}_{|I|}$, we write $\hat{\xi}^k$ for the subarray of elements $\xi_{(K_I;\, I \in \pi)}$ with $\sum_I |K_I| \leqslant k$.

Proposition 2. Consider a π-exchangeable d-array X, admitting representations as in (3.2) in terms of some U-arrays ξ and η. Then $E[\cdot|\hat{\xi}^k] = E[\cdot|\hat{\eta}^k]$ on $L^1(X)$ for every $k \leqslant d$.

Proof. The array X is jointly spreadable, and any representation as in (3.2) can be rearranged to agree with the standard representation of such arrays, as given by Kallenberg,[9] [Thm. 4.1]. The proof of Proposition 4.6 in Kallenberg[10] applies without changes to arrays of this type. □

If the zero order variable $\alpha = \xi_\varnothing$ can be omitted from the representation in (3.2) or (3.3), then X is said to be *ergodic*. In general, we may reduce to the ergodic case by conditioning on α, which leads us to consider the *ergodic distribution* $P[X \in \cdot\,|\alpha]$. The last proposition shows that even the latter is independent of representation. For simple arrays, we define the associated *ergodic representing function* as

$$\varphi(t_1,..., t_d) = f(\alpha, t_1,..., t_d), \qquad t_1,..., t_d \in [0, 1]. \tag{3.4}$$

Note that φ is a measurable process on $[0, 1]^d$. The next result shows that, conversely, any measurable process φ on $[0, 1]^d$ with values in a Borel space S admits an a.s. representation as in (3.4) for some measurable function f: $[0, 1]^{d+1} \to S$ and some $U(0, 1)$ random variable α.

Lemma 8. Given a Borel space S and a measurable space T, consider an S-valued, measurable process X on T and a random element τ in T. Then there exist a measurable function f: $T \times [0, 1] \to S$ and some $U(0, 1)$ random variable $\alpha \perp\!\!\!\perp \tau$ such that $X_\tau = f(\tau, \alpha)$ a.s.

Proof. Let $\vartheta \perp\!\!\!\perp (X, \tau)$ be $U(0, 1)$. By Kallenberg,[11] [Thm. 5.10] there exists some measurable function f: $T \times [0, 1] \to S$ such that the random element $\xi = f(\tau, \vartheta)$ satisfies $(\xi, \tau) =^d (X_\tau, \tau)$. By the same theorem, we may next choose some random variable $\alpha = g(X_\tau, \tau, \vartheta)$ such that $(X_\tau, \tau, \alpha) =^d (\xi, \tau, \vartheta)$. Since S is Borel, the diagonal in S^2 is measurable, and we get $X_\tau = f(\tau, \alpha)$ a.s. □

Given any measurable process φ on $[0, 1)^d$ or $[0, 1]^d$, we may proceed as in (3.3) to construct a *generated* array

$$X_k = \varphi(\xi_{k_1}^{\pi_1},..., \xi_{k_d}^{\pi_d}), \qquad k = (k_1,..., k_d) \in \mathbb{N}^d$$

where $\xi = (\xi_k^i)$ is a U-array indexed by $\mathbb{N}^\pi$ and independent of φ. Note that X is π-exchangeable.

4. ESTIMATION

The purpose of this section is to prove a number of results of relevance to the basic estimation problem for exchangeable arrays described in the introduction. Our main results include an approximation by simple arrays in Theorem 2, some convergence results for empirical distributions in Theorem 3 and Corollary 2, and an estimation of simple representing functions in Theorem 4.

Let us fix an arbitrary partition π of the set $\{1,..., d\}$. The following elementary result shows that any π-exchangeable array can be approximated in distribution by simple arrays. Given any array X indexed by $\{1,..., n\}^d$, we define the associated *grid process* φ on $[0, 1)^d$ by

$$\varphi(t_1,..., t_d) = X_{[nt_1]+1,..., [nt_d]+1}, \qquad t_1,..., t_d \in [0, 1)$$

Recall from Section 3 how φ—like any other measurable process on $[0, 1)^d$—can be used to generate a π-exchangeable d-array.

Proposition 3. Fix a π-exchangeable d-array X in a Polish space S, consider for each $n \in \mathbb{N}$ the grid process φ_n based on $\chi_n X$, and form a π-exchangeable array X_n generated by φ_n. Then $X_n \xrightarrow{d} X$.

The result is essentially a consequence of the asymptotic equivalence of sampling with and without replacement. We may formalize the argument by means of the following lemma, which will also be needed later. Write $U\{1,..., n\}$ for the uniform distribution on the set $\{1,..., n\}$.

Lemma 9. Let $\tau_1, \tau_2,...$ be i.i.d. $U\{1,..., n\}$, and write $\sigma_1,..., \sigma_n$ for their distinct values in the order of first appearance. Then the σ_k are exchangeable, and moreover

$$P \bigcap_{k \leqslant m} \{\sigma_k = \tau_k\} \geqslant 1 - \frac{m(m-1)}{2n}, \qquad 1 \leqslant m \leqslant n \tag{4.1}$$

Proof. Fix an arbitrary permutation p on $\{1,..., n\}$, and note that $p(\sigma_1),..., p(\sigma_n)$ are the distinct values of $p(\tau_1), p(\tau_2),...$ in the order of first appearance. Since the $p(\tau_k)$ are again i.i.d. $U\{1,..., n\}$, we get

$$(p(\sigma_1),..., p(\sigma_n)) \stackrel{d}{=} (\sigma_1,..., \sigma_n)$$

and the first assertion follows. To prove (4.1), we note as in Aldous,[2] (5.6) that

$$
\begin{aligned}
P \bigcap_{k \leqslant m} \{\sigma_k = \tau_k\} &= P\{\tau_1, ..., \tau_m \text{ distinct}\} \\
&= \prod_{k \leqslant m} \frac{n-k+1}{n} \\
&\geqslant 1 - \frac{m(m-1)}{2n}
\end{aligned}
$$

where the last inequality is easily verified by induction. □

Proof of Proposition 3. Choose for every $n \in \mathbb{N}$ some random variables τ_{nk}^I, $I \in \pi$, $k \in \mathbb{N}$, that are i.i.d. $U\{1, ..., n\}$ and independent of X. For each $I \in \pi$, let $\sigma_{n1}^I, ..., \sigma_{nn}^I$ be the distinct values of $\tau_{n1}^I, \tau_{n2}^I, ...$ in the order of first appearance. Define $X \circ \sigma_n = X(\sigma_{nk}^I)$ and $X \circ \tau_n = X(\tau_{nk}^I)$, and note that $X \circ \sigma_n =^d \chi_n X$ and $X \circ \tau_n =^d X_n$. By Lemma 4.2 we get for $1 \leqslant m \leqslant n$

$$
\begin{aligned}
\|\mathscr{L}(\chi_m X_n) - \mathscr{L}(\chi_m X)\| &\leqslant P\{\chi_m(X \circ \sigma_n) \neq \chi_m(X \circ \tau_n)\} \\
&\leqslant 1 - P \bigcap_{I \in \pi} \bigcap_{k \leqslant m} \{\sigma_{nk}^I = \tau_{nk}^I\} \\
&\leqslant dm(m-1)/2n
\end{aligned}
$$

Here the right-hand side tends to 0 as $n \to \infty$ for fixed m. Thus $\chi_m X_n \xrightarrow{d} \chi_m X$, and the assertion follows. □

The approximation by simple arrays may be strengthened as follows. Recall that, for any two measures μ and ν, the equivalence $\mu \sim \nu$ means that $\mu \ll \nu$ and $\nu \ll \mu$.

Theorem 2. Let X be a π-exchangeable d-array in a Borel space S. Then there exist some simple π-exchangeable arrays $X_1, X_2, ...$ such that $\mathscr{L}(\chi_m X_n) \sim \mathscr{L}(\chi_m X)$ for all $m, n \in \mathbb{N}$ and the associated densities tend uniformly to 1 as $n \to \infty$ for fixed m.

Here the *existence* of the approximating arrays X_n can be proved by means of Proposition 3. With a slightly greater effort, we may provide an explicit *construction*. This requires a couple of technical lemmas, and we begin with a discrete approximation involving Bernoulli random variables. By a *Bernoulli array with rate* $p \in [0, 1]$ we mean an indexed collection of independent random variables ξ_k with $P\{\xi_k = 1\} = 1 - P\{\xi_k = 0\} = p$.

Lemma 10. Fix a Bernoulli array X on $\mathbb{N}^d$ with rate $\frac{1}{2}$. Consider for each $m \in \mathbb{N}$ a Bernoulli array (ξ^i_{jk}) on $\mathbb{N}^3$ with rate $p_m = (m^{-1} \log 2)^{1/d}$, and define

$$X^m_k = \max_{j \leqslant m} \prod_{i \leqslant d} \xi^i_{j, k_i}, \qquad k = (k_1, ..., k_d) \in \mathbb{N}^d \tag{4.2}$$

Then

$$\|\mathscr{L}(\chi_n X^m) - \mathscr{L}(\chi_n X)\| \preccurlyeq n^{d+1} p_m d \tag{4.3}$$

Proof. Since the total variation in (4.3) is bounded by 2, we may assume that $n^2 p_m \leqslant \frac{1}{2}$. Introduce the auxiliary array

$$S^m_k = \sum_{j \leqslant m} \prod_{i \leqslant d} \xi^i_{j, k_i} = \sum_{j \leqslant m} Y^m_{jk}, \qquad k = (k_1, ..., k_d) \in \mathbb{N}^d$$

Put $\nu_m = \mathscr{L}(Y^m_j)$ and let Z^m be an infinitely divisible array with Lévy measure $m\nu_m$. By Matthes *et al.*,[14] [Lemma 1.11.2]

$$\begin{aligned} \|\mathscr{L}(\chi_n S^m) - \mathscr{L}(\chi_n Z^m)\| &\preccurlyeq m \left\{ E \sum_k \chi_n Y^m_{jk} \right\}^2 = m(np_m)^{2d} \\ &\preccurlyeq (n^2 p_m)^d \leqslant n^2 p_m \end{aligned} \tag{4.4}$$

Next let $\nu_{m,n} = P[\chi_n Y^m_j \in \cdot\,; A_n]$ where $A_n = \{\sum_k \chi_n Y^m_{jk} = 1\}$, and write $Z^{m,n}$ for a homogeneous Poisson process on $\{1, ..., n\}^d$ with Lévy measure $m\nu_{m,n}$. By Matthes *et al.*,[14] [Lemma 1.10.6]

$$\begin{aligned} \|\mathscr{L}(\chi_n Z^m) - \mathscr{L}(Z^{m,n})\| &\preccurlyeq m\, P \left\{ \sum_k \chi_n Y^m_{jk} > 1 \right\} \\ &\leqslant m \left| E \sum_k \chi_n Y^m_{jk} - P\{\chi_n Y^m_j \neq 0\} \right| \\ &= m\, |(np_m)^d - (1-(1-p_m)^n)^d| \\ &\preccurlyeq m(np_m)^d\, np_m d \preccurlyeq n^{d+1} p_m d \end{aligned} \tag{4.5}$$

Now let Z be a Poisson array on $\mathbb{N}^d$ with rate $\log 2$. By another application of Matthes *et al.*,[14] [Lemma 1.10.6] we get as before

$$\begin{aligned} \|\mathscr{L}(Z^{m,n}) - \mathscr{L}(\chi_n Z)\| &\preccurlyeq \left| E \sum_k Z^{m,n}_k - E \sum_k \chi_n Z_k \right| \\ &= m\, |(1-(1-p_m)^n)^d - (np_m)^d| \\ &\preccurlyeq n^{d+1} p_m d \end{aligned} \tag{4.6}$$

Combining (4.4)–(4.6) gives

$$\|\mathscr{L}(\chi_n S^m) - \mathscr{L}(\chi_n Z)\| \lesssim n^{d+1} p_m d$$

The estimate (4.3) now follows, since $X_k^m = S_k^m \wedge 1$ for all $k \in \mathbb{N}^d$ and we may take $X_k \equiv Z_k \wedge 1$. □

As a second step in the proof of Theorem 2, we show how a U-array on $\mathbb{N}^d$ can be approximated by simple arrays.

Lemma 11. Let Y be a U-array on $\mathbb{N}^d$ and define $Y_r = \sum_{k \geqslant 1} 2^{-k} X_{k+r}$, where the X_k are independent arrays as in (4.2) with parameters $m_k = 2^{kd}$. Then $\mathscr{L}(\chi_n Y_r) \sim \mathscr{L}(\chi_n Y)$ for large r and all n, and the associated densities f_{nr} tend uniformly to 1 as $r \to \infty$ for fixed n.

Proof. Let X be a Bernoulli array on $\mathbb{N}^d$ with rate $\frac{1}{2}$ and write g_{nk} for the density of $\mathscr{L}(\chi_n X_k)$ with respect to $\mathscr{L}(\chi_n X)$. Using the fact that $\chi_n X$ has finite range with positive probabilities 2^{-n^d}, and noting that for $\sum_k |r_k| \leqslant \frac{1}{2}$

$$\left| \prod_k (1 + r_k) - 1 \right| \leqslant \prod_k (1 + |r_k|) - 1$$

$$\leqslant \exp \sum_k |r_k| - 1 \lesssim \sum_k |r_k|$$

we get by Lemma 10

$$\|f_{nr} - 1\| \lesssim \sum_{k > r} \|g_{nk} - 1\|$$

$$\lesssim 2^{n^d} \sum_{k > r} \|\mathscr{L}(\chi_n X_k) - \mathscr{L}(\chi_n X)\|$$

$$\lesssim 2^{n^d} n^{d+1} d \sum_{k > r} 2^{-k} = 2^{n^d - r} n^{d+1} d$$

which tends to 0 as $r \to \infty$ for fixed n. □

Proof of Theorem 2. By Proposition 1 it is enough to consider arrays of the form $X_k \equiv \eta_{\tilde{k}}$, where η is a U-array indexed by the set of all collections $\tilde{k} = (\tilde{k}_I; I \in \pi)$ with $\tilde{k}_I = \{k_i; i \in I\}$. Writing $|(k)| = \prod_{I \in \pi} |\tilde{k}_I|!$, we may then assume that

$$X_k = \left(\max_p Y_{k \circ p} \right)^{|(k)|}, \qquad k \in \mathbb{N}^d \tag{4.7}$$

for some U-array Y on $\mathbb{N}^d$, where the maximum extends over all permutations p of $1,..., d$ that leave the sets $I \in \pi$ invariant.

Now introduce some i.i.d. $U(0, 1)$ random variables ξ_k^I, $I \in \pi$, $k \in \mathbb{N}$, and let $g_1,..., g_d$ be independent $U(0, 1)$ random variables on the Lebesgue unit interval. Then the variables $\gamma_k^i = g_i(\xi_k^{\pi_i})$ form a U-array on the index set $\{1,..., d\} \times \mathbb{N}$, and we may choose some functions $f_1, f_2,...$ on $[0, 1]^d$ such that the arrays

$$Y_k^n = f_n(\gamma_{k_1}^1,..., \gamma_{k_d}^d), \qquad k = (k_1,..., k_d) \in \mathbb{N}^d$$

satisfy the condition of Lemma 11. Imitating (4.7), we define

$$X_k^n = \left(\max_p Y_{k\circ p}^n\right)^{|(k)|}, \qquad k \in \mathbb{N}^d$$

and we note that the distribution of X^n approximates that of X in the required sense. Moreover, it is easy to verify that the arrays X^n are simple and π-exchangeable. □

We proceed to establish a general ergodic theorem for the empirical distributions of an exchangeable array. Given a partition π of $\{1,..., d\}$ and a random array X indexed by $\{1,..., n\}^d$, we define the associated *empirical distribution* as the random measure η such that $\eta f = (n!)^{-|\pi|} \sum_p f(X \circ p)$ for every measurable function $f \geqslant 0$. Here the transformed array $X \circ p$ is defined as in Section 3, and the summation extends over all families $p = (p^I; I \in \pi)$ of permutations of $\{1,..., n\}$.

Theorem 3. In a Polish space S, let X be a π-exchangeable d-array with ergodic distribution η, and consider for each $n \in \mathbb{N}$ the empirical distribution η_n of $\chi_n X$. Then $\eta_n \xrightarrow{w} \eta$ a.s.

Our proof of Theorem 3 relies on the following law of large numbers for exchangeable arrays. Here the *ergodic mean* is defined as the expected value with respect to the ergodic distribution η. By the *average* of a matrix we mean the arithmetic mean of its entries. Two elements X_k and X_l of an array X are said to *communicate* if $p(k) = l$ for some set $p = (p^I; I \in \pi)$ of permutations on $\mathbb{N}$. Thus, whenever i and j belong to the same set $I \in \pi$, we require that $k_i = k_j$ iff $l_i = l_j$. To ensure communication between *all* pairs of elements of an array, we may need to omit certain diagonal elements. Put $\bar{n} = (n,..., n)$.

Lemma 12. Fix a π-exchangeable d-array X of communicating, integrable elements with ergodic mean γ, and consider for each $n \in \mathbb{N}$ the average γ_n of $\chi_n X$. Then $\gamma_n \to \gamma$ a.s.

Proof. Let $\mathscr{I}$ denote the invariant σ-field of X with respect to joint shifts in all indices, and let $\mathscr{J}$ be the corresponding σ-field for the representing array ξ. By Lemma 5 we have $E^{\mathscr{I}} = E^{\mathscr{J}}$ on $L^1(X)$, and moreover $\mathscr{J} = \sigma(\alpha)$ by Kallenberg,[11] [Cor. 6.25]. Thus $P[X \in \cdot \mid \mathscr{I}] = \eta$ a.s.

Now define

$$\mathscr{F}_{-n} = \mathscr{I} \vee \sigma\{\gamma_n, \gamma_{n+1}, \ldots\}, \qquad n \in \mathbb{N}$$

and fix any $A \in \mathscr{F}_{-\infty} = \bigcap_n \mathscr{F}_{-n}$. By Kallenberg,[11] [Lemma 2.16] we may choose some events $A_n \in \sigma(\chi_n X)$ with $P(A \triangle A_n) \to 0$. Here $A_n = \{X \in B_n\}$ for some Borel sets B_n, and the events $A'_n = \{\theta_{\tilde{n}} X \in B_n\}$ satisfy $P(A \triangle A'_n) = P(A \triangle A_n) \to 0$ by the exchangeability of X. Arguing as in Aldous,[2] Prop. (14.6), we also note that $\chi_n X \perp\!\!\!\perp_{\mathscr{I}} \theta_{\tilde{n}} X$ and in particular $A_n \perp\!\!\!\perp_{\mathscr{I}} A'_n$. Letting $n \to \infty$, we get $A \perp\!\!\!\perp_{\mathscr{I}} A$ by the L^1-contractivity of conditional expectations. Thus $P^{\mathscr{I}} A \in \{0, 1\}$ a.s., which implies $A = \{P^{\mathscr{I}} A = 1\}$ a.s. Hence, $\mathscr{I} \subset \mathscr{F}_{-\infty} \subset \bar{\mathscr{I}}$ and so $\bar{\mathscr{F}}_{-\infty} = \bar{\mathscr{I}}$, where the bars denote completion.

Proceeding as in the martingale proof of the law of large numbers [cf. Kallenberg,[11] p. 110], we get a.s. for any $k \leqslant \tilde{n}$

$$\gamma_n = E[\gamma_n \mid \mathscr{F}_{-n}] = E[X_k \mid \mathscr{F}_{-n}] \to E[X_k \mid \mathscr{F}_{-\infty}] = E[X_k \mid \mathscr{I}] = \gamma$$

where the second and third steps hold by exchangeability and reverse martingale convergence, respectively. □

Proof of Theorem 3. Fix any $m \geqslant d$ and define $Y_k = X \circ k$ for all matrices $k = (k_j^I)$ indexed by $\pi \times \{1, \ldots, m\}$ such that the k_j^I are distinct for fixed I. Thus, Y is an infinite array of dimension $m\,|\pi|$ whose entries are S-valued random matrices of order m^d. This defines Y as a measurable function ψ_m of X, and we note that ψ_m has a measurable inverse ψ_m^{-1}.

Now consider an arbitrary array $p = (p^I; I \in \pi)$ such that each p^I is a permutation or shift on $\mathbb{N}$. Letting T_p denote the induced permutation or shift operator on the set of arrays X or Y, we note that

$$\begin{aligned}(T_p \psi_m X)_k &= (T_p Y)_k = Y_{p \circ k} = X \circ (p \circ k) \\ &= (X \circ p) \circ k = (T_p X) \circ k = (\psi_m T_p X)_k\end{aligned}$$

Thus, ψ_m commutes with all the operators T_p, and the same thing is then true even for the inverse map ψ_m^{-1}. In particular, the π-exchangeability of X carries over to Y. Writing $\mathscr{I}_X$ and $\mathscr{I}_Y$ for the invariant σ-fields of X and Y with respect to joint shifts in all indices, it is further clear that $\mathscr{I}_X = \mathscr{I}_Y$. Finally, we note as before that $\eta = P[X \in \cdot \mid \mathscr{I}_X]$ and similarly for the ergodic distribution of Y.

Next we introduce for each $n \geqslant m$ a random matrix $\tau_n = (\tau_{nj}^I) \perp\!\!\!\perp X$ indexed by $\pi \times \{1,..., m\}$, assumed to be uniformly distributed subject to the constraints that the τ_{nj}^I are distinct for each I and $\tau_{nj}^I \leqslant n$. Fix any bounded, measurable function f on S^{m^d}. Using Lemma 12, the equality $\mathscr{I}_X = \mathscr{I}_Y$, and the definitions of Y, η, and η_n, we get a.s. as $n \to \infty$

$$\begin{aligned}\eta_n(f \circ \chi_m) &= E[f(X \circ \tau_n) \mid X] = E[f(Y_{\tau_n}) \mid Y] \\ &\to E[f(Y_k) \mid \mathscr{I}_Y] = E[f(X \circ k) \mid \mathscr{I}_X] = \eta(f \circ \chi_m)\end{aligned}$$

Thus $\eta_n \circ \chi_m^{-1} \xrightarrow{w} \eta \circ \chi_m^{-1}$ a.s., and since m was arbitrary, we may conclude from Kallenberg,[11] [Thm. 3.29] that $\eta_n \xrightarrow{w} \eta$ a.s. □

Using the last theorem, we now show how the grid processes φ_n based on the restrictions $\chi_n X$ may serve as approximations to the representing function of X, in the sense that the ergodic distributions of the generated π-exchangeable arrays converge to the ergodic distribution of X.

Corollary 2. Fix a π-exchangeable d-array X in a Polish space S with ergodic distribution η, consider for each $n \in \mathbb{N}$ the grid process φ_n based on $\chi_n X$, and write ζ_n for the ergodic distribution of a π-exchangeable array generated by φ_n. Then $\zeta_n \xrightarrow{w} \eta$ a.s.

Proof. Let $\sigma_n = (\sigma_{nk}^I)$ and $\tau_n = (\tau_{nk}^I)$ be such as in the proof of Proposition 3, and note that a.s.

$$\eta_n = P[X \circ \sigma_n \in \cdot \mid \chi_n X], \qquad \zeta_n = P[X \circ \tau_n \in \cdot \mid \chi_n X], \qquad n \in \mathbb{N}$$

Using Lemma 9 and Fubini's theorem, we get a.s. for $1 \leqslant m \leqslant n$

$$\begin{aligned}\|(\eta_n - \zeta_n) \circ \chi_m^{-1}\| &\leqslant P[\chi_m(X \circ \sigma_n) \neq \chi_m(X \circ \tau_n) \mid \chi_n X] \\ &\leqslant 1 - P \bigcap_{I \in \pi} \bigcap_{k \leqslant m} \{\sigma_k^I = \tau_k^I\} \\ &\leqslant dm(m-1)/2n\end{aligned}$$

which tends to 0 as $n \to \infty$ for fixed m. For any bounded, continuous function $f: S^{m^d} \to \mathbb{R}$, we obtain a.s. on the set $A = \{\eta_n \xrightarrow{w} \eta\}$

$$|(\zeta_n - \eta)(f \circ \chi_m)| \leqslant |(\eta_n - \eta)(f \circ \chi_m)| + \frac{dm(m-1)}{2n} \|f\| \to 0$$

Thus, $\zeta_n \xrightarrow{w} \eta$ a.s. on A by Kallenberg,[11] [Thm. 3.29] and it remains to note that $PA = 1$ by Theorem 3. □

Finally, we consider the problem of estimating the representing function itself, in the case of a simple, π-exchangeable d-array X. Because of the nonuniqueness of the representation, we must allow certain transformations of the estimating processes φ_n. As before, we choose the latter to be grid processes based on the truncated arrays $\chi_n X$. In this context, two processes φ and ψ with the same grid size n^{-1} will be regarded as *equivalent* (written as $\varphi \sim \psi$) if the generating matrices X and Y on $\{1,..., n\}^d$ satisfy $X \circ p = Y$ for some family $p = (p^I; I \in \pi)$ of random permutations on $\{1,..., n\}$.

To describe the appropriate norms, we note that any partition π of $\{1,..., d\}$ defines an associated *diagonal space* $D_\pi \subset [0, 1]^d$, consisting of all points $(t_1,..., t_d)$ such that $t_i = t_j$ whenever i and j belong to the same subset $I \in \pi$. For any such diagonal space $D = D_\pi$, we write λ_D for the associated Lebesgue measure, obtained by the obvious projection of λ^π onto D. Finally, $\|\cdot\|_D$ is defined as the norm in $L^1(\lambda_D)$. For any two partitions π and π', we note that $D_{\pi'} \supset D_\pi$ iff π' is a refinement of π.

Theorem 4. Fix a bounded, π-exchangeable d-array X in $\mathbb{R}$ with a simple representing process φ, and consider for each $n \in \mathbb{N}$ the grid process φ_n based on $\chi_n X$. Then there exist some processes $\tilde{\varphi}_n \sim \varphi_n$ satisfying

$$E\,\|\tilde{\varphi}_n - \varphi\|_D \to 0, \qquad D \supset D_\pi \tag{4.8}$$

The proof will be based on a simple lemma. To state the result, consider a U-array $\xi = (\xi_k^I)$ indexed by $\pi \times \mathbb{N}$. For each $I \in \pi$ and $n \in \mathbb{N}$, let $\eta_{n1}^I,..., \eta_{nn}^I$ denote the variables $\xi_1^I,..., \xi_n^I$, enumerated in increasing order. Recall that π_i denotes the set $I \in \pi$ with $i \in I$. For each $n \in \mathbb{N}$, we define a process $\tau_n = (\tau_n^1,..., \tau_n^d)$ on $[0, 1)^d$ by

$$\tau_n^i(t) = \eta_{n, [nt_i]+1}^{\pi_i}, \qquad i \leqslant d, \quad t = (t_1,..., t_d) \in [0, 1)^d \tag{4.9}$$

We need some simple properties of the τ_n.

Lemma 13. Fix a partition π of $\{1,..., d\}$, and let the processes $\tau_1, \tau_2,...$ on $[0, 1)^d$ be given by (4.9). Then

(i) $(P \otimes \lambda^d) \circ \tau_n^{-1} \leqslant \lambda^d$ on the nondiagonal part of $[0, 1]^d$;

(ii) $\tau_n(t) \to t$ a.s. for all $t \in [0, 1)^d$.

Proof.

(i) Since $\xi_{k_1}^{\pi_1},..., \xi_{k_d}^{\pi_d}$ are i.i.d. $U(0, 1)$ whenever the pairs $(\pi_1, k_1),..., (\pi_d, k_d)$ are distinct, we get for any nondiagonal Borel set $B \subset [0, 1]^d$

$$
\begin{aligned}
(P\otimes\lambda^d)\circ\tau_n^{-1}B &= n^{-d}\sum_{k_1,\dots,k_d} P\{(\eta^{\pi_1}_{n,k_1},\dots,\eta^{\pi_d}_{n,k_d})\in B\}\\
&= n^{-d}E\sum_{k_1,\dots,k_d} 1_B(\eta^{\pi_1}_{n,k_1},\dots,\eta^{\pi_d}_{n,k_d})\\
&= n^{-d}E\sum_{k_1,\dots,k_d} 1_B(\xi^{\pi_1}_{k_1},\dots,\xi^{\pi_d}_{k_d})\\
&= n^{-d}\sum_{k_1,\dots,k_d} P\{(\xi^{\pi_1}_{k_1},\dots,\xi^{\pi_d}_{k_d})\in B\}\\
&\leqslant P\{(\xi^I_1,\dots,\xi^I_d)\in B\} = \lambda^d B
\end{aligned}
$$

(ii) Here it is enough to take $d=1$. The result then follows easily from the Glivenko–Cantelli lemma. □

Proof of Theorem 4. Define as in (4.9) the processes $\tau_1, \tau_2, \ldots \perp\!\!\!\perp \varphi$ on $[0, 1)^d$ and put $\tilde{\varphi}_n = \varphi\circ\tau_n$. By Fubini's theorem and dominated convergence, it suffices to prove (4.8) in the case when $\varphi = f$ is nonrandom. We may also assume that $\|f\|\leqslant 1$. Fixing any diagonal space $D\supset D_\pi$, we may choose some continuous functions $f_1, f_2,\ldots$ on D such that $\|f_m\|\leqslant 1$ and $\|f-f_m\|_D\to 0$. By Lemma 9 the subdiagonal part of the n^{-1}-grid in D has λ_D-measure $\leqslant d'(d'-1)/2n$, where d' denotes the dimension of D. Thus, in view of Lemma 13 (i),

$$
\begin{aligned}
&E\,\|f\circ\tau_n - f\|_D\\
&\qquad\leqslant E\,\|(f-f_m)\circ\tau_n\|_D + E\,\|f_m\circ\tau_n - f_m\|_D + \|f_m - f\|_D\\
&\qquad\leqslant E\,\|f_m\circ\tau_n - f_m\|_D + 2\,\|f-f_m\|_D + d(d-1)/n \qquad (4.10)
\end{aligned}
$$

Using Lemma 13 (ii) and the continuity of the f_m, we obtain $f_m\circ\tau_n(t)\to f_m(t)$ a.s. for every $t\in D$ and $m\in\mathbb{N}$. By Fubini's theorem and dominated convergence it follows that $E\,\|f_m\circ\tau_n - f_m\|_D\to 0$ as $n\to\infty$ for fixed m. Hence, the right-hand side of (4.10) tends to 0 as $n\to\infty$ and then $m\to\infty$, and the assertion follows. □

REFERENCES

1. Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivar. Anal.* **11**, 581–598.
2. Aldous, D. J. (1985). Exchangeability and related topics. *École d'Été de Prob. Saint-Flour XIII*, Springer, Berlin. *LNM* **1117**, 1–198.
3. Dellacherie, C., and Meyer, P. A. (1975). *Probabilités et Potentiel*, Chap. I–IV. Hermann, Paris.

4. Diaconis, P., and Freedman, D. (1981). On the statistics of vision: The Julesz conjecture. *J. Math. Psychol.* **24**, 112–138.
5. Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. Institute of Advanced Study, Princeton. Preprint.
6. Hoover, D. N. (1982). Row-column exchangeability and a generalized model for probability. In Koch, G., and Spizzichino, F. (eds.), *Exchangeability in Probability and Statistics*, North-Holland, Amsterdam, pp. 281–291.
7. Hoover, D. N. (1989). Tail fields of partially exchangeable arrays. *J. Multivar. Anal.* **31**, 160–163.
8. Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. *J. Multivar. Anal.* **30**, 137–154.
9. Kallenberg, O. (1992). Symmetries on random arrays and set-indexed processes. *J. Theor. Prob.* **5**, 727–765.
10. Kallenberg, O. (1995). Random arrays and functionals with multivariate rotational symmetries. *Prob. Th. Rel. Fields* **103**, 91–141.
11. Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer, New York.
12. Kallenberg, O. (1999). Asymptotically invariant sampling and averaging from stationary-like processes. *Stoch. Proc. Appl.* **82** (to appear).
13. Krengel, U. (1985). *Ergodic Theorems*, de Gruyter, Berlin.
14. Matthes, K., Kerstan, J., and Mecke, J. (1978). *Infinitely Divisible Point Processes*, Wiley, Chichester.

Printed in Belgium