CONTENTS

0. Introduction

If you had asked a probabilist in 1970 what was known about exchangeability, you would likely have received the answer "There's de Finetti's theorem: what else is there to say?" The purpose of these notes is to dispel this (still prevalent) attitude by presenting, in Parts II-IV, a variety of mostly post-1970 results relating to exchangeability. The selection of topics is biased toward my own interests, and away from those areas for which survey articles already exist. Any student who has taken a standard first year graduate course in measure-theoretic probability theory (e.g. Breiman (1968)) should be able to follow most of this article; some sections require knowledge of weak convergence.

In Bayesian language, de Finetti's theorem says that the general infinite exchangeable sequence (Z_i) is obtained by first picking a distribution θ at random from some prior, and then taking (Z_i) to be i.i.d. with distribution θ. Rephrasing in the language of probability theory, the theorem says that with (Z_i) we can associate a random distribution $\alpha(\omega,\cdot)$ such that, conditional on $\alpha = \theta$, the variables (Z_i) are i.i.d. with distribution θ. This formulation is the central fact in the circle of ideas surrounding de Finetti's theorem, which occupies most of Part I. No previous knowledge of exchangeability is assumed, though the reader who finds my proofs overly concise should take time out to read the more carefully detailed account in Chow and Teicher (1978), Section 7.3.

Part II contains results complementary to de Finetti's theorem. Dacunha-Castelle's "spreading-invariance" property and Kallenberg's stopping time property give conditions on an infinite sequence which turn out to be equivalent to exchangeability. Kingman's "paintbox" description of exchangeable random partitions leads to Cauchy's formula for the distribution of cycle lengths in a uniform random permutation, and to results about components of random functions. Continuous-time processes with interchangeable increments are discussed; a notable result is that any continuous-path process on $[0,\infty)$ (resp. $[0,1]$) with interchangeable increments is a mixture of processes which are linear transformations of Brownian motion (resp. Brownian bridge). The subsequence principle reveals exchangeable-like sequences lurking unsuspectedly within arbitrary sequences of random variables. And we discuss exchangeability in abstract spaces, and weak convergence issues.

The class of exchangeable sequences is the class of processes whose distributions are invariant under a certain group of transformations; in Part III related invariance concepts are described. After giving the abstract result on ergodic decompositions of measures invariant under a group of transformations, we specialize to the setting of <u>partial exchangeability</u>, where we study the class of processes $(X_i: i \in I)$ invariant under the action of some group of transformations of the index set I. Whether anything can be proved about partially exchangeable classes in general is a challenging open problem; we can only discuss three particular instances. The most-studied instance, investigated by Hoover and by myself, is partial exchangeability for arrays of random variables, where the picture is fairly complete. We also discuss partial exchangeability on trees of infinite degree, where the basic examples are reversible Markov chains; and on infinite-dimensional cubes, where it appears that the basic examples are random walks, though here the picture remains fragmentary.

Part IV outlines other topics of current research. A now-classical result on convergence of partial sum processes from sampling without replacement to Brownian bridge leads to general questions of convergence for triangular arrays of finite exchangeable sequences, where the present picture is unsatisfactory for applications. Kingman's uses of exchangeability in mathematical genetics will be sketched. The theory of sufficient statistics and mixtures of processes of a specified form will also be sketched--actually, this topic is perhaps the most widely studied relative of exchangeability, but in view of the existing accounts in Lauritzen (1982) and Diaconis and Freedman (1982), I have not emphasized it in these notes. Kallenberg's stopping time approach to continuous-time exchangeability is illustrated by the study of exchangeable subsets of $[0,\infty)$. A final section provides references to work related to exchangeability not elsewhere discussed: I apologize in advance to those colleagues whose favorite theorems I have overlooked.

General references. Chow and Teicher (1978) is the only textbook (known to me) to give more than a cursory mention to exchangeability. A short but elegant survey of exchangeability, whose influence can be seen in these notes, has been given by Kingman (1978a). In 1981 a conference on "Exchangeability in Probability and Statistics" was held in Rome to honor Professor Bruno de Finetti: the conference proceedings (EPS in the References) form a sample of the current interests of workers in exchangeability. Dynkin (1978) gives a concise abstract treatment of the "sufficient statistics" approach in several areas of probability including exchangeability.

The material in Sections 13 and 16 is new, and perhaps a couple of proofs elsewhere may be new; otherwise no novelty is claimed.

Notation and terminology. The mathematical notation is intended to be standard, so the reader should seldom find it necessary to consult the list of notation at the end. As for terminology, "exchangeable" is more popular and shorter than the synonyms "symmetrically dependent" and "interchangeable". I have introduced "directing random measure" in place of Kallenberg's "canonical random measure", partly as a more vivid metaphor and partly for more grammatical flexibility, so one can say "directed by ...". I use "partial exchangeability" in the narrow sense of Section 12 (processes with certain types of invariance) rather than in the wider context of Section 18 (processes with specified sufficient statistics). "Problem" means "unsolved problem" rather than "exercise": if you can solve one, please let me know.

Acknowledgements. My thanks to Persi Diaconis for innumerable invaluable discussions over the last several years; and to the members of the audiences at St. Flour and the Berkeley preview who detected errors and contributed to the presentation. Research supported by National Science Foundation Grant MCS80-02698.

PART I

The purpose of Part I is to give an account of de Finetti's theorem and some straightforward consequences, using the language and techniques of modern probability theory. I have not attempted to assign attributions to these results: historical accounts of de Finetti's work on exchangeability and the subsequent development of the subject can be found in EPS (Foreword, and Furst's article) and in Hewitt and Savage (1955).

1. Definitions and immediate consequences

A finite sequence $(Z_1,\dots,Z_N)$ of random variables is called exchangeable (or N-exchangeable, to indicate the number of random variables) if

$$(1.1) \qquad (Z_1,\dots,Z_N) \overset{\mathcal{D}}{=} (Z_{\pi(1)},\dots,Z_{\pi(N)}) \;;$$

each permutation π of $\{1,\dots,N\}$. An infinite sequence $(Z_1,Z_2,\dots)$ is called exchangeable if

$$(1.2) \qquad (Z_1,Z_2,\dots) \overset{\mathcal{D}}{=} (Z_{\pi(1)},Z_{\pi(2)},\dots)$$

for each finite permutation π of $\{1,2,\dots\}$, that is each permutation for which $\#\{i: \pi(i) \neq i\} < \infty$. Throughout Part I we shall regard random variables Z_i as real-valued; but we shall see in Section 7 that most results remain true whenever the Z_i have any "non-pathological" range space.

There are several obvious reformulations of these definitions. Any finite permutation can be obtained by composing permutations which transpose 1 and $n > 1$; so (1.2) is equivalent to the at first sight weaker condition

(1.3) $\quad (Z_1,\ldots,Z_{n-1},Z_n,Z_{n+1},\ldots) \overset{\mathcal{D}}{=} (Z_n,Z_2,\ldots,Z_{n-1},Z_1,Z_{n+1},\ldots)$;

each $n > 1$. In the other direction, (1.2) implies the at first sight stronger condition

(1.4) $$(Z_1,Z_2,Z_3,\ldots) \overset{\mathcal{D}}{=} (Z_{n_1},Z_{n_2},Z_{n_3},\ldots) ;$$

each sequence (n_i) with distinct elements. In Section 6 we shall see some non-trivially equivalent conditions.

Sampling variables. The most elementary examples of exchangeability arise in sampling. Suppose an urn contains N balls labelled $x_1,\ldots,x_N$. The results $Z_1,Z_2,\ldots$ of an infinite sequence of draws with replacement form an infinite exchangeable sequence; the results $Z_1,\ldots,Z_N$ of N draws without replacement form a N-exchangeable sequence (sequences of this latter type we call urn sequences).

Both ideas generalize. In the first case, (Z_i) is i.i.d. uniform on $\{x_1,\ldots,x_N\}$; obviously any i.i.d. sequence is exchangeable. In the second case we can write

(1.5) $$(Z_1,\ldots,Z_N) = (x_{\pi^*(1)},\ldots,x_{\pi^*(N)})$$

where π^* denotes the uniform random permutation on $\{1,\ldots,N\}$, that is $P(\pi^* = \pi) = 1/N!$ for each π. More generally, let $(Y_1,\ldots,Y_N)$ be arbitrary random variables, take π^* independent of (Y_i), and then

(1.6) $$(Z_1,\ldots,Z_N) = (Y_{\pi^*(1)},\ldots,Y_{\pi^*(N)})$$

defines a N-exchangeable sequence. This doesn't work for infinite sequences, since we cannot have a uniform permutation of a countable infinite set (without abandoning countable additivity--see (13.27)). However we can define a uniform random ordering on a countable infinite set: simply define $i \overset{\omega}{\leq} j$ to mean $\xi_i(\omega) \leq \xi_j(\omega)$, for i.i.d. continuous $(\xi_1, \xi_2, \ldots)$. This trick is useful in several contexts--see (11.9), (17.4), (19.8).

Correlation structure. Exchangeability restricts the possible correlation structure for square-integrable sequences. Let (Z_i) be N-exchangeable. Then there is a correlation $\rho = \rho(Z_i, Z_j)$, $i \neq j$. We assert

(1.7) $\quad \rho \geq \frac{-1}{N-1}$, with equality iff $\sum Z_i$ is a.s. constant.

In particular, $\rho = -1/(N-1)$ for sampling without replacement from an N-element urn. To prove (1.7), linearly scale to make $EZ_i = 0$, $EZ_i^2 = 1$, and then

$$0 \leq E(\sum Z_i)^2 = EZ_i^2 + \sum_{1 \leq i \neq j \leq N}\sum EZ_iZ_j = N + N(N-1)\rho .$$

So $\rho \geq -1/(N-1)$, with equality iff $\sum Z_i = 0$ a.s.

Observe that (1.7) implies

(1.8) $\quad \rho \geq 0$ for an infinite exchangeable sequence.

Conversely, every $\rho \leq 1$ satisfying (1.7) (resp. (1.8)) occurs as the correlation in some N-exchangeable (resp. infinite exchangeable) sequence. To prove this, let (ξ_i) be i.i.d., $E\xi_i = 0$, $E\xi_i^2 = 1$. Define

$$Z_i = \xi_i + c\sum_{j=1}^{N} \xi_j \ ; \quad 1 \leq i \leq N ,$$

for some constant c. Then (Z_i) is N-exchangeable, and a simple computation gives $\rho = 1 - (Nc^2 + 2c + 1)^{-1}$. As c varies we get all values $1 > \rho \geq -(N-1)^{-1}$. (The case $c = -1/N$ which gives $\rho = -1/(N-1)$ will be familiar from statistics!). Of course we can get $\rho = 1$ by setting $Z_1 = \cdots = Z_N$. In the infinite case, take $(\xi_i : i \geq 0)$ i.i.d. and set

$$Z_i = c\xi_0 + \xi_i \; ; \quad i \geq 1$$

for some constant c. Then (Z_i) is an infinite exchangeable sequence with $\rho = c^2/(c^2+1)$, and as c varies we get all values $0 \leq \rho < 1$.

Gaussian exchangeable sequences. Since the distribution of a Gaussian sequence is determined by the covariance structure, the results above enable us to describe explicitly the Gaussian exchangeable sequences. Let $X_0, X_1, X_2, \ldots$ be independent $N(0,1)$. The general N-exchangeable Gaussian sequence is, in distribution, of the form

$$(1.9) \qquad Z_i = a + bX_i + c \sum_{j=1}^{N} X_j \, , \quad 1 \leq i \leq N$$

for some constants a, b, c. The general infinite exchangeable sequence is, in distribution, of the form

$$(1.10) \qquad Z_i = a + bX_0 + cX_i \, , \quad i \geq 1$$

for some constants a, b, c.

Extendibility. An N-exchangeable sequence (Z_i) is M-extendible $(M > N)$ if $(Z_1, \ldots, Z_N) \stackrel{\mathcal{D}}{=} (\hat{Z}_1, \ldots, \hat{Z}_N)$ for some M-exchangeable sequence $(\hat{Z}_i)$. By (1.7) there exist N-exchangeable sequences which are not (N+1)-extendible; for example, sampling without replacement from an urn with N elements. This suggests several problems, which we state rather vaguely.

(1.11) Problem. Find effective criteria for deciding whether a given N-exchangeable sequence is M-extendible.

(1.12) Problem. What proportion of N-exchangeable sequences are M-extendible?

Such problems seem difficult. Some results can be found in Diaconis (1977), Crisma (1982) and Spizzichino (1982).

Combinatorial arguments. Many identities and inequalities for i.i.d. sequences are proved by combinatorial arguments which remain valid for exchangeable sequences. Such results are scattered in the literature; for a selection, see Kingman (1978) Section 1 and Marshall and Olkin (1979).

2. Mixtures of i.i.d. sequences

Everyone agrees on how to say de Finetti's theorem in words:

"An infinite exchangeable sequence is a mixture of i.i.d. sequences."

But there are several mathematical formalizations (at first sight different, though in fact equivalent) of the theorem in the literature, because the concept of "a mixture of i.i.d. sequences" can be defined in several ways. Our strategy is to discuss this concept in detail in this section, and defer discussion of exchangeability and de Finetti's theorem until the next section.

Let $\theta_1,\ldots,\theta_k$ be probability distributions on R, and let $p_1,\ldots,p_k > 0$, $\sum p_i = 1$. Then we can describe a sequence (Y_i) by the two-stage procedure:

(2.1) (i) Pick θ at random from $\{\theta_1,\ldots,\theta_k\}$, $P(\theta = \theta_i) = p_i$;

(ii) then let (Y_i) be i.i.d. with distribution θ.

More generally, write $\mathcal{P}$ for the set of probability measure on $\tilde{R}$, let Θ be a distribution on $\mathcal{P}$, and replace (i) by

(i') Pick θ at random from distribution Θ.

Here we are merely giving the familiar Bayesian idea that (Y_i) is i.i.d. (θ), where θ has a prior distribution Θ. The easiest way to formalize this verbal description is to say

$$(2.2) \qquad P(\underset{\sim}{Y} \in A) = \int_P \theta^\infty(A)\Theta(d\theta) \ ; \quad A \subset R^\infty$$

where $\underset{\sim}{Y} = (Y_1, Y_2, \ldots)$, regarded as a random variable with values in R^∞, and $\theta^\infty = \theta \times \theta \times \cdots$ is the distribution on R^∞ of an i.i.d. (θ) sequence. This describes the distribution of a sequence which is a mixture of i.i.d. sequences. This is a special case of a general idea. Given a family $\{\mu_\gamma : \gamma \in \Gamma\}$ of distributions on a space S, call a distribution ν a mixture of (μ_γ)'s if

$$(2.3) \qquad \nu(\cdot) = \int_\Gamma \mu_\gamma(\cdot)\Theta(d\gamma) \quad \text{for some distribution } \Theta \text{ on } \Gamma \ .$$

But in practice it is much more convenient to use a definition of "(Y_i) is a mixture of i.i.d. sequences" which involves the random variables Y_i explicitly. To do so, we need a brief digression to discuss random measures and regular conditional distributions.

A random measure α is simply a P-valued random variable. So for each ω there is a probability measure $\alpha(\omega)$ and this assigns probability $\alpha(\omega, A)$ to subsets $A \subset R$. To make this definition precise we need to specify a σ-field on P: the natural σ-field is that generated by the maps

$$\theta \to \theta(A) \ ; \quad \text{measurable } A \subset R \ .$$

The technicalities about measurability in P that we need are straightforward and will be omitted. We may equivalently define a random measure as a function $\alpha(\omega, A)$, $\omega \in \Omega$, $A \subset R$, such that

$$\alpha(\omega,\cdot) \text{ is a probability measure; each } \omega \in \Omega .$$

$$\alpha(\cdot,A) \text{ is a random variable; each } A \subset R .$$

Say $\alpha_1 = \alpha_2$ a.s. if they are a.s. equal as random variables in P; or equivalently, if $\alpha_1(\cdot,A) = \alpha_2(\cdot,A)$ a.s. for each $A \subset R$.

Given a real-valued random variable Y and a σ-field F, a <u>regular conditional distribution</u> (r.c.d.) for Y given F is a random measure α such that

$$\alpha(\cdot,A) = P(Y \in A|F) \text{ a.s., each } A \subset R .$$

It is well known that r.c.d.'s exist, are a.s. unique, and satisfy the <u>fundamental property</u>

(2.4) $E(g(X,Y)|F) = \int g(X,y)\alpha(\omega,dy)$ a.s.; $X \in F$, $g(X,Y)$ integrable.

We now come to the key idea of this section. Given a random measure α, it is possible to construct (Y_i) such that conditional on $\alpha = \theta$ (where θ denotes a generic probability distribution), the sequence (Y_i) is i.i.d. with distribution θ. One way of doing so is to formalize the required properties of (Y_i) in an abstract way (2.6) and appeal to abstract existence theorems to show that random variables with the required properties exist. We prefer to give a concrete construction first.

Let $F(\theta,t) = \theta(-\infty,t]$ be the distribution function of θ, and let $F^{-1}(\theta,x) = \inf\{t: F(\theta,t) \geq x\}$ be the inverse distribution function. It is well known that if ξ is uniform on $(0,1)$ ("ξ is $U(0,1)$") then $F^{-1}(\theta,\xi)$ is a random variable with distribution θ. So if (ξ_i) is an i.i.d. $U(0,1)$ sequence then $(F^{-1}(\theta,\xi_i))$ is an i.i.d. (θ) sequence. Now given a random measure α, take (ξ_i) as above, independent of α, and let

(2.5) $$\hat{Y}_i = F^{-1}(\alpha,\xi_i) .$$

This construction captures the intuitive idea that, conditional on $\alpha = \theta$, the variables $\hat{Y}_i$ are i.i.d. (θ). The abstract properties of $(\hat{Y}_i, i \geq 1; \alpha)$ are given in

(2.6) Definition. Let α be a random measure and let $\underset{\sim}{Y} = (Y_i)$ be a sequence of random varibles. Say $\underset{\sim}{Y}$ is a mixture of i.i.d.'s directed by α if

$$(\alpha(\omega))^{\infty} \text{ is a r.c.d. for } \underset{\sim}{Y} \text{ given } \sigma(\alpha) .$$

Plainly this implies that the distribution of $\underset{\sim}{Y}$ is of the form (2.2), where Θ is the distribution of α. We remark that this idea can be abstracted to the general setting of (2.3); X is a mixture of $(\mu_\gamma: \gamma \in \Gamma)$ directed by a random element $\beta: \Omega \to \Gamma$ if $\mu_{\beta(\omega)}$ is a r.c.d. for X given $\sigma(\beta)$. Think of this as the "strong" notion of mixture corresponding to the "weak" notion (2.3).

The condition in (2.6) is equivalent to

(2.6a) $\quad P(Y_i \in A_i,\ 1 \leq i \leq n | \alpha) = \prod_i \alpha(\omega, A_i)\ ; \quad \text{all } A_1, \ldots, A_n, \quad n \geq 1 .$

And this splits into two conditions, as follows.

(2.7) Lemma. Write $F = \sigma(\alpha)$. Then $\underset{\sim}{Y}$ is a mixture of i.i.d.'s directed by α iff

(2.8) $(Y_i: i \geq 1)$ are conditionally independent given F, that is

$P(Y_i \in A_i,\ 1 \leq i \leq n | F) = \prod_i P(Y_i \in A_i | F).$

(2.9) the conditional distribution of Y_i given F is α; that is,

$P(Y_i \in A_i | F) = \alpha(\omega, A_i).$

Readers unfamiliar with the concept of conditional independence defined in (2.8) should consult the Appendix, which lists properties (A1)-(A9) and references.

Lemma 2.7 suggests a definition of "conditionally i.i.d." without explicit refernece to a random measure.

(2.10) Definition. Let (Y_i) be random variables and let F be a σ-field. Say (Y_i) is conditionally i.i.d. given F if (2.8) holds and if

(2.11) $$P(Y_i \in A|F) = P(Y_j \in A|F) \text{ a.s., each } A, \quad i \neq j.$$

Here is a useful technical lemma.

(2.12) Lemma. Suppose (Y_i) are conditionally i.i.d. given F. Let α be a r.c.d. for Y_1 given F. Then

(a) (Y_i) is a mixture of i.i.d.'s directed by α.

(b) $\underset{\sim}{Y}$ and F are conditionally independent given α.

Proof. By (2.11), for each i we have that α is a r.c.d. for Y_i given F. So by (2.8), $P(Y_i \in A_i, 1 \leq i \leq n|F) = \prod_i \alpha(\cdot, A_i)$. Now α is F-measurable, so conditioning on α gives $P(Y_i \in A_i, 1 \leq i \leq n|\alpha) = \prod_i \alpha(\cdot, A_i)$. This implies (a) by (2.6a). And it also implies

$$P(\underset{\sim}{Y} \in A|F) = P(\underset{\sim}{Y} \in A|\alpha) \text{ a.s.,} \quad A \subset R^\infty,$$

which gives (b) by a standard property of conditional independence (A3).

We have been stating results for infinite sequences (Y_i), but the results so far are true for finite sequences also. Suppose now we are told that a sequence (Y_i) is a mixture of i.i.d.'s for some unspecified random measure α. Can we determine α from (Y_i)? For finite sequences the answer is no, in general (4.7). But for infinite sequences we can. Define $\Lambda_n: R^n \longrightarrow P$ and $\Lambda: R^\infty \longrightarrow P$ by

(2.13) $\Lambda_n(x_1,\ldots,x_n) = n^{-1} \sum_i \delta_{x_i}$, the empirical distribution of (x_i);

(2.14) $$\Lambda(\underset{\sim}{x}) = \text{weak-limit}_{n\to\infty} \Lambda_n(x_1,\ldots,x_n)$$
$$= \delta_0 \text{ , say, if the limit does not exist,}$$

(δ_x denotes the degenerate distribution $\delta_x(A) = 1_{(x\in A)}$). If $\underset{\sim}{X} = (X_i)$ is an infinite i.i.d. (θ) sequence, then the Glivenko-Cantelli theorem says that $\Lambda(\underset{\sim}{X}) = \theta$ a.s. Thus for a mixture $(\hat{Y}_i)$ of i.i.d.'s directed by α we have $\Lambda(\hat{\underset{\sim}{Y}}) = \alpha$ a.s. by (2.6) and the fundamental property of r.c.d.'s. Hence

(2.15) Lemma. *If the infinite sequence $\underset{\sim}{Y}$ is a mixture of i.i.d.'s, then it is directed by $\alpha = \Lambda(\underset{\sim}{Y})$, and this directing random measure is a.s. unique.*

So we can talk about "the" directing random measure. Since $\Lambda(\underset{\sim}{x})$ is unchanged by changing a finite number of coordinates of $\underset{\sim}{x}$, we have

(2.16) Lemma. *Let the infinite sequence $\underset{\sim}{Y}$ be a mixture of i.i.d.'s, let α be the directing random measure. Then α is essentially T-measurable, where T is the tail σ-field of $\underset{\sim}{Y}$.*

Keeping the notation of Lemma 2.16, we see that the following are a.s. equal.

(2.17) (a) $P(Y_i \in A_i,\ 1 \le i \le n | Y_m, Y_{m+1}, \ldots)$, $m > n$

(b) $P(Y_i \in A_i,\ 1 \le i \le n | Y_m, Y_{m+1}, \ldots; \alpha)$, $m > n$

(c) $P(Y_i \in A_i,\ 1 \le i \le n | T)$

(d) $P(Y_i \in A_i,\ 1 \le i \le n \mid \alpha)$

(e) $\prod_i \alpha(\cdot, A_i)$.

Indeed, (a) = (b) since the conditioning is the same by Lemma 2.16; (b) = (d) by conditional independence; (d) = (e) by Lemma 2.7; and (b) = (c) = (d) since $\sigma(\alpha) \subset \mathcal{T} \subset \sigma(Y_m, Y_{m+1}, \ldots; \alpha)$ a.s. These identities establish the next lemmas.

(2.18) Lemma. (Y_i) is a mixture of i.i.d.'s if and only if (Y_i) is conditionally i.i.d. given $\mathcal{T}$.

(2.19) Lemma. If the infinite sequence $\underset{\sim}{Y}$ is a mixture of i.i.d.'s then it is directed by

(a) a r.c.d. for Y_1 given $\mathcal{T}$;

(b) a r.c.d. for Y_1 given $(Y_{m+1}, Y_{m+2}, \ldots)$, $m \ge 1$.

Observe that Lemmas 2.15 and 2.19 provide three ways to obtain (in principle, at least) the directing random measure. Each of these ways is useful in some circumstances.

Facts about mixtures of i.i.d.'s are almost always easiest to prove by conditioning on the directing measure. Let us spell this out in detail. For bounded $g: R \to R$ define $\bar{g}: \mathcal{P} \to R$ by

$$\bar{g}(\theta) = \int g \, d\theta \, . \tag{2.20}$$

By (2.9) and the fundamental property of r.c.d.'s

$$E(g(Y_i) \mid \alpha) = \bar{g}(\alpha) \text{ a.s., and hence } Eg(Y_i) = E\bar{g}(\alpha) \, . \tag{2.21}$$

And using the conditional independence (2.8), (2.22) $E(g_1(Y_i)g_2(Y_j)|\alpha) = \bar{g}_1(\alpha)\bar{g}_2(\alpha)$ a.s., $i \neq j$, and hence

$$Eg_1(Y_i)g_2(Y_j) = E\bar{g}_1(\alpha)\bar{g}_2(\alpha) .$$

These extend to unbounded g provided $|g(Y_1)|$, $|g_1(Y_i)g_2(Y_j)|$ are integrable. Let us use these to record some technical facts about moments of Y_i and α. For a distribution θ let $\text{mean}(\theta)$, $\text{var}(\theta)$, $\text{abs}_r(\theta)$ denote the mean, variance and r^{th} absolute moment of θ. Let $V(Y)$ denote the variance of a random variable Y. The next lemma gives properties which follow immediately from (2.21) and (2.22).

(2.23) Lemma.

(a) $E|Y_i|^r = E\ \text{abs}_r(\alpha)$.

(b) *If* $E|Y_i|$ *is finite, then* $EY_i = E\ \text{mean}(\alpha)$.

(c) *If* EY_i^2 *is finite, then*

(i) $EY_iY_j = E(\text{mean}(\alpha))^2$, $i \neq j$;

(ii) $V(Y_i) = E\ \text{abs}_2(\alpha) - (E\ \text{mean}(\alpha))^2 = E\ \text{var}(\alpha) + V(\text{mean}(\alpha))$.

Most classical limit theorems for i.i.d. sequences extend immediately to mixtures of i.i.d. sequences. For instance, writing $S_n = Y_1 + \cdots + Y_n$,

(2.24) $$\lim_{n\to\infty} n^{-1}S_n = \text{mean}(\alpha) \text{ a.s., provided } E|Y_1| < \infty .$$

(2.25) $$\limsup_{n\to\infty} \frac{S_n - n\cdot\text{mean}(\alpha)}{\{\text{var}(\alpha)\cdot 2n \log\log(n)\}^{1/2}} = 1 \text{ a.s., provided } E\,Y_1^2 < \infty .$$

(2.26) $$\frac{S_n - n\cdot\text{mean}(\alpha)}{\{n\ \text{var}(\alpha)\}^{1/2}} \xrightarrow{\mathcal{D}} \text{Normal}(0,1) \text{ as } n \to \infty, \text{ provided } E\,Y_1^2 < \infty .$$

Let us spell out some details. The fundamental property of r.c.d.'s says $P(\lim n^{-1}S_n = \text{mean}(\alpha)|\alpha) = h(\alpha)$, where $h(\theta) = P(\lim n^{-1}(X_1+\cdots+X_n) = \text{mean}(\theta))$ for (X_i) i.i.d. (θ). The strong law of large numbers says $h(\theta) = 1$

provided $abs_1(\theta) < \infty$. So (2.24) holds provided $abs_1(\alpha) < \infty$ a.s., and by Lemma 2.23 this is a consequence of $E|Y_1| < \infty$. The same argument works for (2.25), and for any a.s. convergence theorem for i.i.d. variables. For weak convergence theorems like (2.26), one more step is needed. Let $g: R \rightarrow R$ be bounded continuous. Then

$$E(g(\frac{S_n - n \cdot mean(\alpha)}{\{n\ var(\alpha)\}^{1/2}})|\alpha) = g_n(\alpha)$$

where

$$g_n(\theta) = E\ g(\frac{\sum_1^n X_i - n \cdot mean(\theta)}{\{n\ var(\theta)\}^{1/2}}) \quad \text{for } (X_i) \text{ i.i.d. } (\theta).$$

The central limit theorem says $g_n(\theta) \rightarrow Eg(V)$ when $abs_2(\theta) < \infty$, where V indicates a $N(0,1)$ variable. Hence

$$E\ g(\frac{S_n - n \cdot mean(\alpha)}{\{n\ var(\alpha)\}^{1/2}}) \rightarrow Eg(V)$$

provided $abs_2(\alpha) < \infty$ a.s., which by Lemma 2.23 is a consequence of $E\ Y_1^2 < \infty$. The same technique works for any weak convergence theorem for i.i.d. variables.

The form of results obtained in this way is slightly unusual, in that random normalization is involved, but they can easily be translated into a more familiar form. To ease notation, suppose $EY_1^2 < \infty$ and $mean(\alpha) = 0$ a.s. (which by Lemma 2.23 is equivalent to assuming $EY_i = 0$ and (Y_i) uncorrelated). Then (2.25) translates immediately to

$$\limsup_{n\to\infty} \frac{S_n}{\{2n \log\log(n)\}^{1/2}} = \{var(\alpha)\}^{1/2} \text{ a.s.}$$

And (2.26) translates to

(2.27) $$n^{-1/2}S_n \xrightarrow{\mathcal{D}} V\cdot\{\mathrm{var}(\alpha)\}^{1/2} ;$$

where V is Normal(0,1) independent of α. To see this, observe that the argument for (2.26) gives

$$(\mathrm{var}(\alpha), \frac{S_n - n\cdot\mathrm{mean}(\alpha)}{\{n\ \mathrm{var}(\alpha)\}^{1/2}}) \xrightarrow{\mathcal{D}} (\mathrm{var}(\alpha), V) ,$$

and then the continuous mapping theorem gives (2.27). Finally, keeping the assumptions above, (2.22) shows that $\mathrm{var}(\alpha) = \sigma^2$, constant, iff $EY_i^2 = \sigma^2$ and $EY_i^2Y_j^2 = \sigma^4$, so that these extra assumptions are what is needed to obtain $n^{-1/2}S_n \xrightarrow{\mathcal{D}} N(0,\sigma^2)$. Weak convergence theorems for mixtures of i.i.d. sequences are a simple class of <u>stable</u> convergence theorems, described further in Section 7.

Finally, we should point out there are occasional subtleties in extending results from i.i.d. sequences to mixtures. Consider the weak law of large numbers. The set $\mathcal{S}$ of distributions θ such that

(2.28) $$n^{-1}\sum_1^n X_i \xrightarrow{p} 0 \quad \text{for } (X_i) \text{ i.i.d. } (\theta)$$

is known. For a mixture of i.i.d. sequences to satisfy this weak law, it is certainly <u>sufficient</u> that the directing random measure satisfies $\alpha(\omega) \in \mathcal{S}$ a.s., by the usual conditioning argument. But this is not necessary, because (informally speaking) we can arrange the mixture to satisfy the weak law although α takes values θ such that convergence in (2.28) holds as $n \to \infty$ through some set of integers with asymptotic density one but not when $n \to \infty$ through all integers. (My thanks to Mike Klass for this observation.)

Another instance where the results for mixtures are more complicated is the estimation of L^p norms for weighted sums $\sum a_iX_i$. Such estimates are given by Dacunha-Castelle and Schreiber (1974) in connection with Banach space questions; see also (7.21).

We end with an intriguing open problem.

(2.29) Problem. Let $S_n = \sum_{i=1}^{n} X_i$, $\hat{S}_n = \sum_{i=1}^{n} \hat{X}_i$, where each of the sequences (X_i), $(\hat{X}_i)$ is a mixture of i.i.d. sequences. Suppose $S_n \stackrel{\mathcal{D}}{=} \hat{S}_n$ for each n. Does this imply $(X_i) \stackrel{\mathcal{D}}{=} (\hat{X}_i)$?

3. de Finetti's theorem

Our verbal description of de Finetti's theorem is now a precise assertion, which we restate as

(3.1) de Finetti's Theorem. *An infinite exchangeable sequence* (Z_i) *is a mixture of i.i.d. sequences.*

Remarks

(a) The converse, that a mixture of i.i.d.'s is exchangeable, is plain.

(b) As noted in Section 1, a finite exchangeable sequence may not be to an infinite sequence, and so a finite exchangeable sequence may not be a mixture of i.i.d.'s.

(c) The directing random measure can in principle be obtained from Lemma 2.15 or 2.19.

Most modern proofs of de Finetti's theorem rely on martingale convergence. We shall present both the standard proof (which goes back at least to Loève (1960)) and also a more sophisticated variant. Both proofs contain useful techniques which will be used later in other settings.

First proof of Theorem 3.1. Let

$$G_n = \sigma\{f_n(Z_1,\ldots,Z_n): f_n \text{ symmetric}\}$$
$$H_n = \sigma\{G_n, Z_{n+1}, Z_{n+2},\ldots\} .$$

Then $H_n \supset H_{n+1}$, $n \geq 1$. Exchangeability implies

$$(Z_i,Y) \overset{\mathcal{D}}{=} (Z_1,Y) , \quad 1 \leq i \leq n ,$$

for Y of the form $(f_n(Z_1,\ldots,Z_n), Z_{n+1}, Z_{n+2},\ldots)$, f_n symmetric, and hence for all $Y \in H_n$. So for bounded ϕ,

$$\begin{aligned} E(\phi(Z_1)|H_n) &= E(\phi(Z_i)|H_n) , \quad 1 \leq i \leq n \\ &= E(n^{-1}\sum_{i=1}^{n}\phi(Z_i)|H_n) \end{aligned}$$

(3.2) $$= n^{-1}\sum_{i=1}^{n}\phi(Z_i) , \quad \text{since this is } G_n\text{-measurable.}$$

The reversed martingale convergence theorem implies

(3.3) $$n^{-1}\sum_{i=1}^{n}\phi(Z_i) \rightarrow E(\phi(Z_1)|H) \text{ a.s.,} \quad \text{where } H = \bigcap_n H_n .$$

For bounded $\phi(x_1,\ldots,x_k)$, the argument for (3.2) shows that for $n > k$

$$E(\phi(Z_1,\ldots,Z_k)|H_n) = \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\phi(Z_{j_1},\ldots,Z_{j_k})1_{D_{k,n}}$$

where $D_{k,n} = \{(j_1,\ldots,j_k): 1 \leq j_r \leq n, (j_r) \text{ distinct}\}$. Using martingale convergence, and the fact that $\#D_{k,n}$ is $O(n^{k-1})$ as $n \rightarrow \infty$ for fixed k,

$$n^{-k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\phi(Z_{j_1},\ldots,Z_{j_k}) \rightarrow E(\phi(Z_1,\ldots,Z_k)|H) \text{ a.s.}$$

By considering $\phi(x_1,\ldots,x_k)$ of the form $\phi_1(x_1)\phi_2(x_2)\cdots\phi_k(x_k)$ and using (3.3),

$$E(\prod_{r=1}^{k} \phi_r(Z_r)|H) = \prod_{r=1}^{k} E(\phi_r(Z_1)|H) .$$

This says that (Z_i) is conditionally i.i.d. given H, and as discussed in Section 2 this is one of several equivalent formalizations of "mixture of i.i.d. sequences".

For the second proof we need an easy lemma.

(3.4) Lemma. Let Y be a bounded real-valued random variable, and let $F \subset G$ be σ-fields. If $E(E(Y|G))^2 = E(E(Y|F))^2$, in particular if $E(Y|F) \overset{\mathcal{D}}{=} E(Y|G)$, then $E(Y|G) = E(Y|F)$ a.s.

Proof. This is immediate from the identity

$$E(E(Y|G) - E(Y|F))^2 = E(E(Y|G))^2 - E(E(Y|F))^2 .$$

Second Proof of Theorem 3.1. Write $F_n = \sigma(Z_{n+1}, Z_{n+2}, \ldots)$, so $\bigcap_n F_n = T$, the tail σ-field. We shall show that (Z_i) is conditionally i.i.d. given T. By exchangeability, $(Z_1, Z_2, Z_3, \ldots) \overset{\mathcal{D}}{=} (Z_1, Z_{n+1}, Z_{n+2}, \ldots)$ and so $E(\phi(Z_1)|F_2) \overset{\mathcal{D}}{=} E(\phi(Z_1)|F_n)$ for each bounded $\phi: R \to R$. The reversed martingale convergence theorem says $E(\phi(Z_1)|F_n) \to E(\phi(Z_1)|T)$ a.s. as $n \to \infty$, and so $E(\phi(Z_1)|F_2) \overset{\mathcal{D}}{=} E(\phi(Z_1)|T)$. But Lemma 3.4 now implies there is a.s. equality, and this means (A4)

Z_1 and F_2 are conditionally independent given T.

The same argument applied to $(Z_m, Z_{m+1}, \ldots)$ gives

Z_m and F_{m+1} are conditionally independent given T; $m \geq 1$.

These imply that the whole sequence $(Z_i;\ i \geq 1)$ is conditionally independent given T.

For each $n \geq 1$, exchangeability says $(Z_1, Z_{n+1}, Z_{n+2}, \ldots) \stackrel{\mathcal{D}}{=} (Z_n, Z_{n+1}, Z_{n+2}, \ldots)$, and so $E(\phi(Z_1)|F_n) = E(\phi(Z_n)|F_n)$ a.s. for each bounded ϕ. Conditioning on T gives $E(\phi(Z_1)|T) = E(\phi(Z_n)|T)$ a.s. This is (2.11), and so $(Z_i;\ i \geq 1)$ is indeed conditionally i.i.d. given T.

Spherically symmetric sequences. Another classical result can be regarded as a specialization of de Finetti's theorem. Call a random vector $Y^n = (Y_1, \ldots, Y_n)$ spherically symmetric if $UY^n = Y^n$ for each orthogonal $n \times n$ matrix U. Call an infinite sequence $\underset{\sim}{Y}$ spherically symmetric if Y^n is spherically symmetric for each n. It is easy to check that an i.i.d. Normal $N(0,v)$ sequence is spherically symmetric; and hence so is a mixture (over v) of i.i.d. $N(0,v)$ sequences. On the other hand, computations with characteristic functions (Feller (1971) Section III.4) give

(3.5) Maxwell's Theorem. An independent spherically symmetric sequence has $N(0,\sigma^2)$ distribution, for some $\sigma^2 \geq 0$.

Now a spherically symmetric sequence is exchangeable, since for any permutation π of $\{1, \ldots, n\}$, the map $(y_1, \ldots, y_n) \to (y_{\pi(1)}, \ldots, y_{\pi(n)})$ is a rotation. We shall show that Maxwell's theorem and de Finetti's theorem imply

(3.6) Schoenberg's Theorem. An infinite spherically symmetric sequence $\underset{\sim}{Y}$ is a mixture of i.i.d. $N(0,\sigma^2)$ sequences.

This is apparently due to Schoenberg (1938), and has been rediscovered many times. See Eaton (1981), Letac (1981a) for variations and references. This result also fits naturally into the "sufficient statistics" setting of Section 18.

Proof. Let U be a $n \times n$ orthogonal matrix, let $Y^n = (Y_1,\dots,Y_n)$, $\hat{Y}^{n,m} = (Y_{n+1},\dots,Y_{n+m})$. By considering $\begin{pmatrix} U & 0 \\ 0 & I_m \end{pmatrix}$, an orthogonal $(m+n)\times(m+n)$ matrix, we see $(UY^n,\tilde{Y}^{n,m}) \overset{\mathcal{D}}{=} (Y^n,\tilde{Y}^{n,m})$. Letting $m \to \infty$ gives $(UY^n,\tilde{Y}^n) \overset{\mathcal{D}}{=} (Y^n,\tilde{Y}^n)$, where $\tilde{Y}^n = (Y_{n+1},Y_{n+2},\dots)$. By conditioning on the tail σ-field $\mathcal{T} \subset \sigma(Y^n)$, we see that the conditional distribution of Y^n and of UY^n given $\mathcal{T}$ are a.s. identical. Applying this to a countable dense set of orthogonal $n \times n$ matrices and to each $n \geq 1$, we see that the conditional distribution of $\underset{\sim}{Y}$ given $\mathcal{T}$ is a.s. spherically symmetric. But de Finetti's theorem says that the conditional distribution of $\underset{\sim}{Y}$ given $\mathcal{T}$ is a.s. an i.i.d. sequence, so the result follows from Maxwell's theorem.

Here is a slight variation on de Finetti's theorem. Given an exchangeable sequence Z, say Z _is exchangeable over_ V if

$$(3.7)\quad (V,Z_1,Z_2,\dots) \overset{\mathcal{D}}{=} (V,Z_{\pi(1)},Z_{\pi(2)},\dots)\ , \quad \text{all finite permutations } \pi.$$

Similarly, say Z is exchangeable over a σ-field $\mathcal{G}$ if (3.7) holds for each $V \in \mathcal{G}$.

(3.8) _Proposition_. _Let_ $\underset{\sim}{Z}$ _be an infinite sequence, exchangeable over_ V. _Then_ (a) (Z_i) _is conditionally i.i.d. given_ (V,α), _where_ α _is the directing random measure for_ $\underset{\sim}{Z}$.

(b) $\underset{\sim}{Z}$ _and_ V _are conditionally independent given_ α.

Proof. Let $\hat{Z}_i = (V,Z_i)$. Then $\underset{\sim}{\hat{Z}}$ is exchangeable, so de Finetti's theorem implies

$$(\hat{Z}_i) \text{ is conditionally i.i.d. given } \hat{\alpha}\ ,$$

where $\hat{\alpha}$ is the directing random measure for $\underset{\sim}{\hat{Z}}$. So in particular,

$$(Z_i) \text{ is conditionally i.i.d. given } \hat{\alpha} .$$

But applying Lemma 2.15 to $\hat{\underset{\sim}{Z}}$, we see $\hat{\alpha} = \delta_V \times \alpha$, and so $\sigma(\hat{\alpha}) = \sigma(V,\alpha)$. This gives (a). And (b) follows from (a) and Lemma 2.12(b).

We remark that the conditional independence assertion of (b) is a special case of a very general result, Proposition 12.12. Proposition 3.8 plays an important role in the study of partial exchangeability in Part III. As a simple example, here is a version of de Finetti's theorem for a family of sequences, where each is "internally exchangeable".

(3.9) Corollary. *For* $1 \le j \le k$ *let* $Z^j = (Z_i^j,\ i \ge 1)$ *be exchangeable. Suppose further that for each* j_0 *and each finite permutation* π *we have* $(Z_i^j) \overset{\mathcal{D}}{=} (\hat{Z}_i^j)$, *where*

$$\begin{aligned} \hat{Z}_i^j &= Z_i^j \quad , \quad j \ne j_0 \\ &= \hat{Z}_{\pi(i)}^j \ , \quad j = j_0 \ . \end{aligned}$$

Let α_j *be the directing measure for* Z^j, *and let* $F = \sigma(\alpha_j,\ 1 \le j \le k)$. *Then* (a) $(Z_i^j\colon 1 \le j \le k,\ i \ge 1)$ *are conditionally independent given* F.

(b) α_j *is a r.c.d. for* Z_i^j *given* F.

This result goes back to de Finetti, and has been rediscovered many times. In Bayesian language, the family is obtained as follows.

(i) Pick a k-tuple $(\theta_1,\ldots,\theta_k)$ of distributions according to a prior Θ on $\mathcal{P}^k$;

(ii) then for each j let the sequence $(Z_i^j,\ i \ge 1)$ be i.i.d. (θ_j), independent for different j.

Proof. Fix j. Proposition 3.8 shows

Z^j and $\sigma(Z^m\colon m \neq j)$ are conditionally independent given α_j.

Then de Finetti's theorem for Z^j yields

$Z_1^j, Z_2^j, \ldots;$ $\sigma(Z^m\colon m \neq j)$ are conditionally independent given α_j;
α_j is a r.c.d. for Z_i^j given α_j.

Since $\sigma(\alpha_j) \subset F \subset \sigma(\alpha_j, Z^m\colon m \neq j)$ this is equivalent to

$Z_1^j, Z_2^j, \ldots;$ $\sigma(Z^m\colon m \neq j)$ are conditionally independent given F;
α_j is a r.c.d. for Z_i^j given F.

Since j is arbitrary, this establishes the result.

Here is another application of Proposition 3.8. Call a subset B of R^∞ exchangeable if

$$(x_1, x_2, \ldots) \in B \quad \text{implies} \quad (x_{\pi(1)}, x_{\pi(2)}, \ldots) \in B$$

for each finite permutation π. Given an infinite sequence $\underset{\sim}{X} = (X_i)$, call events $\{\underset{\sim}{X} \in B\}$, B exchangeable, exchangeable events, and call the set of exchangeable events the exchangeable σ-field E_X. It is easy to check that $E_X \supset T_X$ a.s., where T_X is the tail σ-field of $\underset{\sim}{X}$.

(3.10) Corollary. If $\underset{\sim}{Z}$ is exchangeable then $E_Z = T_Z = \sigma(\alpha)$ a.s.

Proof. For $A \in E_Z$ the random variable $V = 1_A$ satisfies (3.7), and so by Proposition 3.4, V and $\underset{\sim}{Z}$ are conditionally independent given α. Hence E_Z and $\underset{\sim}{Z}$ are conditionally independent given α. But $E_Z \subset \sigma(\underset{\sim}{Z})$ and so E_Z and E_Z are conditionally independent given α, which implies

(A6) that $E_Z \subset \sigma(\alpha)$ a.s. But $\sigma(\alpha) \subset T_Z$ a.s. by Lemma 2.16, and $T_Z \subset E_Z$ a.s.

In particular, Corollary 3.10 gives the well-known Hewitt-Savage 0-1 law:

(3.11) Corollary. If $\underset{\sim}{X}$ is i.i.d. then E_X is trivial.

This can be proved by more elementary methods (Breiman (1968), Section 3.9).

There are several other equivalences possible in Corollary 3.10; let us state two.

(3.12) Corollary. If $\underset{\sim}{Z}$ is exchangeable then $\sigma(\alpha)$ coincides a.s. with

(a) the invariant σ-field of $\underset{\sim}{Z}$

(b) the tail σ-field of $(Z_{n_1}, Z_{n_2}, \ldots)$, for any distinct (n_i).

In particular, for an exchangeable process $(Z_i : -\infty < i < \infty)$ the "left tail" and "right tail" σ-fields each coincides a.s. with $\sigma(\alpha)$. In the partial exchangeability setting of Part III, we shall see that subprocesses may generate different σ-fields.

Corollary 3.10 is one generalization of the Hewitt-Savage law. Another type of generalization is to consider which random sequences X have the property that E_X is trivial. For independent sequences, the condition

$$(3.13) \qquad \sum \text{variance}(\phi(X_i)) = 0 \text{ or } \infty \text{ ; each bounded } \phi$$

is necessary in order that E_X be trivial, since if (3.13) fails for ϕ then $\sum(\phi(X_i) - E\phi(X_i))$ defines a non-degenerate E_X-measurable random variable. For finite state space sequences, condition (3.13) is sufficient (see Aldous and Pitman (1979) for this and equivalent conditions). But still open is

(3.14) Problem. For an independent sequence (X_i) taking values in a countable set, is (3.13) sufficient for E_X to be trivial?

Results about E_X for Markovian sequences have been given by Blackwell and Freedman (1964), Grigorenko (1979), and Palacios (1982).

4. Exchangeable sequences and their directing random measures

Convention. In this section $\underset{\sim}{Z} = (Z_i)$ is a real-valued exchangeable infinite sequence directed by some random measure α.

The purpose of this section is to point out some concrete ways of constructing exchangeable sequences, and to investigate how particular properties of $\underset{\sim}{Z}$ correspond to particular properties of α. The results are mostly straightforward consequences of de Finetti's theorem, but will give the reader some experience in manipulating random measures.

There are two ways in which an exchangeable sequence may be considered "degenerate". First, if it is i.i.d. This corresponds to $\alpha = \theta$ a.s., for some fixed distribution θ. Second, if $Z_1 = Z_2 = \cdots$ a.s. This corresponds to $\alpha = \delta_X$ a.s. for some random variable X.

The "simple" exchangeable sequences described at (2.1)(i) are those with $\alpha = \theta_I$ a.s., where $\{\theta_1,\ldots,\theta_k\}$ are distributions and I is a random variable taking values in $\{1,\ldots,k\}$.

One natural way to construct an exchangeable sequence is to take a parametric family of distributions, choose the parameter randomly, and take an i.i.d. sequence whose distribution has this random parameter. For example, let $\mu_{\theta,s}$ denote the Normal(θ,s^2) distribution. For random (Θ,S), we can define an exchangeable sequence (Z_i) which is i.i.d. Normal(θ,s^2) conditional on $(\Theta=\theta,\ S=s)$. This is the exchangeable sequence directed by $\alpha = \mu_{\Theta,S}$. Because the Normal family is a location-scale family, we can construct (Z_i) very simply by putting $Z_i = \Theta + SX_i$, for (X_i) i.i.d. Normal(0,1).

In general, $\underset{\sim}{Z}$ is a <u>mixture</u> of a parametric family (μ_θ) if $\alpha \in (\mu_\theta)$ a.s. It is natural to ask for intrinsic conditions on $\underset{\sim}{Z}$ (rather than α) which determine whether $\underset{\sim}{Z}$ is a mixture of a specified family; results of this kind are given in Section 18. Such processes arise naturally in the Bayesian analysis of parametric statistical problems. For the Bayesian analysis of non-parametric problems, one needs tractable random measures whose values are not restricted to small subsets of P; the most popular are the Dirichlet random measures (Ferguson (1974)), described briefly in Section 10.

Here are two more ways of producing exchangeable sequences.

(4.1) Let $(Y_1, Y_2, \ldots)$ be arbitrary;
let $(X_1, X_2, \ldots)$ be i.i.d., independent of $\underset{\sim}{Y}$, taking values $\{1,2,\ldots\}$;
let $Z_i = Y_{X_i}$.

Then $\underset{\sim}{Z}$ is exchangeable, and using Lemma 2.15 we see $\alpha = \sum_i p_i \delta_{Y_i(\omega)}$, where $p_i = P(X_1 = i)$.

(4.2) Let $X_1, X_2, \ldots$ be i.i.d., distribution θ;
let Y be independent of $\underset{\sim}{X}$, with distribution ϕ;
let $Z_i = f(Y, X_i)$, for some function f.

Then $\underset{\sim}{Z}$ is exchangeable. Indeed, from the canonical construction (2.5) and de Finetti's theorem, every exchangeable sequence is of this form (in distribution). However, exchangeable sequences arising in practice can often be put into the form (4.2) where θ, ϕ, f have some simple form with intuitive significance (e.g. the representation (1.10) for Gaussian exchangeable sequences). To describe the directing random measure for such a sequence, we need some notation.

(4.3) Definition. Given $f: R \to R$ define the induced map $\tilde{f}: \mathcal{P} \to \mathcal{P}$ by

$$\tilde{f}(\mathcal{L}(Y)) = \mathcal{L}(f(Y)) .$$

Given $f: R \times R \to R$ define the induced map $\tilde{f}: R \times \mathcal{P} \to \mathcal{P}$ by

$$\tilde{f}(x,\mathcal{L}(Y)) = \mathcal{L}(f(x,Y)) .$$

This definition and Lemma 2.15 give the next lemma.

(4.4) Lemma. (a) *Let $\underset{\sim}{Z}$ be exchangeable, directed by α, let $f: R \to R$ have induced map $\tilde{f}$, and let $\hat{Z}_i = f(Z_i)$. Then $\underset{\sim}{\hat{Z}}$ is exchangeable and is directed by $\tilde{f}(\alpha)$.*

(b) *Let $\underset{\sim}{Z}$ be of the form (4.2) for some $f: R \times R \to R$, and let $\tilde{f}: R \times \mathcal{P} \to \mathcal{P}$ be the induced map. Then $\underset{\sim}{Z}$ is exchangeable and is directed by $\tilde{f}(Y,\theta)$.*

In particular, for the addition function $f(x,y) = x + y$ we have $\tilde{f}(x,\theta) = \delta_x * \theta$, where $*$ denotes convolution. So (1.10) implies:

(4.5) $\underset{\sim}{Z}$ is Gaussian if and only if $\alpha(\omega,\cdot) = \delta_{X(\omega)} * \theta$, where θ and $\mathcal{L}(X)$ are Normal.

Another simple special case is 0-1 valued exchangeable sequences. Call events $(A_i,\ i \geq 1)$ exchangeable if the indicator random variables $Z_i = 1_{A_i}$ are exchangeable. In this case α must have the form

$X(\omega)\delta_1 + (1 - X(\omega))\delta_0$, for some random variable $0 \le X \le 1$. Informally, conditional on $\{X = p\}$ the events (A_i) are independent and have probability p.

There is a curious connection between exchangeable events and a classical moment problem. Let $p_n = P(A_1 \cap A_2 \cap \cdots \cap A_n)$. Since $P(A_1 \cap \cdots \cap A_n | X) = X^n$ a.s. we have

(a) $p_n = EX^n$.

Now the distribution of the sequence (A_i) is, by exchangeability, determined by the numbers $q_{n,k} = P(A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c)$. But the relation $q_{n,k} = q_{n-1,k} - q_{n,k+1}$ shows that the numbers $(q_{n,k})$ are determined by (p_n) $(= q_{n,n})$. But the distribution of (A_i) determines the distribution of X, so

(b) the numbers (p_n) determine $L(X)$.

Since any distribution θ on $[0,1]$ is possible for X, facts (a) and (b) imply

(4.6) a distribution θ on $[0,1]$ is determined by its moments

$$p_n = \int x^n \theta(dx), \quad n \ge 1.$$

This is the classical "Hausdorff moment problem". The sequence (p_n) is completely monotone; for further discussion of monotonicity and exchangeability, see Kallenberg (1976) Chapter 9; Daboni (1982); Kimberling (1973).

(4.7) Remark. Given N we can find different distributions θ_1, θ_2 on $[0,1]$ such that $\int x^n \theta_1(dx) = \int x^n \theta_2(dx)$, $n \le N$. Consider the corresponding finite exchangeable sequences $(1_{A_1}, \ldots, 1_{A_N})$. These have the same distribution, by the argument above. Thus a finite exchangeable sequence may be extendible to more than one infinite exchangeable sequence.

We now consider covariance properties. Let $\underset{\sim}{Z}$ be exchangeable, and suppose $EZ_1^2 < \infty$. Let $\rho = E(Z_i - EZ_i)(EZ_j - EZ_j)$, $i \neq j$, be the covariance. In (1.8) it was proved directly that $\rho \geq 0$. But de Finetti's theorem gives more information: by Lemma 2.23,

(4.8) $\rho = 0$ if and only if $\text{mean}(\alpha) = c$ a.s., some constant c.

Of course, from de Finetti's theorem

(4.9) $$\text{mean}(\alpha) = E(Z_1|\alpha) = E(Z_1|T) \text{ a.s.}$$

In particular, an exchangeable $\underset{\sim}{Z}$ with $EZ_1 = 0$ is uncorrelated if and only if the random variable in (4.9) is a.s. zero. Curiously, this implies the (generally stronger) property that $\underset{\sim}{Z}$ is a martingale difference sequence.

(4.10) Lemma. *Suppose $\underset{\sim}{Z}$ is exchangeable, $EZ_i = 0$. Then $\underset{\sim}{Z}$ is a martingale difference sequence if and only if* $\text{mean}(\alpha) = 0$ a.s.

Proof. By conditional independence,

$$E(Z_n|Z_1,\ldots,Z_{n-1},\alpha) = E(Z_n|\alpha) = \text{mean}(\alpha) \text{ a.s.}$$

So if $\text{mean}(\alpha) = 0$ a.s. then $E(Z_n|Z_1,\ldots,Z_{n-1}) = 0$ a.s., and so $\underset{\sim}{Z}$ is a martingale difference sequence. Conversely, if $\underset{\sim}{Z}$ is a martingale difference sequence then $E(Z_n|Z_1,\ldots,Z_{n-1}) = 0$ a.s., so by exchangeability $E(Z_1|Z_2,\ldots,Z_n) = 0$ a.s. The martingale convergence theorem now implies $E(Z_1|Z_i, i > 1) = 0$ a.s., and since $\sigma(\alpha) \subset \sigma(Z_i, i > 1)$ we have $\text{mean}(\alpha) = E(Z_1|\alpha) = 0$ a.s.

Here is another instance where for exchangeable sequences one property implies a generally stronger property.

(4.11) Lemma. If an infinite exchangeable sequence $(Z_1,Z_2,\ldots)$ is pairwise independent then it is i.i.d.

Proof. Fix some bounded function $f: R \to R$. The sequence $\hat{Z}_i = f(Z_i)$ is pairwise independent, and hence uncorrelated. By Lemma 4.4(a), $\hat{\underset{\sim}{Z}}$ is directed by $\tilde{f}(\alpha)$, and now (4.8) implies $\text{mean}(\tilde{f}(\alpha))$ is a.s. constant, c_f say. In other words:

$$\int f(x)\alpha(\cdot,dx) = c_f \text{ a.s.; each bounded } f.$$

Standard arguments (7.12) show this implies $\alpha = \theta$ a.s., where θ is a distribution with $\int f(x)\theta(dx) \equiv c_f$. So $\underset{\sim}{Z}$ is i.i.d. (θ).

(4.12) Example. Fix $N > 2$. Let $(Y_1,\ldots,Y_N)$ be uniform on the set of sequences $(y_1,\ldots,y_N)$ of 1's and 0's satisfying $\sum y_i = 0 \bmod (2)$. Then

(a) $(Y_1,\ldots,Y_N)$ is N-exchangeable;

(b) $(Y_1,\ldots,Y_{N-1})$ are independent.

So Lemma 4.11 is not true for finite exchangeable sequences. And by considering $X_i = 2Y_i - 1$, we see that a finite exchangeable sequence may be uncorrelated but not a martingale difference sequence.

Our next topic is the Markov property.

(4.13) Lemma. For an infinite exchangeable sequence $\underset{\sim}{Z}$ the following are equivalent.

(a) $\underset{\sim}{Z}$ is Markov.

(b) $\sigma(\alpha) \subset \sigma(Z_1)$ a.s.

(c) $\sigma(\alpha) = \bigcap_i \sigma(Z_i)$ a.s.

Remark. When the support of α is some countable set (θ_j) of distributions, these conditions are equivalent to

(d) the distributions (θ_j) are mutually singular.

It seems hard to formalize this in the general case.

Proof. $\underset{\sim}{Z}$ is Markov if and only if for each bounded $\phi: R \to R$ and each $n \geq 2$,

(a') $E(\phi(Z_n)|Z_1,\ldots,Z_{n-1}) = E(\phi(Z_n)|Z_{n-1})$ a.s.

Suppose this holds. Then by exchangeability,

$$E(\phi(Z_2)|Z_1,Z_3,\ldots,Z_{n-1}) = E(\phi(Z_2)|Z_1) \text{ a.s.}$$

So by martingale convergence

$$E(\phi(Z_2)|Z_1) = E(\phi(Z_2)|Z_1,Z_3,Z_4,\ldots) = E(\phi(Z_2)|\alpha) .$$

In particular, $\alpha(\cdot,A) = P(Z_2 \in A|\alpha)$ is essentially $\sigma(Z_1)$-measurable, for each A. This gives (b).

If (b) holds then by symmetry $\sigma(\alpha) \subset \sigma(Z_i)$ a.s. for each i, and so $\sigma(\alpha) \subset \cap\, \sigma(Z_i)$ a.s. And Corollary 3.10 says $\sigma(\alpha) = \mathcal{T} \supset \cap\, \sigma(Z_i)$ a.s., which gives (c).

If (c) holds then

$$\begin{aligned} E(\phi(Z_n)|Z_1,\ldots,Z_{n-1}) &= E(\phi(Z_n)|Z_1,\ldots,Z_{n-1},\alpha) \quad \text{by (c)} \\ &= E(\phi(Z_n)|Z_{n-1},\alpha) \quad \text{by conditional independence} \\ &= E(\phi(Z_n)|Z_{n-1}) \quad \text{by (c)} \end{aligned}$$

and this is (a').

This is one situation where the behavior of finite exchangeable sequences is the same as for the infinite case, by the next result.

(4.14) Lemma. *Any finite Markov exchangeable sequence* $(Z_1,\ldots,Z_N)$ *extends to an infinite Markov exchangeable sequence* $\underset{\sim}{Z}$, *provided* $N \geq 3$.

This cannot be true for $N = 2$, since a 2-exchangeable sequence is vacuously Markov.

Proof. The given $(Z_1,\ldots,Z_N)$ extends to an infinite sequence $\underset{\sim}{Z}$ whose distribution is specified by the conditions

(4.15) (i) $\underset{\sim}{Z}$ is Markov;

(ii) $(Z_i,Z_{i+1}) \overset{\mathcal{D}}{=} (Z_1,Z_2)$.

We must prove $\underset{\sim}{Z}$ is exchangeable. Suppose, inductively, that $(Z_1,\ldots,Z_m)$ is m-exchangeable for some $m \geq N$. Then $(Z_2,\ldots,Z_m) \overset{\mathcal{D}}{=} (Z_1,\ldots,Z_{m-1})$, so by (i) and (ii) we get

$$(4.16) \qquad (Z_2,\ldots,Z_{m+1}) \overset{\mathcal{D}}{=} (Z_1,\ldots,Z_m) .$$

So these vectors are m-exchangeable. Hence $(Z_2,\ldots,Z_{m+1}) \overset{\mathcal{D}}{=} (Z_2,\ldots,Z_{m-1},Z_{m+1},Z_m)$ and so by the Markov property (at time 2)

$$(4.17a) \qquad (Z_1,\ldots,Z_{m+1}) \overset{\mathcal{D}}{=} (Z_1,\ldots,Z_{m-1},Z_{m+1},Z_m) .$$

Similarly, using the Markov property at time m,

$$(4.17b) \qquad (Z_1,\ldots,Z_{m+1}) \overset{\mathcal{D}}{=} (Z_2,Z_1,Z_3,\ldots,Z_{m+1}) .$$

Finally, for any permutation π of $(2,\ldots,m)$ we assert

$$(4.17c) \qquad (Z_1,Z_{\pi(2)},\ldots,Z_{\pi(m)},Z_{m+1}) \overset{\mathcal{D}}{=} (Z_1,\ldots,Z_{m+1}) .$$

For let $Y = (Z_2,\ldots,Z_m)$, $\hat{Y} = (Z_{\pi(2)},\ldots,Z_{\pi(m)})$. Then the triples (Z_1,Y,Z_{m+1}) and $(Z_1,\hat{Y},Z_{m+1})$ are Markov, and (4.16) shows $(Z_1,Y) \overset{\mathcal{D}}{=} (Z_1,\hat{Y})$

and $(Y,Z_{m+1}) \overset{\mathcal{D}}{=} (\hat{Y},Z_{m+1})$, which gives (4.17c). Now (4.17a-c) establish the (m+1)-exchangeability of $(Z_1,\ldots,Z_{m+1})$.

Let us digress slightly to present the following two results, due to Carnal (1980). Informally, these can be regarded as extensions of Lemmas 4.13 and 4.14 to non-exchangeable sequences.

(4.18) Lemma. *Let X_1, X_2, X_3 be random variables such that for any ordering (i,j,k) of $(1,2,3)$, X_i and X_j are conditionally independent given X_k. Then X_1, X_2, X_3 are conditionally independent given* $F = \sigma(X_1)\cap\sigma(X_2)\cap\sigma(X_3)$.

(4.19) Lemma. *For an infinite sequence $\underset{\sim}{X}$, the following are equivalent:*

(a) $(X_{n_1},X_{n_2},\ldots)$ *is Markov, for any distinct* $n_1,n_2,\ldots$

(b) $(X_i;\ i\geq 1)$ *are conditionally independent given* $G = \bigcap_i \sigma(X_i)$.

Proof of Lemma 4.18. We shall prove that for any ordering (i,j,k),

(a) $\sigma(X_i)\cap\sigma(X_j) = F$ a.s.;

(b) X_i and $\sigma(X_j,X_k)$ are conditionally independent given $\sigma(X_j)\cap\sigma(X_k)$.

These imply that X_i and $\sigma(X_j,X_k)$ are conditionally independent given F, and the lemma follows.

For $A\in\sigma(X_i)$ and $B\in\sigma(X_j)$, conditional independence given $P(A\cap B|X_k) = P(A|X_k)\cdot P(B|X_k)$. So for $A\in\sigma(X_i)\cap\sigma(X_j)$ we have $P(A|X_k) = \{P(A|X_k)\}^2$, so $A\in\sigma(X_k)$ a.s., giving (a).

For bounded $\phi: R\rightarrow R$, conditional independence gives

$$E(\phi(X_i)|X_j,X_k) = E(\phi(X_i)|X_j) = E(\phi(X_i)|X_k)\ .$$

So $E(\phi(X_i)|X_j,X_k)$ is essentially $\sigma(X_j)\cap\sigma(X_k)$-measurable, proving (b).

Proof of Lemma 4.19. Suppose (a) holds. For distinct i, j, k, the hypothesis of Lemma 4.18 holds, so by its conclusion

$$E(\phi(X_i)|X_j) = E(\phi(X_i)|\sigma(X_j)\cap\sigma(X_j)\cap\sigma(X_k)) .$$

Since this holds for each $k \neq i, j$,

$$E(\phi(X_i)|X_j) \in \bigcap_{m=1} \sigma(X_m) = G \text{ a.s.}$$

In other words, X_i and X_j are conditionally independent given G. But the Markov hypothesis implies that X_i and $\sigma(X_k;\ k \neq i)$ are conditionally independent given X_j. These last two facts imply (A7) that X_i and $\sigma(X_k;\ k \neq i)$ are conditionally independent given G, and the result follows.

5. Finite exchangeable sequences

As mentioned in Section 1, the basic way to obtain an N-exchangeable sequence is as an urn process: take N constants $y_1,\ldots,y_N$, not necessarily distinct, and put them in random order.

(5.1) $$\underset{\sim}{Y} = (Y_1,\ldots,Y_N) = (y_{\hat{\pi}(1)},\ldots,y_{\hat{\pi}(N)}) ;$$

$\hat{\pi}$ the uniform random permutation on $\{1,\ldots,N\}$. In the notation of (2.13), $\underset{\sim}{Y}$ has empirical distribution

(5.2) $$\Lambda_N(\underset{\sim}{Y}) = \frac{1}{N}\sum \delta_{y_i} = \Lambda_N(\underset{\sim}{y}) .$$

Conversely, it is clear that:

(5.3) if $\underset{\sim}{Y}$ is N-exchangeable and satisfies (5.2) then $\underset{\sim}{Y}$ has distribution (5.1).

Let U_N denote the set of distributions $L(\underset{\sim}{Y})$ for urn processes $\underset{\sim}{Y}$.

Let $\mathcal{U}_N^*$ denote the set of empirical distributions $\frac{1}{N}\sum\delta_{y_i}$. Let $\Phi\colon \mathcal{U}_N \to \mathcal{U}_N^*$ be the natural bijection $\Phi(L(\underset{\sim}{Y})) = \Lambda_N(\underset{\sim}{Y})$. The following simple result is a partial analogue of de Finetti's theorem in the finite case.

(5.4) <u>Lemma</u>. <u>Let</u> $\underset{\sim}{Z} = (Z_1,\ldots,Z_N)$ <u>be</u> <u>N-exchangeable</u>. <u>Then</u> $\Phi^{-1}(\Lambda_N(\underset{\sim}{Z}))$ <u>is a regular conditional distribution for</u> $\underset{\sim}{Z}$ <u>given</u> $\Lambda_N(\underset{\sim}{Z})$.

In words: conditional on the empirical distribution, the N (possibly repeated) values comprising the empirical distribution occur in random order. In the real-valued case we can replace "empirical distribution" by "order statistics", which convey the same information.

<u>Proof</u>. For any permutation π of $\{1,\ldots,N\}$,

$$(Z_1,\ldots,Z_N,\Lambda_N(\underset{\sim}{Z})) \overset{\mathcal{D}}{=} (Z_{\pi(1)},\ldots,Z_{\pi(N)},\Lambda_N(\underset{\sim}{Z})) .$$

So if $\beta(\omega,\cdot)$ is a regular conditional distribution for $\underset{\sim}{Z}$ given $\Lambda_N(\underset{\sim}{Z})$ then (a.s. ω)

(a) the distribution $\beta(\omega,\cdot)$ is N-exchangeable.

But from the fundamental property of r.c.d.'s, for a.s. ω we have

(b) the N-vector with distribution $\beta(\omega,\cdot)$ has empirical distribution $\Lambda_N(\underset{\sim}{Z}(\omega))$.

And (5.3) says that (a) and (b) imply $\beta(\omega,\cdot) = \Phi^{-1}(\Lambda_N(\underset{\sim}{Z}(\omega)))$ a.s.

Thus for some purposes the study of finite exchangeable sequences reduces to the study of "sampling without replacement" sequences of the form (5.1). Unlike de Finetti's theorem, this idea is not always useful: for example, it does not seem to help with the weak convergence problems discussed in Section 20.

From an abstract viewpoint, the difference between finite and infinite exchangeability is that the group of permutations on a finite set is compact, whereas on an infinite set it is non-compact. Lemma 5.4 has an analogue for distributions invariant under a specified compact group of transformations; see (12.15).

We know that an N-exchangeable sequence need not be a mixture of i.i.d.'s, that is to say it need not extend to an infinite exchangeable sequence. But we can ask how "close" it is to some mixture of i.i.d.'s. Let us measure closeness of two distributions μ, ν by the total variation distance

$$(5.5) \qquad \|\mu-\nu\| = \sup_A |\mu(A)-\nu(A)| .$$

The next result implies that an M-exchangeable sequence which can be extended to an N-exchangeable sequence, where N is large compared to M^2, is close to a mixture of i.i.d.'s.

(5.6) Proposition. *Let* $\underset{\sim}{Y}$ *be N-exchangeable. Then there exists an infinite exchangeable sequence* $\underset{\sim}{Z}$ *such that, for* $1 \le M \le N$,

$$\|L(Y_1,\ldots,Y_M) - L(Z_1,\ldots,Z_M)\| \le 1 - \prod_{i=1}^{M-1} (1 - i/N) \le \frac{M(M-1)}{2N} .$$

This is one formalization of the familiar fact that sampling with replacement and without replacement are almost equivalent when the sample size is small compared to the population size. See Proposition 20.6 for another formalization.

Proof. We quote the straightforward estimate

$$(5.7) \qquad \|L(V) - L(V|B)\| \le 1 - P(B) \; ; \quad \text{all random variables } V, \text{ events } B$$

where $L(V|B)$ is the conditional distribution. Let $(I_1, I_2, \ldots)$ be i.i.d. uniform on $\{1,2,\ldots,N\}$. Let $B_{M,N}$ be the event $\{I_1,\ldots,I_M$ all distinct$\}$. Then

$$(5.8) \qquad P(B_{M,N}) = \prod_{i=1}^{M-1} (1 - i/N) .$$

Let $Z_i = Y_{I_i}$, $i \geq 1$. Then (Z_i) is an infinite exchangeable sequence. Since, by exchangeability of Y,

$$(Y_{j_1},\ldots,Y_{j_M}) \overset{\mathcal{D}}{=} (Y_1,\ldots,Y_M) \; ; \quad \text{any distinct } (j_k) ,$$

we see that

$$L((Z_1,\ldots,Z_M)|B_{M,N}) = L(Y_1,\ldots,Y_M) .$$

Now (5.7) and (5.8) establish the first inequality of the Proposition; the second is calculus.

Better bounds can be obtained if there are bounds on the number of possible values of Y_i, but more delicate arguments are required. We quote a result from Diaconis and Freedman (1980a), which also contains Proposition 5.6 and further discussion.

(5.9) Proposition. *Let $(Y_1,\ldots,Y_N)$ be N-exchangeable, taking values in a set of cardinality c. Then there exists an infinite exchangeable sequence $\underset{\sim}{Z}$ such that for $1 \leq M \leq N$*

$$\| L(Y_1,\ldots,Y_M) - L(Z_1,\ldots,Z_M) \| \leq cM/N .$$

Diaconis and Freedman (unpublished) also have a similar result relating to Schoenberg's theorem.

(5.10) Proposition. Let $(Y_1,\ldots,Y_N)$ be spherically symmetric. Then there exists a sequence $\underset{\sim}{Z}$ which is a mixture of i.i.d. $N(0,\sigma^2)$ sequences such that

$$\|L(Y_1,\ldots,Y_M) - L(Z_1,\ldots,Z_M)\| \le bM/N \ , \quad 1 \le M \le N \ ,$$

where b is a constant not depending on (Y_i).

The obvious way of getting M-exchangeable sequences from N-exchangeable sequences (N > M) is by taking the first M variables; Proposition 5.6 and the discussion of extendibility in Section 1 show that the M-exchangeable sequences obtainable in this way are restricted. Here is another way of getting new exchangeable sequences from old. Let $N, K \ge 1$. Let $Z = (Z_i : 1 \le i \le KN)$ be exchangeable, and let $f: R^N \to R$ be a function. Define

$$(5.11) \qquad \hat{Y}_j = f(Z_{(j-1)N+1},\ldots,Z_{jN}) \ , \quad 1 \le j \le K \ .$$

Then $\hat{Y} = (\hat{Y}_j : 1 \le j \le K)$ is exchangeable. Now any given K-exchangeable sequence Y may or may not have the property

(5.12) for each N there exists a NK-exchangeable Z and a function f such that $Y \overset{\mathcal{D}}{=} \hat{Y}$, for $\hat{Y}$ defined at (5.11).

This can be regarded as a non-linear analogue of "infinite divisibility".

(5.13) Problem. Give an intrinsic characterization of K-exchangeable sequences Y with property (5.12).

(5.14) Example. Let (Y_1,Y_2,Y_3) be the urn sequence from urn $\{a,b,c\}$. Then (5.12) fails for N = 2. Here is an outline of the argument. Suppose (5.12) held, so some $(f(Z_1,Z_2),f(Z_3,Z_4),f(Z_5,Z_6))$ is a random ordering of $\{a,b,c\}$. Using Lemma 5.4, we may suppose (Z_i) is an urn process, with urn (z_i) say. Since $f(Z_5,Z_6)$ is determined by (Z_1,Z_2,Z_3,Z_4), we may take f symmetric. Now picture a, b, c as colors; consider the complete graph on the 6 points (z_i) and paint the edge (z_i,z_j) with color $f(z_i,z_j)$. Then edges with distinct endpoints must have different colors, and it is easy to verify this is impossible.

(5.15) Example. The Gaussian exchangeable sequences (1.9) do have property (5.12).

Finally, let us mention a curious result given in Dellacherie and Meyer (1980), V.51: any finite exchangeable sequence is a "mixture" of i.i.d. sequences, if we allow the mixing measure to be a signed measure.

PART II

In Part II we present those extensions and analogues of de Finetti's theorem which are close in spirit to the theorem itself; subsequent parts will take us further afield.

6. Properties equivalent to exchangeability

In Section 1 we pointed out some conditions which were trivially equivalent to exchangeability. The next result collects together some remarkable, non-trivial equivalences.

(6.1) Theorem. *For an infinite sequence of random variables* $Z = (Z_i)$, *each of the following conditions is equivalent to exchangeability:*

(A) $Z \stackrel{\mathcal{D}}{=} (Z_{n_1}, Z_{n_2}, \dots)$ *for each increasing sequence* $1 \le n_1 < n_2 < \cdots$ *of constants.*

(B) $Z \stackrel{\mathcal{D}}{=} (Z_{T_1+1}, Z_{T_2+1}, Z_{T_3+1}, \dots)$ *for each increasing sequence* $0 \le T_1 < T_2 < \cdots$ *of stopping times.*

(C) $Z \stackrel{\mathcal{D}}{=} (Z_{T+1}, Z_{T+2}, Z_{T+3}, \dots)$ *for each stopping time* $T \ge 0$

(*Stopping times are relative to the filtration* $F_n = \sigma(Z_1, \dots, Z_n)$, F_0 *trivial.*)

Remarks

(a) Ryll-Nardzewski (1957) proved (A) implies exchangeability. Property (A), under the name "spreading-invariance", arises naturally in the work of Dacunha-Castelle and others who have studied certain Banach space problems using probabilistic techniques. A good survey of this area is Dacunha-Castelle (1982).

(b) The fact that (B) and (C) are equivalent to exchangeability is due to Kallenberg (1982a), who calls property (C) "strong stationarity". The idea of expressing exchangeability-type properties in terms of stopping times seems a promising technique for the study of exchangeability concepts for continuous-time processes, where there is a well-developed technical machinery involving stopping times. See Section 17 for one such study.

(c) Stopping times of the form $T+1$ are predictable stopping times.

(d) The difficult part is proving these conditions imply exchangeability. Let us state the (vague) question

(6.2) Problem. What hypotheses *prima facie* weaker than exchangeability do in fact imply exchangeability?

The best result known seems to be that obtained by combining Lemma 6.5 and Proposition 6.4 below, which are taken from Aldous (1982b).

For the proof of Theorem 6.1 we need the following extension of Lemma 3.4.

(6.3) Lemma. *Let $(\mathcal{G}_n)$ be an increasing sequence of σ-fields, let $\mathcal{G} = \bigvee_n \mathcal{G}_n$, and let $F \subset \mathcal{G}$. Let Y be a bounded random variable such that for each n there exists $F_n \subset F$ such that $E(Y|F_n) \stackrel{\mathcal{D}}{=} E(Y|\mathcal{G}_n)$. Then $E(Y|F) = E(Y|\mathcal{G})$ a.s.*

Proof. Write $\|U\| = EU^2$. Then $\|E(Y|\mathcal{G}_n)\| = \|E(Y|F_n)\| \le \|E(Y|F)\|$. Since $E(Y|\mathcal{G}_n) \to E(Y|\mathcal{G})$ in L^2 by martingale convergence, we obtain $\|E(Y|\mathcal{G})\| \le \|E(Y|F)\|$. But $F \subset \mathcal{G}$ implies $\|E(Y|F)\| \le \|E(Y|\mathcal{G})\|$. So $\|E(Y|\mathcal{G})\| = \|E(Y|F)\|$, and now Lemma 3.4 establishes the result.

(6.4) Proposition. *Let $\underset{\sim}{X}$ be an infinite sequence with tail σ-field $\mathcal{T}$. Suppose that for each $j, k \ge 1$ there exist $n_1,\dots,n_k$ such that $n_i > i$*

<u>and</u> $(X_j, X_{j+1}, \ldots, X_{j+k}) \overset{\mathcal{D}}{=} (X_j, X_{j+n_1}, \ldots, X_{j+n_k})$. <u>Then</u> $(X_i;\ i \geq 1)$ <u>are conditionally independent given</u> $\mathcal{T}$.

<u>Proof</u>. Fix $m, n \geq 1$, let $F = \sigma(X_m, X_{m+1}, \ldots)$ and let $G_n = \sigma(X_2, \ldots, X_n)$. By repeatedly applying the hypothesis, there exist $q_2, \ldots, q_n$ such that $q_i > m$ and $(X_1, \ldots, X_n) \overset{\mathcal{D}}{=} (X_1, X_{q_2}, \ldots, X_{q_n})$. So for bounded $\phi: R \to R$ we have $E(\phi(X_1)|G_n) \overset{\mathcal{D}}{=} E(\phi(X_1)|F_n)$, where $F_n = \sigma(X_{q_2}, \ldots, X_{q_n}) \subset F$. Applying Lemma 6.3,

$$\begin{aligned} E(\phi(X_1)|X_2, X_3, \ldots) &= E(\phi(X_1)|X_m, X_{m+1}, \ldots) \text{ a.s.} \\ &= E(\phi(X_1)|\mathcal{T}) \text{ a.s. by martingale convergence.} \end{aligned}$$

This says that X_1 and $\sigma(X_2, X_3, \ldots)$ are conditionally independent given $\mathcal{T}$. But for each j the sequence $(X_j, X_{j+1}, \ldots)$ satisfies the hypotheses of the Proposition, so X_j and $\sigma(X_{j+1}, X_{j+2}, \ldots)$ are conditionally independent given $\mathcal{T}$. This establishes the result.

(6.5) <u>Lemma</u>. <u>Let</u> X <u>be an infinite sequence with tail</u> <u>σ-field</u> $\mathcal{T}$. <u>Suppose</u>

$$(X_1, X_{n+1}, X_{n+2}, X_{n+3}, \ldots) \overset{\mathcal{D}}{=} (X_n, X_{n+1}, X_{n+2}, X_{n+3}, \ldots); \ \underline{\text{each}}\ n \geq 1.$$

<u>Then the random variables</u> X_i <u>are conditionally identically distributed given</u> $\mathcal{T}$, <u>that is</u> $E(\phi(X_1)|\mathcal{T}) = E(\phi(X_n)|\mathcal{T})$, <u>each</u> $n \geq 1$, ϕ <u>bounded</u>.

<u>Proof</u>. By hypothesis $E(\phi(X_1)|X_{n+1}, X_{n+2}, \ldots) = E(\phi(X_n)|X_{n+1}, X_{n+2}, \ldots)$. Condition on $\mathcal{T}$.

<u>Proof of Theorem 6.1</u>. It is well known (and easy) that an i.i.d. sequence has property (B). It is also easy to check that any stopping time T on (F_n) can be taken to have the form $T = t(Z_1, Z_2, \ldots)$, where the function $t(\underset{\sim}{x})$ satisfies the condition

(*) if $t(\underset{\sim}{x}) = n$ and $x_i' = x_i$, $i \le n$, then $t(\underset{\sim}{x}') = n$.

Let Z be exchangeable, directed by α say, and let $T_r = t_r(Z)$ be an increasing sequence of stopping times. Conditional on α, the variables Z_i are i.i.d. and the times (T_r) are stopping times, since the property (*) is unaffected by conditioning. Thus we can apply the i.i.d. result to see that conditional on α the distributions of Z and $(Z_{T_1+1}, Z_{T_2+1}, \ldots)$ are identical; hence the unconditional distributions are identical. Thus exchangeability implies (B).

Plainly property (B) implies (A) and (C). It remains to show that each of the properties (A), (C) implies both the hypotheses of (6.4) and (6.5), so that (Z_i) are conditionally i.i.d. given $\mathcal{T}$, and hence exchangeable. For (A) these implications are obvious. So suppose (C) holds. Let $j, k, n \ge 1$. Applying (C) to the stopping times S, T, where $S = j$ and

$$T = j \quad \text{on } \{Z_j \in F\},$$
$$= j+n \quad \text{on } \{Z_j \notin F\},$$

we have $(Z_{j+1}, Z_{j+2}, \ldots) \overset{\mathcal{D}}{=} (Z_{T+1}, Z_{T+2}, \ldots)$. Since these vectors are identical on $\{Z_j \in F\}$, the conditional distributions given $\{Z_j \notin F\}$ must be the same. That is, conditional on $\{Z_j \notin F\}$ the distributions of $(Z_{j+1}, Z_{j+2}, \ldots)$ and of $(Z_{j+n+1}, Z_{j+n+2}, \ldots)$ are the same. Since F is arbitrary, we obtain

(6.6) $$(Z_j, Z_{j+1}, Z_{j+2}, \ldots, Z_{j+k}) \overset{\mathcal{D}}{=} (Z_j, Z_{j+n+1}, \ldots, Z_{j+n+k}),$$

and this implies the hypothesis of (6.4). Finally, property (C) with $T = 1$ shows Z is stationary, so

$$(Z_{j+n}, Z_{j+n+1}, \ldots, Z_{j+n+k}) \overset{\mathcal{D}}{=} (Z_j, Z_{j+1}, \ldots, Z_{j+k})$$
$$\overset{\mathcal{D}}{=} (Z_j, Z_{j+n+1}, \ldots, Z_{j+n+k}) \quad \text{by (6.6)}$$

and this gives the hypothesis of (6.5).

Remark. Here we have used Proposition 6.3 for sequences which eventually turn out to be exchangeable. However, it can also give information for sequences which are in fact not exchangeable; see (14.7).

Finite exchangeable sequences. For a finite sequence $(Z_1,\dots,Z_N)$ conditions (A)-(C) do not imply exchangeability. For instance, when $N = 2$ they merely imply $Z_1 \stackrel{\mathcal{D}}{=} Z_2$. On the other hand an exchangeable sequence obviously satisfies (A); what is less obvious is that (B) (and hence (C)) holds in the finite case, where the argument used in the infinite case based on de Finetti's theorem cannot be used.

(6.7) Proposition. Let $(Z_1,\dots,Z_N)$ be exchangeable and let $0 \leq T_1 < T_2 < \cdots < T_k < N$ be stopping times. Then $(Z_{T_1+1},\dots,Z_{T_k+1}) \stackrel{\mathcal{D}}{=} (Z_1,\dots,Z_k)$.

This result is related to a gambling game called "play red". I shuffle a standard deck of cards (52 cards; 26 red and 26 black) and slowly deal them out, face up so you can see them. At some time, you bet that the next card will be red; and you must make this bet sometime before all the cards are dealt. What is your best strategy for deciding when to make the bet? One strategy is to bet on the first card, which gives you chance 1/2 of winning: is there a better strategy? By counting the colors of the cards already dealt, you can know the proportion of red cards remaining in the deck, so a natural strategy is to wait until this proportion is greater than 1/2 and then bet; intuitively, this should give you a chance greater than 1/2 of winning. However, this intuitive argument is wrong. Let Z_i be the i^{th} card dealt, and let T be the time you decide to bet; then Proposition 6.7 says that the next card Z_{T+1} has the same distribution as Z_1, that is uniform over the deck of 52 cards.

Proposition 6.7 is due to Kallenberg (1982a). The fact that $Z_{T_1+1} \stackrel{\mathcal{D}}{=} Z_1$ is easy. For the process

$$P(Z_{i+1} \in A | Z_1, \ldots, Z_i) = P(Z_N \in A | Z_1, \ldots, Z_i)$$

is a martingale, so for a stopping time $T < N$

$$\begin{aligned} P(Z_{T+1} \in A) &= E\, P(Z_{T+1} \in A | Z_1, \ldots, Z_T) \\ &= E\, P(Z_N \in A | Z_1, \ldots, Z_T) \\ &= P(Z_N \in A) \quad \text{by the optional sampling theorem.} \end{aligned}$$

However, making an honest proof of the k-stopping-times result requires some effort; we take a slightly different approach. Recall that (Z_i) is exchangeable over V if $(V, Z_1, \ldots, Z_N) \stackrel{\mathcal{D}}{=} (V, Z_{\pi(1)}, \ldots, Z_{\pi(N)})$ for all permutations π. The following lemma is immediate.

(6.8) Lemma. Let $(Z_1, \ldots, Z_N)$ be exchangeable over V, let $0 \le i \le N$, let $A \in \sigma(V, Z_1, \ldots, Z_i)$ and let $V^i = (V, Z_1, \ldots, Z_i)$, $Z^i_j = Z_{i+j}$, $1 \le j \le N-i$. Then conditional on A, $(Z^i_1, \ldots, Z^i_{N-1})$ is exchangeable over V^i.

We now establish Proposition 6.7 by proving, by induction on $k \ge 1$, the following more general fact.

(6.9)(k) Assertion. Whenever (Z_i) is exchangeable over V and $0 \le T_1 < \cdots < T_k$ are stopping times relative to $\mathcal{G}_n = \sigma(V, Z_1, \ldots, Z_n)$, then

$$(V, Z_{T_1+1}, \ldots, Z_{T_k+1}) \stackrel{\mathcal{D}}{=} (V, Z_{N-k+1}, \ldots, Z_N) .$$

Proof. $P(V\in A, Z_{T_1+1}\in B) = \sum_i P(V\in A, Z_{i+1}\in B, T_1 = i)$

$$= \sum_i P(V\in A, Z_1^i\in B|T_1 = i)P(T_1 = i)$$

$$= \sum P(V\in A, Z_{N-i}^i\in B|T_1 = i)P(T_1 = i) \quad \text{by (6.8)}$$

$$= \sum_i P(V\in A, Z_N\in B, T_1 = i)$$

$$= P(V\in A, Z_N\in B) ,$$

establishing (6.9) for $k = 1$. Suppose (6.9) holds for some k. We shall prove it for $k+1$. Consider first the special case where $T_1 = 0$. Then the sequence $(Z_2,\ldots,Z_N)$ is exchangeable over (V,Z_1), so

$$(V,Z_1,Z_{T_2+1},\ldots,Z_{T_{k+1}+1}) \overset{\mathcal{D}}{=} (V,Z_1,Z_{N-k+1},\ldots,Z_N) \quad \text{by (6.9) for } k$$
$$\overset{\mathcal{D}}{=} (V,Z_{N-k},\ldots,Z_N) \quad \text{by exchangeability over } V,$$

establishing (6.9) for $k+1$ in the special case $T_1 = 0$. In the general case, fix i, and define V^i, Z_j^i as at (6.8). On the set $\{T_1 = i\}$ we have $T_j = i+\hat{T}_j$, where $\hat{T}_1 = 0$ and $\hat{T}_j$ is a stopping time with respect to $G_n^i = \sigma(V^i,Z_1^i,\ldots,Z_n^i) = G_{i+n}$. So by (6.8) and the special case,

$$(V^i,Z^i_{\hat{T}_1+1},\ldots,Z^i_{\hat{T}_{k+1}+1}) \overset{\mathcal{D}}{=} (V^i,Z^i_{N-i-k},\ldots,Z^i_{N-i}), \quad \text{conditional on } \{T_1 = i\}.$$

This implies

$$(V,Z_{T_1+1},\ldots,Z_{T_{k+1}+1}) \overset{\mathcal{D}}{=} (V,Z_{N-k},\ldots,Z_N), \quad \text{conditional on } \{T_1 = i\}.$$

Since this holds for each i, it holds unconditionally, establishing (6.9) for $k+1$.

7. Abstract spaces

Let S be an arbitrary measurable space. For a sequence $\underset{\sim}{Z} = (Z_1, Z_2, \ldots)$ of S-valued random variables the definition (1.2) of "exchangeable" and the definition (2.6) of "mixture of i.i.d.'s" make sense. So we can ask whether de Finetti's theorem is true for S-valued sequences, i.e. whether these definitions are equivalent. Dubins and Freedman (1979) give an example to show that for general S de Finetti's theorem is false: an exchangeable sequence need not be a mixture of i.i.d. sequences. See also Freedman (1980). But, loosely speaking, de Finetti's theorem _is_ true for "non-pathological" spaces. One way to try to prove this would be to examine the proof of the theorem for R and consider what abstract properties of the range space S were needed to make the proof work for S. However, there is a much simpler technique which enables results for real-valued processes to be extended without effort to a large class of abstract spaces. We now describe this technique.

(7.1) _Definition_. Spaces S_1, S_2 are _Borel-isomorphic_ if there exists a bijection $\phi: S_1 \to S_2$ such that ϕ and ϕ^{-1} are measurable. A space S is a _Borel_ (or _standard_) space if it is Borel-isomorphic to some Borel-measurable subset of R.

It is well known (see e.g. Breiman (1968) A7) that any Polish (i.e. complete separate metric) space is Borel; in particular R^n, R^∞ and the familiar function spaces $C(0,1)$ and $D(0,1)$ are Borel. Restricting attention to Borel spaces costs us some generality; for instance, the general compact Hausdorf space is not Borel, and it is known (Diaconis and Freedman (1980a)) that de Finetti's theorem is true for compact Hausdorf spaces, but has the great advantage that results extend automatically from the real-valued setting to the Borel space-valued setting.

We need some notation. Let $P(S)$ denote the set of probability measures on S. As at (4.3), for functions $f\colon S_1 \to S_2$ or $g\colon S_1 \times S_2 \to S_3$ define the _induced_ maps $\tilde{f}\colon P(S_1) \to P(S_2)$, $\tilde{g}\colon S_1 \times P(S_2) \to P(S_3)$ by

$$\tilde{f}(L(Y)) = L(f(Y)) \ , \quad \tilde{g}(x, L(Y)) = L(g(x,Y)) \ . \tag{7.3}$$

(7.4) _Proposition. Let_ $\underset{\sim}{Z}$ _be an infinite exchangeable sequence, taking values in a Borel space_ S. _Then_ $\underset{\sim}{Z}$ _is a mixture of i.i.d. sequences._

Proof. Let $\phi\colon S \to B$ be an isomorphism as in (7.1) between S and a Borel subset B of R. Let $\underset{\sim}{\hat{Z}}$ be the real-valued sequence $(\phi(Z_i))$. Then $\underset{\sim}{\hat{Z}}$ is exchangeable, so by the result (3.1) for the real-valued case, $\underset{\sim}{\hat{Z}}$ is a mixture of i.i.d. sequences, directed by a random measure $\hat{\alpha}$, say. Since $\hat{Z}_i \in B$ we have $\hat{\alpha}(\cdot, B) = 1$ a.s., so we may regard $\hat{\alpha}$ as $P(B)$-valued. The map $\psi = \phi^{-1}\colon B \to S$ induces a map $\hat{\psi}\colon P(B) \to P(S)$, and $\alpha = \hat{\psi}(\hat{\alpha})$ defines a random measure on S. It is straightforward to check that $\underset{\sim}{Z}$ is a mixture of i.i.d.'s directed by α.

Exactly the same arguments show that all our results for real-valued exchangeable sequences which involve only "measure-theoretic" properties of R can be extended to S-valued sequences. We shall not write them all out explicitly. Let us just mention two facts.

(7.5) There exists a regular conditional distribution for any S-valued random variable given any σ-field.

(7.6) Let ξ be $U(0,1)$. For any distribution μ on S there exists $f\colon (0,1) \to S$ such that $f(\xi)$ has distribution μ.

Topological spaces. To discuss convergence results we need a topology on the range space S. We shall simply make the

Convention. All abstract spaces S mentioned are assumed to be Polish.

Roughly speaking, convergence results for real-valued exchangeable processes extend to the Polish space setting.

Let us record some notation and basic facts about weak convergence in a Polish space S. We assume the reader has some familiarity with this topic (see e.g. Billingsley (1968); Parthasarathy (1967)).

For bounded $f: S \to R$ write

(7.7) $$\bar{f}(\theta) = \int f(x)\theta(dx) .$$

Let $C(S)$ be the set of bounded continuous functions $f: S \to R$. Give $P(S)$ the topology of weak convergence:

$$\theta_n \to \theta \quad \text{iff} \quad \bar{f}(\theta_n) \to \bar{f}(\theta); \quad \text{each} \quad f \in C(S) .$$

The space $P(S)$ itself is Polish: if d is a bounded complete metric on S then

(7.8) $$\tilde{d}(\mu,\nu) = \inf\{Ed(X,Y): L(X) = \mu,\ L(Y) = \nu\}$$

defines a complete metrization of $P(S)$.

(7.9) Skorohod Representation Theorem. *Given $\theta_n \to \theta$, we can construct random variables $X_n \to X$ a.s. and such that $L(X_n) = \theta_n$, $L(X) = \theta$.*

A sequence (θ_n) is relatively compact iff it is tight, that is for each $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset S$ such that $\inf_n \theta_n(K_\varepsilon) \geq 1 - \varepsilon$. There exists a countable subset H of $C(S)$ which is convergence-determining:

(7.10) $$\text{if } \lim_n \bar{h}(\theta_n) = \bar{h}(\theta), \quad h \in H, \quad \text{then } \theta_n \to \theta.$$

In particular H is <u>determining</u>:

(7.11) $$\text{if } \bar{h}(\theta) = \bar{h}(\mu), \quad h \in H, \quad \text{then } \theta = \mu.$$

For a random measure α on S, that is to say a $P(S)$-valued random variable, and for bounded $h: S \to R$, the expression $\bar{h}(\alpha)$ gives the real-valued random variable $\int h(x)\alpha(\cdot,dx)$. By (7.11), if

(7.12) $$\bar{h}(\alpha_1) = \bar{h}(\alpha_2) \text{ a.s.}, \quad h \in H,$$

where H is a countable determining class, then $\alpha_1 = \alpha_2$ a.s.

For a random measure α on S define

(7.13) $$\bar{\alpha}(A) = E\alpha(\cdot,A), \quad A \subset S,$$

so $\bar{\alpha}$ is a distribution on S.

Here is a technical lemma.

(7.14) <u>Lemma</u>. <u>Let</u> (α_n) <u>be random measures on</u> S.

(a) <u>If</u> $(\bar{\alpha}_n)$ <u>is tight on</u> $P(S)$ <u>then</u> $(L(\alpha_n))$ <u>is tight in</u> $P(P(S))$.

(b) <u>If</u> (α_n) <u>is a martingale, in the sense that</u>

$$E(\alpha_{n+1}(\cdot,B)|F_n) = \alpha_n(\cdot,B) \text{ a.s.}; \quad B \subset S, \quad n \geq 1,$$

<u>for some increasing</u> σ-<u>fields</u> (F_n), <u>then there exists a random measure</u> β <u>such that</u> $\alpha_n \to \beta$ <u>a.s., that is</u>

$$P(\omega: \alpha_n(\omega,\cdot) \to \beta(\omega,\cdot) \text{ in } P(S)) = 1 .$$

<u>Proof</u>. (a) Fix $\varepsilon > 0$. By hypothesis there exist compact $K_j \subset S$ such that

(7.15) $$\bar{\alpha}_n(K_j^c) \le \varepsilon 2^{-2j}; \quad j, n \ge 1.$$

So by Markov's inequality

(7.16) $$P(\alpha_n(\cdot, K_j^c) > 2^{-j}) \le \varepsilon 2^{-2j}/2^{-j} = \varepsilon 2^{-j}; \quad j, n \ge 1.$$

So, setting

(7.17) $$\Theta = \{\theta: \theta(K_j^c) \le 2^{-j}, \text{ all } j \ge 1\},$$

we have from (7.16)

$$P(\alpha_n \in \Theta) \ge 1 - \varepsilon; \quad n \ge 1.$$

Since Θ is a compact subset of $P(S)$, this establishes (a).

(b) For each $h \in C(S)$ the sequence $\bar{h}(\alpha_n)$ is a real-valued martingale. So for a countable convergence-determining class H we have (a.s.)

$$\lim_{n\to\infty} \bar{h}(\alpha_n(\omega)) \text{ exists, each } h \in H.$$

Thus it suffices to prove that a.s.

(7.18) the sequence of distributions $\alpha_n(\omega, \cdot)$ is tight.

By the martingale property, $\bar{\alpha}_n$ does not depend on n. Take (K_j) as at (7.15). Using the maximal inequality for the martingale $\alpha_n(\cdot, K_j^c)$ gives

$$P(\alpha_n(\cdot, K_j^c) > 2^{-j} \text{ for some } n) \le \varepsilon 2^{-j}.$$

So for Θ as at (7.17),

$$P(\omega: \alpha_n(\omega, \cdot) \in \Theta \text{ for all } n) \ge 1 - \varepsilon .$$

This establishes (7.18).

Weak convergence of exchangeable processes. First observe that the class of exchangeable processes is closed under weak convergence. To say this precisely, suppose that for each $k \geq 1$ we have an infinite exchangeable (resp. N-exchangeable) sequence $\underset{\sim}{Z}^k = (Z_1^k)$. Think of $\underset{\sim}{Z}^k$ as a random element of S^∞ (resp. S^N), where this product space has the product technology. If $\underset{\sim}{Z}^k \xrightarrow{\mathcal{D}} \underset{\sim}{X}$, which in the infinite case is equivalent to

$$(7.19) \qquad (Z_1^k,\ldots,Z_m^k) \xrightarrow{\mathcal{D}} (X_1,\ldots,X_m) \quad \text{as} \quad k \to \infty; \quad \text{each} \quad m \geq 1,$$

then plainly $\underset{\sim}{X}$ is exchangeable. Note that by using interpretation (7.19) we can also talk about $\underset{\sim}{Z}^k \xrightarrow{\mathcal{D}} \underset{\sim}{X}$ where $\underset{\sim}{Z}^k$ is N^k-exchangeable, $N_k \to \infty$, and $\underset{\sim}{X}$ is infinite exchangeable.

Note also that tightness of a family $(\underset{\sim}{Z}^k)$ of exchangeable processes is equivalent to tightness of (Z_1^k). Given some class of exchangeable processes, one can consider the "weak closure" of the class, i.e. the (necessarily exchangeable) processes which are weak limits of processes from the given class.

We know that the distribution of an infinite (resp. finite) exchangeable process $\underset{\sim}{Z}$ is determined by the distribution of the directing random measure (resp. empirical distribution) α. The next result shows that weak convergence of exchangeable processes is equivalent to weak convergence of these associated random measures. Kallenberg (1973) gives this and more general results.

(7.20) Proposition. *Let Z be an infinite exchangeable sequence directed by α. For $k \geq 1$ let $\underset{\sim}{Z}^k$ be exchangeable, and suppose either*

(a) *each $\underset{\sim}{Z}^k$ is infinite, directed by α_k, say; or*

(b) *$\underset{\sim}{Z}^k$ is N_k-exchangeable, with empirical distribution α_k, and $N_k \to \infty$*

Then $\underset{\sim}{Z}^k \xrightarrow{\mathcal{D}} \underset{\sim}{Z}$ *if and only if* $\alpha_k \xrightarrow{\mathcal{D}} \alpha$, *that is to say* $L(\alpha_k) \to L(\alpha)$ *in* $P(P(S))$.

Proof. (a) Recall the definition (7.8) of $\tilde{d}$. It is easy to check that the infimum in (7.8) is attained by some distribution $L(X,Y)$ which may be taken to have the form $g(\theta,\mu)$ for some measurable $g\colon P(S) \times P(S) \to P(S \times S)$. To prove the "if" assertion, we may suppose $\alpha_k \to \alpha$ a.s., by the Skorohod representation (7.9). Then $\tilde{d}(\alpha_k,\alpha) \to 0$ a.s. For each k let $(\underset{\sim}{V}^k,\underset{\sim}{W}^k) = ((V_i^k,W_i^k);\ i \geq 1)$ be the S^2-valued infinite exchangeable sequence directed by $g(\alpha_k,\alpha)$. Then

(i) $\underset{\sim}{V}^k \stackrel{\mathcal{D}}{=} \underset{\sim}{Z}^k$; $\underset{\sim}{W}^k \stackrel{\mathcal{D}}{=} \underset{\sim}{Z}$; each $k \geq 1$.

Also $E(d(V_1^k,W_1^k)\,|\,g(\alpha_k,\alpha)) = \tilde{d}(\alpha_k,\alpha)$, and so

(ii) $Ed(V_1^k,W_1^k) \to 0$ as $k \to \infty$.

Properties (i) and (ii) imply $\underset{\sim}{Z}^k \xrightarrow{\mathcal{D}} \underset{\sim}{Z}$.

Conversely, suppose $Z_k^k \xrightarrow{\mathcal{D}} \underset{\sim}{Z}$. Since $\bar{\alpha}_k = L(Z_1^k)$, Lemma 7.14 shows that (α_k) is tight. If $\hat{\alpha}$ is a weak limit, the "if" assertion of the Proposition implies $\hat{\alpha} \stackrel{\mathcal{D}}{=} \alpha$. So $\alpha_k \xrightarrow{\mathcal{D}} \alpha$ as required.

(b) Let $\hat{\underset{\sim}{Z}}^k$ be the infinite exchangeable sequence directed by α_k. By Proposition 5.6, for fixed $m \geq 1$ the total variation distance $\| L(\hat{Z}_1^k,\ldots,\hat{Z}_m^k) - L(Z_1^k,\ldots,Z_m^k) \|$ tends to 0 as $k \to \infty$. So $\underset{\sim}{Z}^k \xrightarrow{\mathcal{D}} \underset{\sim}{Z}$ iff $\hat{\underset{\sim}{Z}}^k \xrightarrow{\mathcal{D}} \underset{\sim}{Z}$, and part (b) follows from part (a).

Proposition 7.20 is of little practical use in the finite case, e.g. in proving central limit theorems for triangular arrays of exchangeable variables, because generally finite exchangeable sequences are presented in such a way that the distribution of their empirical distribution is not manifest. Section 20 presents more practical results. Even in the infinite case, there are open problems, such as the following.

Let $\underset{\sim}{Z}$ be an infinite exchangeable real-valued sequence directed by α. For constants $(a_1,\ldots,a_m)$ we can define an exchangeable sequence $\underset{\sim}{Y}$ by taking weighted sums of blocks of $\underset{\sim}{Z}$:

$$Y_i = \sum_{j=1}^{m} a_j Z_{j+(i-1)m} .$$

By varying $(m;a_1,\ldots,a_m)$ we obtain a class of exchangeable sequences; let $C(\underset{\sim}{Z})$ be the weak closure of this class.

(7.21) Problem. Describe explicitly which exchangeable processes are in $C(\underset{\sim}{Z})$.

This problem arises in the study of the Banach space structure of subspaces of L^1; see Aldous (1981b). There it was shown that, under a uniform integrability hypothesis, $C(\underset{\sim}{Z})$ must contain a sequence of the special form (VY_i), where (Y_i) is i.i.d. symmetric stable, and V is independent of (Y_i). This implies that every infinite-dimensional linear subspace of L^1 contains a subspace linearly isomorphic to some ℓ_p space. Further information about Problem 7.21 might yield further information about isomorphisms between subspaces of L^1.

(7.22) Stable convergence. For random variables $X_1, X_2, \ldots$ defined on the same probability space, say X_n converges stably if for each non-null event A the conditional distributions $L(X_n|A)$ converge in distribution to some limit, μ_A say. Plainly stable convergence is stronger than convergence in distribution and weaker than convergence in probability. This concept is apparently due to Rényi (1963), but has been rediscovered by many authors; a recent survey of stability and its applications is in Aldous and

Eagleson (1978). Rényi and Révész (1963) observed that exchangeable processes provide an example of stable convergence. Let us briefly outline this idea.

Copying the usual proof of existence of regular conditional distributions, one readily obtains

(7.23) Lemma. *Suppose (X_n) converges stably. Then there exists a random measure $\beta(\omega,\cdot)$ which represents the limit distributions μ_A via*

$$P(A)\mu_A(B) = \int 1_A(\omega)\beta(\omega,B)P(d\omega); \quad A \subset \Omega, \quad B \subset S.$$

Let us prove

(7.24) Lemma. *Suppose (Z_n) is exchangeable, directed by α. Then (Z_n) converges stably, and the representing random measure $\beta = \alpha$.*

Proof. Let $f \in C(S)$ and $A \in \sigma(Z_1,\ldots,Z_m)$. Then for $n > m$

$$\begin{aligned} P(A)E(f(Z_n)|A) &= E\,1_A\,E(f(Z_n)|Z_1,\ldots,Z_m,\alpha) \\ &= E\,1_A\,E(f(Z_n)|\alpha) \quad \text{by conditional independence} \\ &= E\,1_A \int f(x)\alpha(\omega,dx)\ . \end{aligned}$$

Thus $P(A)E(f(Z_n|A)) \to E\,1_A \int f(x)\alpha(\omega,dx)$ as $n \to \infty$ for $A \in \sigma(Z_1,\ldots,Z_m)$, and this easily extends to all A. Thus $L(Z_n|A) \to \mu_A$, where

$$P(A)\mu_A(\cdot) = E\,1_A\,\alpha(\omega,\cdot)$$

as required.

Note that our proof of Lemma (7.24) did not use the general result (7.23). It is actually possible to first prove the general result (7.23) and then use the type of argument above to give another proof of de Finetti's theorem; see Rényi and Révész (1963).

If X_n, defined on (Ω,F,P), converges stably, then we can extend the space to construct a "limit" variable X^* such that the representing measure β is a regular conditional distribution for X^* given F. Then (see e.g. Aldous and Eagleson (1978))

$$(7.25) \qquad (Y,X_n) \xrightarrow{\mathcal{D}} (Y,X^*); \quad \text{all} \quad Y \in F.$$

Classical weak convergence theorems for exchangeable processes are stable. For instance, let (Z_i) be a square-integrable exchangeable sequence directed by α. Let $S_n = n^{-1/2} \sum_1^n (Z_i - \text{mean}(\alpha))$. Then S_n converges stably, and its representing measure $\beta(\omega,\cdot)$ is the Normal $N(0,\text{var}(\alpha))$ distribution. If we construct a $N(0,1)$ variable W independent of the original probability space, then not only do we have $S_n \xrightarrow{\mathcal{D}} S^* = \{\text{var}(\alpha)\}^{1/2} W$ as at (2.27), but also by (7.25)

$$(Y,S_n) \xrightarrow{\mathcal{D}} (Y,S^*); \quad \text{each } Y \text{ in the original space.}$$

8. The subsequence principle

Suppose we are given a sequence (X_i) of random variables whose distributions are tight. Then we know we can pick out a subsequence $Y_i = X_{n_i}$ which converges in distribution. Can we say more, e.g. can we pick (Y_i) to have some tractable kind of dependence structure? It turns out that we can: informally,

(A) we can find a subsequence (Y_i) which is similar to some exchangeable sequence $\underset{\sim}{Z}$.

Now we know from de Finetti's theorem that infinite exchangeable sequences are mixtures of i.i.d. sequences, and so satisfy analogues of the classical limit theorems for i.i.d. sequences. So (A) suggests the equally informal assertion

(B) we can find a subsequence (Y_i) which satisfies an analogue of any prescribed limit theorem for i.i.d. sequences.

Historically, the prototype for (B) was the following result of Komlós (1967).

(8.1) Proposition. If $\sup_i E|X_i| < \infty$ then there exists a subsequence (Y_i) such that $N^{-1}\sum_1^N Y_i \to V$ a.s., for some random variable V.

This is (B) for the strong law of large numbers. Chatterji (1974) formulated (B) as the subsequence principle and established several other instances of it. A weak form of (A), in which (Y_i) is asymptotically exchangeable in the sense

$$(Y_{j+1}, Y_{j+2}, \ldots) \xrightarrow{\mathcal{D}} (Z_1, Z_2, \ldots) \quad \text{as} \quad j \to \infty,$$

arose independently from several sources: Dacunha-Castelle (1974), Figiel and Sucheston (1976), and Kingman (unpublished), who was perhaps the first to note the connection between (A) and (B). We shall prove this weak form of (A) as Theorem 8.9. Unfortunately this form is not strong enough to imply (B); we shall discuss stronger results later.

The key idea in our proof is in (b) below. An infinite exchangeable sequence $\underset{\sim}{Z}$ has the property (stronger than the property of stable convergence) that the conditional distribution of Z_{n+1} given $(Z_1, \ldots, Z_n)$ converges to the directing random measure; the key idea is a kind of converse, that any sequence with this property is asymptotically exchangeable. Our arguments are rather pedestrian; the proof of Dacunha-Castelle (1974) uses ultrafilters to obtain limits, while Figiel and Sucheston (1976) use Ramsey's combinatorial theorem to prove a result for general Banach spaces which is readily adaptable to our setting.

Suppose random variables take values in a Polish space S.

(8.2) Lemma. Let $\underset{\sim}{Z}$ be an infinite exchangeable sequence directed by α.

(a) Let α_n be a regular conditional distribution for Z_{n+1} given $(Z_1,\ldots,Z_n)$. Then $\alpha_n \to \alpha$ a.s.

(b) Let $\underset{\sim}{X}$ be an infinite sequence, let α_n be a regular conditional distribution for X_{n+1} given $(X_1,\ldots,X_n)$, and suppose $\alpha_n \to \alpha$ a.s. Then

(8.3) $$(X_{n+1},X_{n+2},\ldots) \xrightarrow{\mathcal{D}} (Z_1,Z_2,\ldots) \text{ as } n\to\infty.$$

Proof. (a) Construct Z_0 so that $(Z_i;\ i\geq 0)$ is exchangeable. Let $h \in C(S)$, and define $\bar{h}$ as at (7.7). Then

$$\begin{aligned}
\bar{h}(\alpha_n) &= E(h(Z_{n+1})|Z_1,\ldots,Z_n) \\
&= E(h(Z_0)|Z_1,\ldots,Z_n) && \text{by exchangeability} \\
&\to E(h(Z_0)|Z_i;\ i\geq 1) \text{ a.s.} && \text{by martingale convergence} \\
&= E(h(Z_0)|\alpha) \\
&= \bar{h}(\alpha).
\end{aligned}$$

Apply (7.10).

(b) Given $\underset{\sim}{X}$ and α, let $F_m = \sigma(X_1,\ldots,X_m)$, $F = \sigma(X_i;\ i\geq 1)$ and construct $\underset{\sim}{Z}$ such that $\underset{\sim}{Z}$ is an infinite exchangeable sequence directed by α and also

(8.4) $\underset{\sim}{Z}$ and F are conditionally independent given α.

We shall prove, by induction on k, that

(8.5) $$(V,X_{n+1},\ldots,X_{n+k}) \xrightarrow{\mathcal{D}} (V,Z_1,\ldots,Z_k) \text{ as } n\to\infty; \text{ each } V\in F;$$

for each k. This will establish (b).

Suppose (8.5) holds for fixed $k \geq 0$. Let $f: S^k \times S \to R$ be bounded continuous. Define $\bar{f}: S^k \times P(S) \to R$ by

$$\bar{f}(x_1,\ldots,x_k,L(Y)) = Ef(x_1,\ldots,x_k,Y) .$$

Note $\bar{f}$ is continuous. By the fundamental property of conditional distributions,

$$(8.6) \qquad E(f(X_{n+1},\ldots,X_{n+k},X_{n+k+1})|F_{n+k}) = \bar{f}(X_{n+1},\ldots,X_{n+k},\alpha_{n+k})$$

$$(8.7) \qquad E(f(Z_1,\ldots,Z_k,Z_{k+1})|F,Z_1,\ldots,Z_k) = \bar{f}(Z_1,\ldots,Z_k,\alpha), \quad \text{using (8.4).}$$

Fix $m \geq 1$ and $A \in F_m$. By inductive hypothesis

$$(\alpha,1_A,X_{n+1},\ldots,X_{n+k}) \xrightarrow{\mathcal{D}} (\alpha,1_A,Z_1,\ldots,Z_k) \quad \text{as} \quad n \to \infty .$$

Since $\alpha_{n+k} \to \alpha$ a.s.,

$$(8.8) \qquad (\alpha,1_A,X_{n+1},\ldots,X_{n+k},\alpha_{n+k}) \xrightarrow{\mathcal{D}} (\alpha,1_A,Z_1,\ldots,Z_k,\alpha) \quad \text{as} \quad n \to \infty.$$

Now

$$\begin{aligned} Ef(X_{n+1},\ldots,X_{n+k+1})1_A &= E\bar{f}(X_{n+1},\ldots,X_{n+k},\alpha_{n+k})1_A, \quad n \geq m, \quad \text{by (8.6)} \\ &\to E\bar{f}(Z_1,\ldots,Z_k,\alpha)1_A \quad \text{as } n \to \infty, \text{ by (8.8)} \\ &\qquad \text{and continuity of } \bar{f}; \\ &= Ef(Z_1,\ldots,Z_{k+1})1_A \quad \text{by (8.7).} \end{aligned}$$

Since this convergence holds for all f, we see that the inductive assertion (8.5) holds for $k+1$ when $V = 1_A$, $A \in F_m$. But m is arbitrary, so this extends to all $V \in F$.

(8.9) Theorem. Let $\underset{\sim}{X}$ be a sequence of random variables such that $\mathcal{L}(X_i)$ is tight. Then there exists a subsequence $Y_i = X_{n_i}$ such that

$$(Y_{j+1}, Y_{j+2}, \dots) \xrightarrow{\mathcal{D}} (Z_1, Z_2, \dots) \quad \text{as} \quad j \to \infty$$

for some exchangeable $\underset{\sim}{Z}$.

We need one preliminary. A standard fact from functional analysis is that the unit ball of a Hilbert space is compact in the weak topology (i.e. the topology generated by the dual space): applying this fact to the space L^2 of random variables gives

(8.10) Lemma. Let (V_i) be a uniformly bounded sequence of real-valued random variables. Then there exists a subsequence (V_{n_i}) and a random variable V such that $EV_{n_i} 1_A \to EV1_A$ for all events A.

Proof of Theorem 8.9. By approximating, we may suppose each X_i takes values in some finite set S_i. Let (h_j) be a convergence-determining class. By Lemma 8.10 and a diagonal argument, we can pick a subsequence (X_n) such that as $n \to \infty$

$$(8.13) \qquad Eh_j(X_n)1_A \to EV_j 1_A; \quad \text{each } A, j .$$

We can now pass to a further subsequence in which

$$(8.14) \qquad \left| E(h_j(X_{n+1}) | A) - E(V_j | A) \right| \le 2^{-n}$$

for each $n \geq 1$, each $1 \leq j \leq n$ and each atom A of the finite σ-field $F_n = \sigma(X_1, \dots, X_n)$ with $P(A) > 0$. Let α_n be a regular conditional distribution for X_{n+1} given F_n. We shall prove $\alpha_n \to \beta$ a.s. for some random measure β, and then Lemma 8.2(b) establishes the theorem. Note

(8.15) $$E(h_j(X_{n+1}|F_n) = \bar{h}_j(\alpha_n) .$$

Fix $m \geq 1$ and an atom A of F_m. By (8.13),

(8.16) $$L(X_n|A) \to \mu_A, \text{ say, where } \bar{h}_j(\mu_A) = E(V_j|A) .$$

Let β_m be the random measure such that $\beta_m(\omega,\cdot) = \mu_A(\cdot)$ for $\omega \in A$. So $\bar{h}_j(\beta_m) = E(V_j|F_m)$, and so by (8.14) and (8.15)

(8.17) $$|h_j(\alpha_n)-h_j(\beta_n)| \leq 2^{-n} ; \quad 1 \leq j \leq n .$$

We assert that (β_m) forms a martingale, in the sense of Lemma 7.14. For an atom A of F_m is a finite union of atoms A_k of F_{m+1}, and by (8.16) $\mu_A(B) = \sum P(A_k|A)\mu_{A_k}(B)$, $B \subset S$, which implies $E(\beta_{m+1}(\cdot,B)|F_m) = \beta_m(\cdot,B)$. Now by Lemma 7.14 we have $\beta_m \to \beta$ a.s., for some random measure β. And (8.17) implies $\bar{h}_j(\alpha_n) \to \bar{h}_j(\beta)$ a.s. for each j, and so $\alpha_n \to \beta$ a.s. as required.

Let us return to discussion of the subsequence principle. Call (Y_i) <u>almost exchangeable</u> if we can construct exchangeable (Z_i) such that $\sum|Y_i-Z_i| < \infty$ a.s. (we are now taking real-valued sequences). Plainly such a (Y_i) will inherit from (Z_i) the property of satisfying analogues of classical limit theorems. So if we can prove

(8.18) Every tight sequence (X_i) has an almost exchangeable subsequence (Y_i)

then we would have established a solid form of the subsequence principle (B). Unfortunately (8.18) is false. See Kingman (1978) for a counterexample, and Berkes and Rosenthal (1983) for more counterexamples and discussion of which sequences (X_i) do satisfy (8.18).

Thus we need a property weaker than "almost exchangeability" but stronger than "asymptotically exchangeable". Let $\varepsilon_k \downarrow 0$. Let (X_n) be such that for each k we can construct exchangeable $(Z_j^k,\ j \geq k)$ such that $P(|X_j - Z_j^k| > \varepsilon_k) \leq \varepsilon_k$ for each $j \geq k$. This property (actually, a slightly stronger but more complicated version) was introduced by Berkes and Péter (1983), who call such (X_n) <u>strongly exchangeable at infinity</u> with rate (ε_k). They prove

(8.19) <u>Theorem</u>. <u>Let</u> (X_i) <u>be tight, and let</u> $\varepsilon_k \downarrow 0$. <u>Then there exists a subsequence</u> (Y_i) <u>which is strongly exchangeable at infinity with rate</u> (ε_k).

(Again, they actually prove a slightly stronger result). From this can be deduced results of type (B), such as Proposition 8.1 and, to give another example, the analogue of the law of the iterated logarithm:

(8.20) <u>Proposition</u>. <u>If</u> $\sup EX_i^2 < \infty$ <u>then there exists a subsequence</u> (Y_i) <u>and random variables</u> V, S <u>such that</u> $\limsup_{n\to\infty} (2N \log\log(N))^{-1/2} \sum_{i=1}^{N} (Y_i - V) = S$ a.s.

A different approach to the subsequence principle is to abstract the idea of a "limit theorem". Let $A \subset P(R) \times R^\infty$ be the set

$$\{(\theta, \underset{\sim}{x}) : \text{mean}(\theta) = \infty \text{ or } \lim N^{-1} \sum_{1}^{N} x_i = \text{mean}(\theta)\} .$$

Then the strong law of large numbers is the assertion

(8.21) $$P((\theta, X_1, X_2, \ldots) \in A) = 1 \quad \text{for } (X_i) \text{ i.i.d. } (\theta) .$$

Similarly, any a.s. limit theorem for i.i.d. variables can be put in the form of (8.21) for some set A, which we call a <u>statute</u>. Call A a <u>limit statute</u> if also

$$\text{if} \quad (\theta,\underset{\sim}{x}) \in A \quad \text{and if} \quad \sum|\hat{x}_i - x_i| < \infty \quad \text{then} \quad (\theta,\underset{\sim}{\hat{x}}) \in A \; .$$

Then Aldous (1977) shows

(8.22) Theorem. Let A be a limit statute and (X_i) a tight sequence. Then there exists a subsequence (Y_i) and a random measure α such that

$$(\alpha, Y_1, Y_2, \ldots) \in A \quad \text{a.s.}$$

Applying this to the statutes describing the strong law of large numbers or the law of the iterated logarithm, we recover Propositions 8.1 and 8.20. To appreciate (8.22), observe that for an exchangeable sequence (Z_i) directed by α we have $(\alpha, Z_1, Z_2, \ldots) \in A$ a.s. for each statute A, by (8.21). So for an almost exchangeable sequence (Y_i) and a limit statute A we have $(\alpha, Y_1, Y_2, \ldots) \in A$ a.s. Thus (8.22) is a consequence of (8.18), when (8.18) holds; what is important is that (8.22) holds in general while (8.18) does not.

The proofs of Theorems 8.19 and 8.22 are too technical to be described here: interested readers should consult the original papers.

9. Other discrete structures

In Part III we shall discuss processes $(X_i: i \in I)$ invariant under specified transformations of the index set I. As an introduction to this subject, we now treat some simple cases where the structure of the invariant processes can be deduced from de Finetti's theorem. We have already seen one result of this type, Corollary 3.9.

Two exchangeable sequences. Consider two infinite S-valued sequences (X_i), (Y_i) such that

(9.1) the sequence (X_i, Y_i), $i \geq 1$, of pairs is exchangeable.

Then this sequence of pairs is a mixture of i.i.d. bivariate sequences, directed by some random measure α on $S \times S$, and the marginals $\alpha_X(\omega)$, $\alpha_Y(\omega)$ are the directing measures for (X_i) and for (Y_i). Corollary 3.9 says that the stronger condition

(9.2) $$(X_1, X_2, \ldots; Y_1, Y_2, \ldots) \stackrel{\mathcal{D}}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots; Y_{\sigma(1)}, Y_{\sigma(2)}, \ldots)$$
for all finite permutations π, σ

holds iff $\alpha(\omega) = \alpha_X(\omega) \times \alpha_Y(\omega)$.

If we wish to allow switching X's and Y's, consider the following possible conditions:

(9.3) $$(X_1, X_2, X_3, \ldots; Y_1, Y_2, Y_3, \ldots) \stackrel{\mathcal{D}}{=} (Y_1, Y_2, \ldots; X_1, X_2, \ldots),$$

(9.4) $$(X_1, X_2, X_3, \ldots; Y_1, Y_2, Y_3, \ldots) \stackrel{\mathcal{D}}{=} (Y_1, X_2, X_3, \ldots; X_1, Y_2, Y_3, \ldots) .$$

Let $h(x,y) = (y,x)$; let $\tilde{h}: \mathcal{P}(S \times S) \to \mathcal{P}(S \times S)$ be the induced map, and let $\mathcal{S}$ be the set of symmetric (i.e. $\tilde{h}$-invariant) measures on $S \times S$.

(9.5) Proposition.

(a) Both (9.1) and (9.3) hold iff $\alpha \stackrel{\mathcal{D}}{=} \tilde{h}(\alpha)$.

(b) Both (9.1) and (9.4) hold iff $\alpha(\omega) \in \mathcal{S}$ a.s.

(c) Both (9.2) and (9.3) hold iff $\alpha(\omega) = \alpha_X(\omega) \times \alpha_Y(\omega)$ a.s., where $(\alpha_X, \alpha_Y) \stackrel{\mathcal{D}}{=} (\alpha_Y, \alpha_X)$.

(d) Both (9.2) and (9.4) hold iff $\alpha(\omega) = \alpha_X(\omega) \times \alpha_Y(\omega)$ a.s., where $\alpha_X = \alpha_Y$ a.s., that is iff the whole family $(X_1, X_2, \ldots; Y_1, Y_2, \ldots)$ is exchangeable.

This is immediate from the remarks above and the following lemma, applied to $Z_i = (X_i, Y_i)$.

(9.6) Lemma. Let $h: S \to S$ be measurable, let $\tilde{h}: P(S) \to P(S)$ be the induced map, and let P_h be the set of distributions μ which are h-invariant: $\tilde{h}(\mu) = \mu$. Let $\underset{\sim}{Z}$ be an infinite exchangeable S-valued sequence directed by α.

(i) $\underset{\sim}{Z} \overset{\mathcal{D}}{=} (h(Z_1), h(Z_2), h(Z_3), \ldots)$ iff $\alpha \overset{\mathcal{D}}{=} \tilde{h}(\alpha)$.

(ii) $\underset{\sim}{Z} \overset{\mathcal{D}}{=} (h(Z_1), Z_2, Z_3, Z_4, \ldots)$ iff $\alpha = \tilde{h}(\alpha)$ a.s., that is $\alpha \in P_h$ a.s.

Proof. Lemma 4.4(a) says that $(h(Z_i))$ is an exchangeable sequence directed by $\tilde{h}(\alpha)$, and this gives (i). For (ii), note first

$$\alpha \text{ is a r.c.d. for } Z_1 \text{ given } \alpha;$$
$$\tilde{h}(\alpha) \text{ is a r.c.d. for } h(Z_1) \text{ given } \alpha.$$

Writing $W = (Z_2, Z_3, \ldots)$, we have by Lemma 2.19

$$\alpha \text{ is a r.c.d. for } Z_1 \text{ given } W;$$
$$\tilde{h}(\alpha) \text{ is a r.c.d. for } h(Z_1) \text{ given } W.$$

Now $(Z_1, W) \overset{\mathcal{D}}{=} (h(Z_1), W)$ iff the conditional distribution for Z_1 and $h(Z_1)$ given W are a.s. equal: this is (ii).

It is convenient to record here a technical result we need in Section 13.

(9.7) Lemma. Let (X_i), (Y_i) be exchangeable. Suppose that for each subset A of $\{1, 2, \ldots\}$ the sequence Z defined by

$$Z_i = X_i, \quad i \in A; \quad Z_i = Y_i, \quad i \notin A$$

<u>satisfies</u> $Z \stackrel{\mathcal{D}}{=} X$. <u>Then the directing random measures</u> α_Z, α_X <u>satisfy</u> $\alpha_Z = \alpha_X$ <u>a.s. for each</u> Z.

<u>Remark</u>. This says that conditional on $\alpha = \theta$, the vectors (X_i, Y_i) are indepedent as i varies and have marginal distributions θ.

<u>Proof</u>. In the notation of Lemma 2.15, $\alpha_X = \Lambda(X_1, X_2, \ldots)$. Now a function of infinitely many variables may be approximated by functions of finitely many variables, so there exist functions g_k such that

$$E\, \tilde{d}(\alpha_X, g_k(X_1, \ldots, X_k)) = \delta_k \,, \tag{9.8}$$

where $\tilde{d}$ is a bounded metrisation of $P(S)$, and $\delta_k \to 0$ as $k \to \infty$. Fix $\underset{\sim}{Z}$ and define $\underset{\sim}{Z}^k$ by

$$\begin{aligned} Z_i^k &= X_i, \quad i \le k \\ &= Z_i, \quad i > k. \end{aligned}$$

By hypotheses $\underset{\sim}{Z}^k \stackrel{\mathcal{D}}{=} \underset{\sim}{X}$, so by (9.8) $E\, \tilde{d}(\alpha_{Z^k}, g_k(Z_1^k, \ldots, Z_k^k)) = \delta_k$. But $\alpha_{Z^k} = \alpha_Z$ a.s. because α is tail-measurable; and $Z_i^k = X_i$ for $i \le k$; so by (9.8)

$$E\, d(\alpha_Z, \alpha_X) \le 2\delta_k \,.$$

Since k is arbitrary, $\alpha_Z = \alpha_X$ a.s.

<u>A stratified tree</u>. We now discuss a quite different structure, a type of stratified tree. For each $n \in \mathbb{Z}$ let $I_n = \{j2^n : j \ge 0\}$, and let $I = \{(n,i) : n \in \mathbb{Z}, i \in I_n\}$. The set I has a natural tree structure--see the diagram. A point (n,i) has a set of "descendants", the points (m,i') such that $m \le n$ and $i \le i' < i + 2^{-n}$. Given n and $i_1, i_2 \in I_n$ we can define a map $\gamma : I \to I$ which switches the descendants of (n, i_1) with those of (n, i_2):

$$\gamma(m,i) = (m,i) \quad \text{if } (m,i) \text{ is not a descendant of } (n,i_1) \text{ or } (n,i_2)$$
$$= (m,i+i_2-i_1) \quad \text{if } (m,i) \text{ is a descendant of } (n,i_1)$$
$$= (m,i+i_1-i_2) \quad \text{if } (m,i) \text{ is a descendant of } (n,i_2).$$

Let Γ be the set of maps γ of this form. We want to consider processes $\underset{\sim}{X} = (X_i : i \in I)$ invariant under Γ; that is

(9.9) $$\underset{\sim}{X} \overset{\mathcal{D}}{=} (X_{\gamma(i)}, i \in I), \quad \text{each } \gamma \in \Gamma.$$

Suppose also that each $X_{n,i}$ is a function of its immediate descendants:

(9.10) $$X_{n,i} = f_n(X_{n-1,i}, X_{n-1,i+2^{n-1}}) .$$

(9.11) <u>Lemma</u>. <u>Under hypotheses</u> (9.10) <u>and</u> (9.10), <u>there is a σ-field</u> F <u>such that for each</u> n <u>the family</u> $(X_{n,i}; i \in I_n)$ <u>is conditionally i.i.d. given</u> F.

<u>Proof</u>. For fixed n the family $(X_{n,i}; i \in I_n)$ is exchangeable, and so has directing random measure α_n, say. Now consider $k < n$. The variables $(X_{n,i}; i \in I_n)$ are functions of the variables $(X_{k,i}; i \in I_k)$ which are conditionally i.i.d. given α_k, and hence $(X_{n,i}; i \in I_n)$ are conditionally i.i.d. given α_k. Appealing to Lemma 2.12 we see

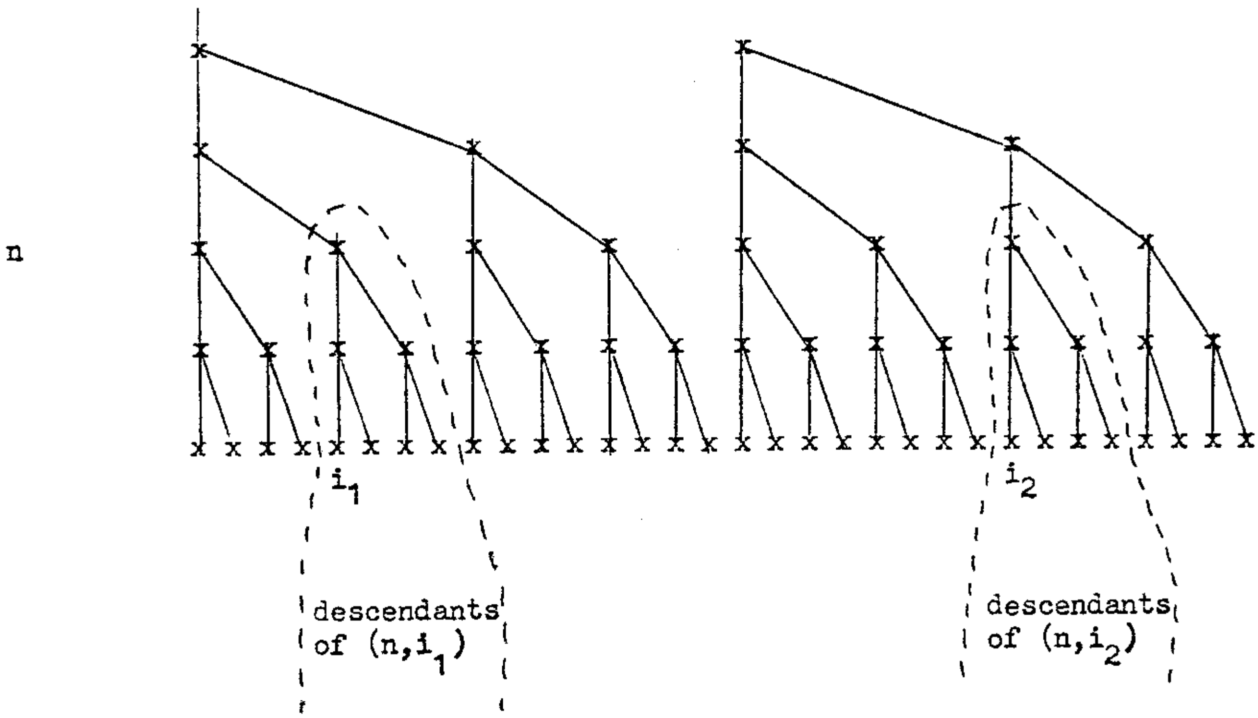
$$(X_{n,i}; i \in I_n) \text{ and } \alpha_k \text{ are c.i. given } \alpha_n$$
$$\alpha_n \in \sigma(\alpha_k) \text{ a.s.}$$

Setting $F = \bigvee_{k<-\infty} \sigma(\alpha_k)$, we obtain

$$(X_{n,i}; i \in I_n) \text{ and } F \text{ are c.i. given } \alpha_n.$$

Since the family $(X_{n,i}; i \in I_n)$ is conditionally i.i.d. given α_n, the result follows.

(9.12) Problem. What is the analogue of Lemma 9.11 in the finite case, where we set $I_n = \{j2^n: 0 \leq j \leq 2^{-n}, n \leq 0\}$?



10. Continuous-time processes

de Finetti's theorem can be extended to uncountable families of random variables. But a process $(X_t: t \geq 0)$ with i.i.d. values has sample paths which are either constant or non-measurable, and this makes the concept of an exchangeable process $(X_t: t \geq 0)$ rather uninteresting. However, by considering instead the continuous analogue of processes of partial sums of exchangeable sequences, we get the concept of processes with interchangeable increments, and this leads to the simplest continuous-time analogues of discrete-time results. This theory has been developed in detail by Kallenberg (1973, 1974, 1975), who gives the results through (10.19) and many further results.

Consider real-valued processes $X = (X_t)$, $0 \leq t < \infty$ or $0 \leq t \leq 1$, with sample paths in the space D ($= D(0,\infty)$ or $D(0,1)$) of functions which are right-continuous with left limits. Say X has <u>interchangeable increments</u> if

(10.1a) for each $\delta > 0$ the sequence $(Z_i) = (X_{i\delta} - X_{(i-1)\delta})$ of increments is exchangeable;

(10.1b) $X_0 = 0$.

Assumption (b) entails no real loss of generality, since we can replace X_t by $X_t - X_0$ and preserve (a). Informally, think of interchangeable increments processes as integrals $X_t = \int_0^t Z_s \, ds$ of some underlying exchangeable "generalized process" Z.

Here is an alternative definition.

Given disjoint intervals $(a_1,b_1]$, $(a_2,b_2]$ of equal length, let $T\colon \mathbb{R}^+ \to \mathbb{R}^+$ be the natural map which switches these intervals:

(10.2)
$$T(t) = t; \quad t \notin \bigcup_i (a_i,b_i]$$
$$T(a_1+t) = a_2 + t, \quad T(a_2+t) = a_1 + t; \quad 0 < t \leq b_i - a_i.$$

Let $\mathcal{T}$ be the set of such maps T. Let $\mathcal{B}$ be the set of finite unions of disjoint intervals. With any real-valued point function $f = (f(t)\colon t \geq 0)$ we can associate a set function $(f(B)\colon B \in \mathcal{B})$ defined by

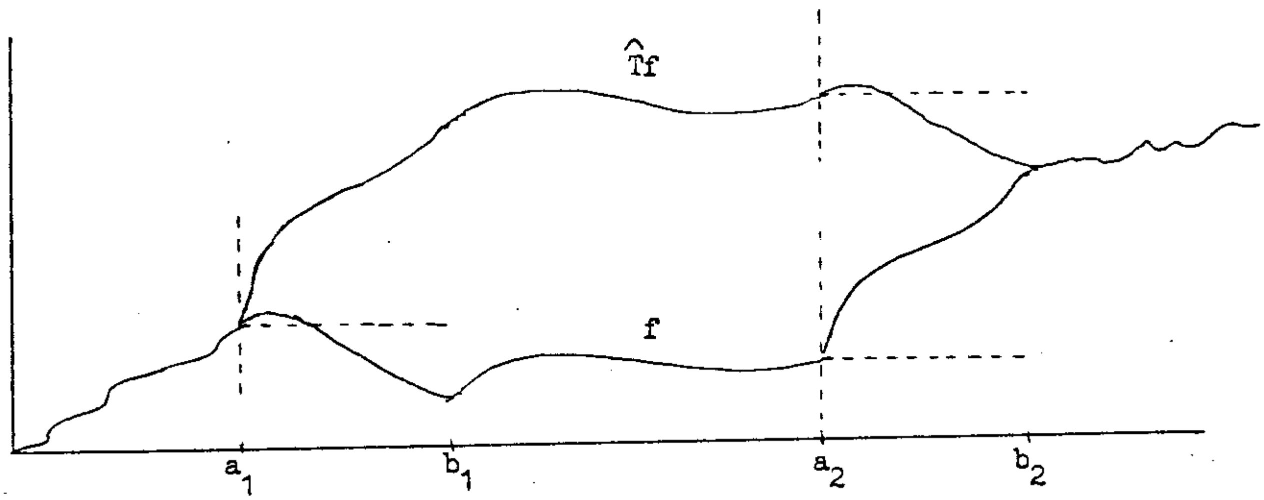
$$f(B) = \sum_i (f(t_i) - f(s_i)); \quad B = \cup(s_i,t_i], \text{ disjoint.}$$

Given a function f and a map T, let $\hat{T}f$ be the function whose associated set function is

$$(\hat{T}f)(B) = f(T(B)) .$$

The diagram is more informative than the formulas: $\hat{T}f$ is obtained from f by switching the increment of f over $(a_1,b_1]$ with the increment over $(a_2,b_2]$. Now $\hat{T}$ maps D to D. It is easy to see that definition (10.1a) is equivalent to

(10.3) $$\hat{T}(X) \overset{\mathcal{D}}{=} X; \quad \text{each } T \in \mathcal{T}.$$



Recall the theory of Lévy processes, i.e. processes with stationary independent increments. For a Lévy process X, the distribution of X_1 is infinitely divisible; conversely, to any infinitely divisible distribution there corresponds a Lévy process for which X_1 has the given distribution. A continuous-path Lévy process is of the form

(10.4a) $$X_t = at + bB_t; \quad B_t \text{ Brownian motion; } a, b \text{ constants.}$$

A Lévy process which is a counting process is of the form

(10.4b) $$X_t = N_{\lambda t}; \quad N_t \text{ Poisson process of rate 1; } \lambda \text{ constant.}$$

It turns out that processes with interchangeable increments on the time-interval $[0,\infty)$ (resp. $[0,1]$) are analogous to infinite (resp. finite) exchangeable sequences. Here is the analogue of de Finetti's theorem, first noted by Bühlmann (1960).

(10.5) Proposition. A process X with interchangeable increments on the time-interval $[0,\infty)$ is a mixture of Lévy processes.

Proof. For $n \in \mathbb{Z}$, $j \geq 0$ set $i = j2^n$ and set $X_{n,i} = (X(i+t)-X(i): 0 \leq t \leq 2^n)$. So $X_{n,i}$ is a random element of $D[0,2^n]$. We assert that $(X_{n,i})$ satisfies the hypotheses of Lemma (9.11). In fact (9.10) is immediate, and assertion (9.9) is a reformulation of the property of interchangeable increments for intervals with dyadic rational endpoints. Now Lemma (9.11) concludes that, conditional on a certain σ-field F, the family $(X_{n,j2^n}; j \geq 0)$ is i.i.d. for each n. By approximating arbitrary intervals by intervals with dyadic rational endpoints, we deduce that conditional on F the process X is Lévy.

Using (10.4a,b) we get

(10.6) Corollary. Let X be a process with interchangeable increments on the time-interval $[0,\infty)$. If X has continuous paths then

(a) $X_t = \alpha t + \beta B_t$; B_t Brownian motion; α, β r.v.'s independent of B.

If X is a counting process, then

(b) $X_t = N_{\Lambda t}$; N_t the Poisson process of rate 1; Λ a r.v. independent of N.

We should also mention the analogue of Theorem 6.1. For a process X write X^t for the process $X^t_u = X_{t+u} - X_t$. The strong Markov property shows that X is a Lévy process iff

$$X^T \text{ is independent of } \sigma(X_s,\ s \le T);\ X^T \overset{\mathcal{D}}{=} X; \text{ for each stopping time } T.$$

Kallenberg (1982a,b) defines a process X to have <u>strongly stationary increments</u> if

(10.7) $$X^T \overset{\mathcal{D}}{=} X; \quad \text{each stopping time } T,$$

and shows that this property is equivalent to the interchangeable increments property. The proof requires only Theorem 6.1 and the arguments of Proposition 10.5.

Now consider processes on the time-interval $[0,1]$. Two processes with interchangeable increments are

(a) the counting process N^m of m draws from the uniform distribution:

(10.8) $$N_t^m = \sum_{i=1}^m 1_{(t \le \xi_i)}, \quad (\xi_i) \text{ i.i.d. } U(0,1).$$

(b) the Brownian bridge B^0.

It turns out that these are the basic examples of counting processes (resp. continuous-path processes) with interchangeable increments on $[0,1]$.

(10.9) <u>Lemma</u>. <u>Let</u> X <u>be a counting process on</u> $[0,1]$ <u>with interchangeable increments, and suppose</u> $X_1 = m$. <u>Then</u> $X \overset{\mathcal{D}}{=} N^m$, <u>for</u> N^m <u>as at</u> (10.7).

<u>Proof</u>. Let $D_k = \{i2^{-k}: 0 \le i < 2^k\}$ and let $\phi_k(x) = \max\{t \le x: t \in D_k\}$. Let $0 \le \theta_1 < \cdots < \theta_m \le 1$ be the jump times of X, and let $\xi_{(1)},\ldots,\xi_{(m)}$ be the increasing ordering of (ξ_i). Let A_k be the event that the random variables $\phi_k(\theta_i)$, $1 \le i \le m$, are distinct; let $\hat{A}_k$ be the corresponding event for (ξ_i). Then

(i) $P(A_k) \to 1$ and $P(\hat{A}_k) \to 1$ as $k \to \infty$;

(ii) $\phi_k(\theta_i) \to \theta_i$ a.s. and $\phi_k(\xi_i) \to \xi_i$ a.s. as $k \to \infty$;

(iii) the distribution of $(\phi_k(\theta_1),\dots,\phi_k(\theta_m))$ conditional on A_k is the same as the distribution of $(\phi_k(\xi_{(1)}),\dots,\phi_k(\xi_{(m)}))$ conditional on $\hat{A}_k$, because each is the distribution of the order statistics of m draws without replacement from D_k.

Properties (i)-(iii) imply $(\theta_1,\dots,\theta_m) \overset{\mathcal{D}}{=} (\xi_{(1)},\dots,\xi_{(m)})$, so $X \overset{\mathcal{D}}{=} N^m$.

From Lemma 10.9 we see the form of the general counting process with interchangeable increments on $[0,1]$: let M have an arbitrary distribution on $\{0,1,2,\dots\}$ and then, conditional on $\{M=m\}$, let X have distribution (10.8).

We now consider continuous-path processes. Let $(B^0_t\colon 0 \le t \le 1)$ be the Brownian bridge, that is the Gaussian process with $B^0_0 = B^0_1 = 0$ and

$$EB^0_s B^0_t = s(1-t), \quad 0 \le s \le t \le 1 \,. \tag{10.10}$$

For fixed k put $Z_i = B^0_{i/k} - B^0_{(i-1)/k}$. Then $(Z_1,\dots,Z_k)$ is Gaussian, and using (10.10) we compute $EZ_i^2 = (k-1)k^{-2}$, $EZ_iZ_j = -k^{-2}$ $(i \ne j)$, and so (Z_i) is exchangeable. It follows that B^0 has interchangeable increments. From this we can construct more continuous-path interchangeable increments processes by putting

$$X_t = \alpha B^0_t + \beta t; \quad \text{where } (\alpha,\beta) \text{ is independent of } B^0. \tag{10.11}$$

Note that Brownian motion on $[0,1]$ is of this form, for $\alpha = 1$ and β having $N(0,1)$ distribution.

(10.12) Theorem. *Every continuous-path interchangeable increments process* $(X_t\colon 0 \le t \le 1)$ *is of the form* (10.11).

This is a non-trivial result. The natural method of proof (Kallenberg (1973)) is via weak convergence of discrete approximations. However, we shall later (Section 20) use Theorem 10.12 as a starting point for proving weak convergence results, and so for aesthetic reasons we would like a proof of Theorem 10.12 independent of weak convergence ideas. The following proof uses martingale techniques, and assumes

(10.13) $$EX_t^2 < \infty, \text{ each } t.$$

We shall later indicate how to remove this restriction.

(10.14) Lemma. Let $(X_t: 0 \le t \le 1)$ be a continuous-path process adapted to a filtration (G_t) and satisfying

(i) $X_0 = X_1 = 0$

(ii) $E(X_u - X_t | G_t) = -\frac{u-t}{1-t} X_t, \quad t \le u < 1$

(iii) $\text{var}(X_u - X_t | G_t) = \frac{(u-t)(1-u)}{1-t}, \quad t \le u < 1.$

Then $X \overset{\mathcal{D}}{=} B^0$ and X is independent of $G_0^+ = \bigcap_{t>0} G_t$.

Proof. Write

(10.15) $$Y_t = (1+t)X_{t/(1+t)}; \quad F_t = G_{t/(1+t)}; \quad 0 \le t < \infty.$$

Then $Y_0 = 0$, Y has continuous paths and by (ii) and (iii),

$$E(Y_u - Y_t | F_t) = 0, \quad \text{var}(Y_u - Y_t | F_t) = u - t; \quad t \le u.$$

So by Lévy's Theorem (see e.g. Doob (1953) p. 78) Y is Brownian motion. Inverting (10.15) gives $X_t = (1-t)Y_{t/(1-t)}$, which implies X is Brownian bridge. Finally, for any event A in G_0^+ the process X conditioned on A satisfies the hypotheses of the lemma, so the conditioned process is also Brownian bridge; this proves independence.

Proof of Theorem 10.12. Consider first the special case

$$X_1 = 0 .$$

For $m \geq 1$ let $D_m = \{i2^{-m}: 0 \leq i \leq 2^m\}$. Let $V_{m,j} = (X(j2^{-m}+u) - X(j2^{-m}): u \leq 2^{-m})$, considered as a random element of $D(0,2^{-m})$. For $t \in D_m$ let G_t^m be the σ-field generated by $(X_s: s \leq t)$ and the empirical distribution of $(V_{m,j}: j2^{-m} \geq t)$. Then conditional on G_t^m the variables $(X_{(j+1)2^{-m}} - X_{j2^{-m}}: j2^{-m} \geq t)$ are an urn process, in the language of section 5. The elementary formulas for means and variances when sampling without replacement (20.1) show that for $u \in D_m$, $u \geq t$,

$$\text{(10.16)} \qquad \begin{aligned} E(X_u - X_t | G_t^m) &= -\frac{u-t}{1-t} X_t \\ \operatorname{var}(X_u - X_t | G_t^m) &= \frac{(u-t)(1-u)}{(1-t)(1-t-2^{-m})} \left\{Q_1^m - Q_t^m + \frac{2^{-m}X_t^2}{1-t}\right\} \end{aligned}$$

where

$$\text{(10.17)} \qquad Q_t^m = \sum_{i=0}^{t2^m-1} (X_{(i+1)2^{-m}} - X_{i2^{-m}})^2; \quad t \in D_m .$$

For fixed t, the σ-fields G_t^m are decreasing as m increases; for $t \in D = \bigcup_m D_m$ let $G_t = \bigcap_m G_t^m$, and for general t let $G_t = \bigcap_{\substack{u>t \\ u \in D}} G_u$. Suppose we can prove there exists a random variable $\alpha \geq 0$ such that

$$\text{(10.18)} \qquad Q_t^m \xrightarrow[p]{} \alpha t; \quad \text{each } t \in D.$$

Then reverse martingale convergence in (10.16) shows that for $t \leq u$ $(t, u \in D)$

$$\begin{aligned} E(X_u - X_t | G_t) &= -\frac{u-t}{1-t} X_t \\ \operatorname{var}(X_u - X_t | G_t) &= \frac{(u-t)(1-u)}{1-t} \alpha . \end{aligned}$$

These extend to all $t \le u$, by approximating from above. Note $\alpha \in \mathcal{G}_0^+$. On $\{\alpha > 0\}$ set $V_t = \alpha^{-1/2} X_t$. Then V satisfies the hypotheses of Lemma 10.13, and so V is Brownian bridge, independent of α. Since $X_t = \alpha^{1/2} V_t$, this establishes the theorem in the special case $X_1 = 0$. For the general case, set $\hat{X}_t = X_t - tX_1$, define $\hat{\mathcal{G}}_t^m$ using $\hat{X}$ as $\mathcal{G}_t^m$ was defined using X, and include X_1 in the σ-field. The previous argument gives that $V_t = \alpha^{-1}\hat{X}_t$ is Brownian bridge independent of $\mathcal{G}_0^+ \supset \sigma(\alpha, X_1)$, and then writing $X_t = \alpha V_t + X_1 t$ establishes the theorem in the general case.

To prove (10.18), we quote the following lemma, which can be regarded as a consequence of maximal inequalities for sampling without replacement (20.5) or as the degenerate weak convergence result (20.10).

(10.19) Lemma. For each $m \ge 1$ let $(Z_{m,1}, \dots, Z_{m,k_m})$ be exchangeable. If

(a) $\sum_i Z_{m,i} = 0$ for each m,

(b) $\sum_i Z_{m,i}^2 \xrightarrow[p]{} 0$ as $m \to \infty$,

then

(c) $\max_j \left| \sum_{i=1}^{j} Z_{m,i} \right| \xrightarrow[p]{} 0$ as $m \to \infty$.

Proof of (10.18). Set $Z_{m,i} = (X_{(i+1)2^{-m}} - X_{i2^{-m}})^2 - 2^{-m} Q_1^m$. Then (a) is immediate. For (b),

$$\begin{aligned} \sum_i Z_{m,i}^2 &\le \delta_m \sum_i |Z_{m,i}|; \quad \delta_m = \max_i |Z_{m,i}| \\ &\le \delta_m \cdot 2Q_1^m \\ &\xrightarrow[p]{} 0 \text{ since } \delta_m \xrightarrow[p]{} 0 \text{ by continuity and } Q_1^m \text{ converges a.s.} \\ &\qquad \text{by reverse martingale convergence in (10.16).} \end{aligned}$$

So conclusion (c) says

$$\max_{t \in D_m} |Q_t^m - tQ_1^m| \xrightarrow[p]{} 0 \quad \text{as} \quad m \to \infty ,$$

and this is (10.18).

Remark. To remove the integrability hypothesis (10.13), note first that for non-integrable variables we can define conditional expectations "locally": $E(U|F) = V$ means that for every $A \in F$ for which $V1_A$ is integrable, we have that $U1_A$ is integrable and $E(U1_A|F) = V1_A$. In the non-integrable case, (10.16) remains true with this interpretation. To establish convergence of the "local" martingale Q_1^m it is necessary to show that $(Q_1^m: m \geq 1)$ is tight, and for this we can appeal to results on sampling without replacement in the spirit of (10.19). However, there must be some simpler way of making a direct proof of Theorem 10.12.

Let us describe briefly the general case of processes $(X_t: 0 \leq t \leq 1)$ with interchangeable increments, and refer the reader to Kallenberg (1973) for the precise result. Given a finite set $J = (x_i)$, there is one process with interchangeable increments, with jump sizes (x_i), and constant between jumps

$$X_t^J = \sum_i x_i 1_{(t \geq \xi_i)} \text{ ; where } (\xi_i) \text{ are independent } U(0,1).$$

This sum can also be given an interpretation for certain infinite sets J, as a L^2-limit. The resulting processes X^J are the "pure jump" processes; taking constants a, b and taking a Brownian bridge B^0 independent of (ξ_i), and putting

$$X_t = X_t^J + aB_t^0 + bt$$

gives a process X with interchangeable increments. These are the "ergodic" processes in the sense of Section 12; the general process with interchangeable increments is obtained by first choosing (J,a,b) at random from some prior distribution.

(10.20) The Dirichlet process. An interesting and useful instance of a process with interchangeable increments and increasing discontinuous paths is the family of Dirichlet processes, which we now describe.

Fix $a > 0$. Recall that the Gamma(b,1) distribution has density $\frac{1}{\Gamma(b)} x^{b-1}e^{-x}$ on $\{x \geq 0\}$. Since this distribution is infinitely divisible, there exists a Lévy process (X_t) such that X_1 has Gamma(a,1) distribution, and hence X_t has Gamma(at,1) distribution. Call X the Gamma(a) process. Here is an alternative description. Let ν be the measure on $(0,\infty)$ with density

$$\nu(dx) = ax^{-1}e^{-x}dx \; . \tag{10.21}$$

Then $\nu(\varepsilon,\infty) < \infty$ for $\varepsilon > 0$, but $\nu(0,\infty) = \infty$. Let $\nu \times \lambda$ be the product of ν and Lebesgue measure on $Q = \{(x,t): x > 0, t \geq 0\}$. Let N be a Poisson point process on Q with intensity $\nu \times \lambda$. Then N is distributed as the times and sizes of the jumps of X:

$$N \stackrel{\mathcal{D}}{=} \{(x,t): X_t - X_{t-} = x\} \; .$$

So we can construct X from N by adding up jumps:

$$X_t = \sum x \, 1_{\{(x,s) \in N, \; s \leq t\}} \; .$$

The Dirichlet(a) process is $Y_t = X_t/X_1$, $0 \leq t \leq 1$. Thus Y has increasing paths, interchangeable increments, $Y_0 = 0$, $Y_1 = 1$. (The relation between the Dirichlet process and the Gamma process is reminiscent of the relation between Brownian bridge and Brownian motion). The marginal distribution of Y_t is Beta$(at, a(1-t))$. As $a \to \infty$ the distribution of (Y_t) converges to the deterministic process t; as $a \to 0$ it converges to the single jump process $1_{(t \geq \xi)}$, ξ uniform on $[0,1]$. For $0 < t_1 < t_2 < \cdots < t_k = 1$ the increments $(Y_{t_1}, Y_{t_2} - Y_{t_1}, \ldots, Y_{t_k} - Y_{t_{k-1}})$ have the distribution $\mathcal{D}(at_1, a(t_2 - t_1), \ldots, a(t_k - t_{k-1}))$, where $\mathcal{D}(\alpha_1, \ldots, \alpha_k)$ is the distribution on the simplex $\{(y_1, \ldots, y_k): y_i \geq 0, \sum y_i = 1\}$ with density

$$\frac{\Gamma(\sum \alpha_i)}{\Pi \Gamma(\alpha_i)} \Pi y_i^{\alpha_i - 1}$$

with respect to Lebesgue measure on the simplex. In particular, the increments

(10.22) $$(Z_1, Z_2, \ldots, Z_k) = (Y_{1/k}, Y_{2/k} - Y_{1/k}, \ldots, Y_1 - Y_{(k-1)/k})$$

have density

$$\frac{\Gamma(a)}{\{\Gamma(a/k)\}^k} (\Pi z_i)^{-1+a/k}$$

on the simplex, and this is k-exchangeable.

Dirichlet processes have been extensively studied as prior distributions in Bayesian statistics. For we can view $Y_t(\omega)$ as the distribution function of a random measure $\beta(\omega, \cdot)$ on $[0,1]$ which is specified by the requirement

(10.23) $(\beta(\cdot, A_1), \ldots, \beta(\cdot, A_k))$ has distribution $\mathcal{D}(a\lambda(A_1), \ldots, a\lambda(A_k))$ for each partition (A_i).

This is the Dirichlet random measure associated with the measure $a\lambda(\cdot)$ on $[0,1]$. Given any finite measure α on a Borel space S, we can construct a Dirichlet random measure β satisfying (10.23) with α in place of $a\lambda(\cdot)$, by simply applying a function $f: [0,1] \to S$ which takes $a\lambda(\cdot)$ into α.

The advantage of using random measures of this form as priors is that posterior distributions can be handled analytically. For a survey of the statistical results, see Ferguson (1973, 1974).

By construction the Dirichlet random measure takes values in the set of purely atomic distributions. It is interesting to note that a random measure on $[0,1]$ with the property

(10.24) $(\beta(\cdot,A_1),\dots,\beta(\cdot,A_k))$ is exchangeable, for partitions (A_i) with $\lambda(A_i) = 1/k$

and which takes values in the set of continuous distributions, must be of the trivial form $\beta(\omega,\cdot) = \lambda(\cdot)$. For consider the distribution function $F_t(\omega) = \beta(\omega,[0,t])$. By (10.24), (F_t) has interchangeable increments, so if it is continuous then by Theorem 10.12 we must have $F_t = \alpha B_t + \beta t$ for some (α,β). But $F_1 = 1$ forces $\beta = 1$, and for F to be increasing we must have $\alpha = 0$.

Our uses of Dirichlet processes arise from entirely different considerations, the Kingman-Watterson treatment of the infinite allele model and the Ewens sampling formula, to be described in Sections 19 and 11. We need one more concept. The countable set of jump sizes $(Y_t - Y_{t-})$ of a sample path $Y(\omega)$ of a Dirichlet(a) process can be arranged in decreasing order as $(D_1,D_2,\dots)$. This describes a distribution on the infinite-dimensional simplex $\{(x_1,x_2,\dots): x_i \geq 0, \sum x_i = 1\}$ which is called the Poisson-Dirichlet(a) distribution. The name is explained by an alternative description using the measure ν of (10.21). Take a Poisson point process of intensity ν on $(0,\infty)$, and let $(V_1,V_2,\dots)$ be the points of this process, arranged in decreasing order. Then $S = \sum V_i < \infty$ a.s., because $ES = \int_{0+}^{\infty} x\, \nu(dx) = a$;

and $(D_1, D_2, \ldots) = (V_1/S, V_2/S, \ldots)$ has the Poisson-Dirichlet distribution. Explicit though complicated expressions for the marginals are given in Watterson (1976); see also Kingman (1980), Section 3.6.

As a brief indication of the kind of calculations which can be done, consider the "intensity" of the points (D_i):

$$\phi(y)dy = P(\text{some } D_i \in (y, y+dy)) .$$

We shall show

$$\phi(y) = ay^{-1}(1-y)^{a-1}, \quad 0 < y < 1 . \tag{10.25}$$

Consider first the intensity

$$\psi(v,s)dv\, ds = P(\text{some } V_i \in (v, v+dv),\ S = \textstyle\sum V_i \in (s, s+ds)) .$$

Conditional on some V being in $(v, v+dv)$, the remaining (V_i) still form a Poisson process of intensity ν, so their sum still has the Gamma$(a,1)$ density f_S, say. Thus

$$\begin{aligned} \psi(v,s)dv\, ds &= \nu(dv)f_S(s-v)ds \\ &= av^{-1}e^{-v}(s-v)^{a-1}e^{-(s-x)}dvds/\Gamma(a) . \end{aligned} \tag{10.26}$$

Now $D_i = V_i/S$, so the intensity

$$\begin{aligned} P(\text{some } D_i \in (y, y+dy),\ S \in (s, s+ds)) &= \psi(sy, s)\, s\, dy\, ds, \quad \text{putting } y = v/s \\ &= ay^{-1}(1-y)^{a-1}dy \cdot s^{a-1}e^{-s}/\Gamma(a) \quad \text{by (10.26)}. \end{aligned}$$

This product form shows that the intensity $\phi(y)$ satisfies (10.25), independent of S.

Observing now that at most one of the D_i can be greater than 1/2 (since $\sum D_i = 1$), we see from (10.25) that the density of D_1 is

$$f_{D_1}(y) = ay^{-1}(1-y)^{a-1} \quad \text{on} \quad 1/2 < y < 1 .$$

However, the expression for the density over $0 < y < 1/2$ is more complicated.

11. Exchangeable random partitions

In this section we describe recent work of Kingman and others.

Let $\mathcal{S}_N$ be the (finite) set of all partitions $A = (A_i)$ of $\{1,2,\ldots,N\}$. Call the sets A_i the components of A. By a random partition $\mathcal{R}$ we simply mean a random element of $\mathcal{S}_N$. Each permutation π of $\{1,2,\ldots,N\}$ acts on subsets $B \subset \{1,2,\ldots,N\}$ by $\pi(B) = \{\pi(i): i \in B\}$ and so acts on partitions by $\pi(A_1,A_2,\ldots) = (\pi(A_1),\pi(A_2),\ldots)$. Thus we call $\mathcal{R}$ exchangeable if $\pi(\mathcal{R}) \overset{\mathcal{D}}{=} \mathcal{R}$ for each π.

It is easy to describe the general exchangeable partition of a finite set (this is analogous to the description of the general finite exchangeable sequence in Section 5). Let $I_N = \{(n_i): n_1 \geq n_2 \geq \cdots \geq 0, \sum n_i = N\}$. Define $L_N: \mathcal{S}_N \to I_N$ by

(11.1) $\quad L_N((A_i))$ is the decreasing rearrangement of $(\#A_i)$.

Let π^* be the uniform random permutation. For any partition A, it is easy to see that $\pi^*(A)$ is uniform on $\{B: L_N(B) = L_N(A)\}$. Thus the general exchangeable partition $\mathcal{R}$ is obtained as follows:

(i) take a random element (η_i) of I_N;

(ii) conditional on $(\eta_i) = (n_i)$, let $\mathcal{R}$ be uniform on $\{A: L_N(A) = (n_i)\}$.

In particular, the distribution of $\mathcal{R}$ is determined by the distribution of $L_N(\mathcal{R})$.

An alternative description of partitions is sometimes useful. Given a partition $\mathcal{R}$, define

(11.2) $\quad R_{i,j} = \{\omega\colon i \text{ and } j \text{ in the same component of } \mathcal{R}(\omega)\}$,

and then

(11.3) $\quad R_{i,i} = \Omega;\ \ R_{i,j} = R_{j,i};\ \ R_{i,k} \supset R_{i,j} \cap R_{j,k};\ \ 1 \le i,j,k \le N$.

Conversely any family of events $(R_{i,j}\colon 1 \le i,j \le N)$ satisfying (11.3) defines a random partition $\mathcal{R}$, in which $\mathcal{R}(\omega)$ is the partition into equivalence classes of the equivalence relation

$$i \overset{\omega}{\sim} j \quad \text{iff} \quad \omega \in R_{i,j} .$$

Furthermore, $\mathcal{R}$ is exchangeable iff

(11.4) $\quad (R_{i,j}\colon 1 \le i,j \le N) \overset{\mathcal{D}}{=} (R_{\pi(i),\pi(j)}\colon 1 \le i,j \le N);$ each permutation π.

Consider now partitions of $\{1,2,3,\ldots\}$. The set $\mathcal{S}_\infty$ of all partitions is uncountable; we define a random partition to be a map $\mathcal{R}\colon \Omega \to \mathcal{S}_\infty$ such that the sets $R_{i,j}$ defined at (11.2) are measurable. As before, a random partition of $\{1,2,\ldots\}$ can be described as a family of events $\{R_{i,j}\colon 1 \le i,j < \infty\}$ satisfying (11.3) for each N. Note that one way to construct random partitions of $\{1,2,3,\ldots\}$ is by appealing to the Kolmogorov extension theorem: if for each N we have an exchangeable partition $\mathcal{R}^N$ satisfying the consistency conditions

(11.5) $\quad (R^{N+1}_{i,j}\colon 1 \le i,j \le N) \overset{\mathcal{D}}{=} (R^N_{i,j}\colon 1 \le i,j \le N);$ each $N \ge 1$,

then there exists an exchangeable random partition $\mathcal{R}$ of $\{1,2,3,\ldots\}$ such that

(11.6) $\quad (R_{i,j}\colon 1 \le i,j \le N) \overset{\mathcal{D}}{=} (R^N_{i,j}\colon 1 \le i,j \le N);$ each $N \ge 1$.

Our aim is to prove an analogue of de Finetti's theorem for exchangeable partitions of $\{1,2,3,\ldots\}$. The role of i.i.d. sequences is played by the "paintbox processes" which we now describe. Let μ be a distribution on $[0,1]$; think of μ as partly discrete and partly continuous. Let (X_i) be i.i.d. (μ). Let $R(\omega)$ be the partition with components $\{i: X_i(\omega) = x\}$, $0 \leq x \leq 1$. In other words, $R_{i,j} = \{\omega: X_i(\omega) = X_j(\omega)\}$. Clearly R is exchangeable. Kingman suggests the following mental picture: think of real numbers $0 \leq x \leq 1$ as labelling the colours of the spectrum; imagine colouring objects 1,2,3,... at random by painting object i with colour X_i; then we obtain a partition into sets of identically-coloured objects.

Clearly the distribution of R in this construction depends only on the sizes of the atoms of μ. Let $p_j = \mu(x_j)$, where (x_j) are the atoms of μ arranged so that (p_j) is decreasing (and put $p_j = 0$ if there are less than j atoms). This defines a map

(11.7) $$L(\mu) = (p_j)$$

from $P[0,1]$ into the set of possible sequences

$$\nabla = \{(p_1,p_2,\ldots): p_1 \geq p_2 \geq \cdots \geq 0, \textstyle\sum p_j \leq 1\} .$$

When $L(\mu) = \underset{\sim}{p}$, call the exchangeable partition R above the <u>paintbox</u>$(\underset{\sim}{p})$ <u>process</u>, and denote its distribution by ψ_p. The following facts are easy consequences of the strong law $N^{-1}\#(i \leq N: X_i = x_j) \to p_j$ a.s.

(11.8) <u>Lemma</u>. <u>Let</u> R <u>be a paintbox</u>$(\underset{\sim}{p})$ <u>process</u>.

(a) $N^{-1}L_N(R^N) \to (p_1,p_2,\ldots)$ <u>in</u> ∇ <u>a.s., where</u> R^N <u>is the restriction of</u> R <u>to</u> $\{1,2,\ldots,N\}$.

(b) <u>Let</u> C_1 <u>be the component of</u> $R(\omega)$ <u>containing</u> 1. <u>Then</u>

$$N^{-1}\#(C_1 \cap \{1,\dots,N\}) \xrightarrow{\mathcal{D}} p_J 1_{(J>0)} ,$$

where $P(J=j) = p_j$, $P(J=0) = 1 - \sum p_j$.

(c) $P(1,2,\dots,r$ in same component$) = \sum_{j\geq 1} p_j^r$; $r \geq 2$.

Here is the analogue of de Finetti's theorem, due to Kingman (1978b, 1982a).

(11.9) Proposition. Let $\mathcal{R}$ be an exchangeable partition of $\{1,2,\dots\}$, and let $\mathcal{R}^N$ be its restriction to $\{1,2,\dots,N\}$. Then

(a) $N^{-1}L_N(\mathcal{R}^N) \xrightarrow{a.s.} (D_1,D_2,\dots) = D$ for some random element D of ∇;

(b) ψ_D is a regular conditional distribution for $\mathcal{R}$ given $\sigma(\psi_D)$.

So (b) says that conditional on $D = \underset{\sim}{p}$, the partition $\mathcal{R}$ has the paintbox($\underset{\sim}{p}$) distribution ψ_p. As discussed in Section 2, this is the "strong" notion of $\mathcal{R}$ being a mixture of paintbox processes.

Proof. Let (ξ_i) be i.i.d. uniform on $(0,1)$, independent of $\mathcal{R}$. Throwing out a null set, we can assume the values $(\xi_i(\omega): i \geq 1)$ are distinct. Define

$$F_i(\omega) = \min\{j: i \text{ and } j \text{ in same component of } \mathcal{R}(\omega)\} \leq i$$

$$Z_i = \xi_{F_i} .$$

So for each ω the partition $\mathcal{R}(\omega)$ is precisely the partition with components $\{i: Z_i(\omega) = z\}$, $0 \leq z \leq 1$. We assert (Z_i) is exchangeable. For $(Z_i) = g((\xi_i),\mathcal{R})$ for a certain function g, and $(Z_{\pi(i)}) = g((\xi_{\pi(i)}),\pi(\mathcal{R}))$, and $((\xi_{\pi(i)}),\pi(\mathcal{R})) \overset{\mathcal{D}}{=} ((\xi_i),\mathcal{R})$ by exchangeability and by independence of $\mathcal{R}$ and (ξ_i).

Let α be the directing random measure for (Z_i). Then conditional on $\alpha = \mu$ the sequence (X_i) is i.i.d. (μ) and so $\mathcal{R}$ has the paintbox distribution $\psi_{L(\mu)}$. In other words $\psi_{L(\alpha)}$ is a regular conditional distribution for R given α, and this establishes (b) for $D = L(\alpha)$. And then (a) follows from Lemma 11.8(a) by conditioning on D.

Remarks. Kingman used a direct martingale argument, in the spirit of the first proof of de Finetti's theorem in Section 3. Our trick of labelling components by external randomization enables us to apply de Finetti's theorem. Despite the external randomization, (a) shows that D is a function of R.

Yet another proof of Proposition 11.9 can be obtained from deeper results on partial exchangeability--see (15.23).

The Ewens sampling formula. Proposition 11.5 is a recent result, so perhaps some presently unexpected applications will be found in future. The known applications involve situations where certain extra structure is present. An exchangeable random partition R on $\{1,2,3,\ldots\}$ can be regarded as a sequence of exchangeable random partitions R^N on $\{1,2,\ldots,N\}$ satisfying the consistency condition (11.5). Let us state informally another "consistency" condition which may hold. Fix N and $r < N$. Pick a partition $R^N = (A_i)$, and suppose $1 \in A_j$, where $\#A_j = r$. The remaining sets $(A_i\colon i \neq j)$ form a partition of $\{1,2,\ldots,N\}\backslash A_j$; the new condition is that this partition $(A_i\colon i \neq j)$ should be distributed as R^{N-r}.

To formalize this, we introduce more notation. For a partition $A = (A_i)$ of $\{1,2,\ldots,N\}$ define

(11.10) $a(A) = (a_1,\ldots,a_N)$, where a_j is the number of sets A_i for which $\#A_i = j$.

Of course $a(A)$ gives precisely the same information about A as does $L_N(A)$ used earlier. Given an exchangeable random partition R^N, let

(11.11) $$Q_N(a_1,\ldots,a_N) = P(a(R^N) = (a_1,\ldots,a_N)) .$$

By exchangeability, Q_N determines the distribution of R^N. Let $B \subset \{1,2,\ldots,N\}$ be such that $1 \in B$, $\#B = r$. Let $\hat{R}$ denote the partition of R^N with the set containing 1 removed. The condition we want is

(11.12) $$P\big(a(\hat{R}) = (a_1,\ldots,a_{r-1},a_r-1,a_{r+1},\ldots)\,|\,B \in R^N\big)$$
$$= Q_{N-r}(a_1,\ldots,a_{r-1},a_r-1,a_{r+1},\ldots) .$$

To rephrase this condition, observe that the left side equals

$$P\big(a(R^N) = a\,|\,B \in R^N\big), \quad \text{for} \quad a = (a_1,\ldots,a_r,\ldots)$$
$$= P\big(a(R^N) = a\big) \cdot P\big(B \in R^N\,|\,a(R^N) = a\big) \cdot 1/P(B \in R^N)$$
$$= Q_N(a) \cdot a_r / \binom{N}{r} \cdot 1/P\big(\{1,2,\ldots,r\} \in R^N\big) .$$

Thus condition (11.12) implies

$$\frac{a_r Q_N(a_1,\ldots,a_r,\ldots)}{Q_{N-r}(a_1,\ldots,a_r-1,\ldots)} \quad \text{depends only on} \quad (N,r) .$$

This is the basis for the proof of (11.16) below.

Now consider, for some exchangeable R, the chance that 1 and 2 belong to the same set in the partition:

(11.13) $$P\big(R^2 = \{1,2\}\big) = Q_2(0,1) = 1/(1+\theta), \quad \text{say, for some} \quad 0 \le \theta \le \infty .$$

There are two extreme cases: if $\theta = 0$ then R is a.s. the trivial partition $\{1,2,\ldots\}$; if $\theta = \infty$ then R is a.s. the discrete partition $(\{1\},\{2\},\{3\},\ldots)$. The interesting case is $0 < \theta < \infty$. It is a remarkable fact that if (11.12) holds then θ determines the distribution of R, and some explicit formulas can be obtained. We quote the following result (see Kingman (1980), Sections 3.5-3.7).

(11.14) Theorem. Let R be an exchangeable random partition on $\{1,2,\ldots\}$ satisfying (11.12). Define θ by (11.13), and suppose $0 < \theta < \infty$. Let $D = (D_1, D_2, \ldots)$ be as in Proposition 11.9. Then

(11.15) $(D_1, D_2, \ldots)$ has the Poisson-Dirichlet(θ) distribution,

(11.16) $$Q_N(a_1,\ldots,a_N) = \frac{N!}{\theta(\theta+1)\cdots(\theta+N-1)} \prod_{r=1}^{N} \frac{\theta^{a_r}}{r^{a_r} a_r!}$$

Equation (11.16) arose in genetics as in the Ewens sampling formula--see section 19.

As an example of the calculations which are possible, let R be an exchangeable random partition and let C_1 be the component of R containing 1. By Proposition 11.9 and (11.8b),

(11.17a) $$N^{-1}\#(C_1 \cap \{1,2,\ldots,N\}) \xrightarrow{\mathcal{D}} T = D_J 1_{(J \geq 1)} ,$$

where $P(J = j \mid D) = D_j$. The limit distribution T can be described by

$$P(T \in (t, t+dt)) = t \sum_{j \geq 1} P(D_j \in (t, t+dt)) .$$

So under the hypotheses of Theorem 11.14 we can apply (10.24) to obtain the density

(11.17b) $$P(T \in (t, t+dt)) = \theta(1-t)^{\theta-1} dt , \quad 0 < t < 1 .$$

Equation (11.16) readily yields various special cases:

(11.18a) $$P(1,2,\ldots,N \text{ in same component}) = \prod_{i=2}^{N} \frac{i-1}{\theta+i-1}$$

(11.18b) $$P(1,2,\ldots,N \text{ in distinct components}) = \prod_{i=1}^{N} \frac{\theta}{\theta+i-1}$$

Finally, we remark that there is a "sequential" description of the partitions R^N of Theorem 11.14; following Jim Pitman, we call this

(11.19) <u>The Chinese restaurant process</u>. Imagine people $1,2,\ldots,N$ arriving sequentially at an initially empty restaurant with a large number of large tables. Person j either sits at the same table as person i (with probability $1/(j-1+\theta)$, for each $i < j$), or else sits at an empty table (with probability $\theta/(j-1+\theta)$). Then the partition "people at each table" has the distribution of Theorem 11.14.

Although Theorem 11.14 was motivated by a problem in genetics, it has some purely mathematical applications. The first was been noted by Kingman and others.

<u>Cycle length in random permutations</u>. Let π_N^* be the uniform random permutation on $\{1,2,\ldots,N\}$, and let R^N be the partition into cycles of π_N^*. We claim that the exchangeable random partitions R^N, $N \geq 1$ are consistent in the sense of (11.5). To see this, for a permutation σ of $\{1,\ldots,N+1\}$ define a permutation $\hat{\sigma} = g(\sigma)$ of $\{1,\ldots,N\}$ by deleting $N+1$ from the cycle representation of σ:

$$\begin{aligned} \hat{\sigma}(i) &= \sigma(i) && \text{if } \sigma(i) \neq N+1 \\ &= \sigma(N+1) && \text{if } \sigma(i) = N+1 . \end{aligned}$$

Then

$$g(\pi_{N+1}^*) \stackrel{\mathcal{D}}{=} \pi_N^* \ ;$$

a pair $i,j \leq N$ are in the same cycle of $g(\pi_{N+1}^*(\omega))$ iff they are in the same cycle of $\pi_{N+1}^*(\omega)$; and this implies consistency.

Next, consider $1 \in B \subset \{1,2,\ldots,N\}$. Then conditional on B being a cycle of π_N^*, the restriction of π_N^* to $\{1,\ldots,N\}\backslash B$ is distributed uniformly on the set of all permutations of $\{1,\ldots,N\}\backslash B$. This establishes the consistency property (11.12).

So Theorem 11.14 is applicable. To evaluate θ, let $X_0 = 1$, $X_i = \pi_N^*(X_{i-1})$. Then $P(X_i = 1 | X_1,\ldots,X_{i-1} \notin \{1,2\}) = P(X_i = 2 | X_1,\ldots,X_{i-1} \notin \{1,2\}) = (N-i+1)^{-1}$, and so $P(1 \text{ and } 2 \text{ are in same cycle of } \pi_N^*) = \frac{1}{2}$. Thus $\theta = 1$.

From (11.16) we obtain a classical theorem of Cauchy.

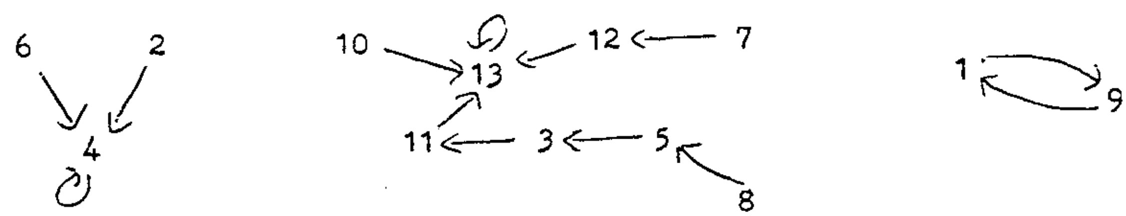
(11.20) Corollary. Let $a_r \geq 0$, $\sum r a_r = N$. The number of permutations of $\{1,\ldots,N\}$ with exactly a_r cycles of length r (each $r \geq 1$) is

$$N! \prod_{r \geq 1} \frac{1}{r^{a_r} a_r!}.$$

And if $(M_{N,1}, M_{N,2},\ldots)$ are the lengths of the cycles of a uniform random permutation of $\{1,\ldots,N\}$, arranged in decreasing order, then (11.15) says that $N^{-1}(M_{N,1}, M_{N,2},\ldots)$ converges in distribution to the Poisson-Dirichlet(1) process.

Remarks. There is a large literature on random permutations; we mention only a few recent papers related to the discussion above. Vershik and Schmidt (1977) give an interesting "process" description of the Poisson-Dirichlet(1) limit of $N^{-1}(M_{N,1}, M_{N,2},\ldots)$. Ignatov (1982) extends their ideas to general θ. Kerov and Vershik (1982) use the ideas of Theorem 11.14 in the analysis of some deeper structure of random permutations (e.g. Young's tableaux).

<u>Components of random functions</u>. A function $f: \{1,\ldots,N\} \to \{1,\ldots,N\}$ defines a directed graph with edges $i \to f(i)$ for each i. Thus f induces a partition of $\{1,\ldots,N\}$ into the components of this graph. The partition can be described by saying that i and j are in the same component iff $f^{k*}(i) = f^{m*}(j)$ for some $k,m \geq 1$; where $f^{k*}(i)$ denotes the k-fold iteration $f(f(\ldots f(i)\ldots))$.



If we now let F_N be a random function, uniform over the set of all N^N possible functions, then we get a random partition R^N of $\{1,\ldots,N\}$ into the components of F_N. Clearly R^N is exchangeable. We shall outline how Theorem 11.14 can be used to get information about the asymptotic (as $N \to \infty$) sizes of the components.

<u>Remark</u>. Many different questions can be asked about iterating random functions. For results and references see e.g. Knuth (1981), pp. 8, 518-520; Pavlov (1982); Pittel (1983). I do not know references to the results outlined below, though I presume they are known.

Given that a specified set $B \subset \{1,\ldots,N\}$ is a component of F_N, it is clear that F_N restricted to $\{1,\ldots,N\}\backslash B$ is uniform on the set of functions from $\{1,\ldots,N\}\backslash B$ into itself. Thus

$$R^N, \quad N \geq 1, \quad \text{satisfies the consistency condition (11.12).}$$

However, the consistency condition (11.5) is not satisfied exactly, but rather holds in an asymptotic sense. To make this precise, for $K < N$ let

$R^{N,K}$ be the restriction of R^N to $\{1,\ldots,K\}$. Then

(11.21) Lemma. $R^{N,K} \xrightarrow{\mathcal{D}} \hat{R}^K$, say, as $N \to \infty$.

For each N the family $(R^{N,K};\ K < N)$ is consistent in the sense of (11.5), and so Lemma 11.21 implies that $(\hat{R}^K;\ K \geq 1)$ is consistent in that sense. It is then not hard to show

(11.22) Lemma. $(\hat{R}^K;\ K \geq 1)$ is consistent in sense (11.12).

Then Theorem 11.14 is applicable to $(\hat{R}^K)$. To identify θ, we need

(11.23) Lemma. $P(1$ and 2 in same component of $R^N) \to 2/3$ as $N \to \infty$.

Then Lemma 11.21 implies $P(\hat{R}^2 = \{1,2\}) = 2/3$, and so (11.13) identifies

$$\theta = \frac{1}{2}.$$

Now Theorem 11.14 gives information about $(\hat{R}^K)$. For instance, writing $\hat{C}_1^K$ for the component of $\hat{R}^K$ containing 1, (11.17) says

(11.24) $K^{-1}\ \#\hat{C}_1^K \xrightarrow{\mathcal{D}} T$, where T has density $f(t) = \frac{1}{2}(1-t)^{-1/2}$, $0 < t < 1$.

We are really interested in the corresponding assertion

$$N^{-1}\ \#C_1^N \xrightarrow{\mathcal{D}} T \tag{11.25}$$

where C_1^N is the component of the random function F_N containing 1. Let us indicate how to pass from (11.24) to (11.25). Lemma 11.21 shows

(a) $\left\| L(\#\hat{C}_1^K) - L(\#(C_1^N \cap \{1,\ldots,K\})) \right\|_0 \to 0$ as $N \to \infty$; K fixed,

where $\|\ \|_0$ is total variation distance (5.5). By exchangeability, $\#(C_1^N \cap \{2,\ldots,K\})$ is distributed as the sum of $K-1$ draws without replacement from an urn with $\#C_1^N - 1$ "1"s and $N - \#C_1^N$ "0"s. Let $V_{N,K}$ be the corresponding sum with replacement. As $N \to \infty$, sampling with or without replacement become equivalent, so

(b) $\|L(\#(C_1^N \cap \{2,\ldots,K\})) - L(V_{N,K})\|_0 \to 0$ as $N \to \infty$; K fixed.

Now $V_{N,K}$ is conditionally Binomial$(K-1,(N-1)^{-1}(\#C_1^N-1))$ given $\#C_1^N$, so by the weak law of large numbers

(c) $\lim\limits_{K\to\infty} \limsup\limits_{N\to\infty} E|K^{-1}V_{N,K} - N^{-1}\#C_1^N| = 0$.

Properties (a)-(c) lead from (11.24) to (11.25).

The same argument establishes the result corresponding to (11.15); if $(M_1^N, M_2^N, \ldots)$ are the sizes of the components of F_N arranged in decreasing order, then $N^{-1}(M_1^N, M_2^N, \ldots)$ converges in distribution to the Poisson-Dirichlet$(\frac{1}{2})$ process.

Let us outline the proofs of Lemmas 11.21 and 11.23. Let $X_0 = 1$, $X_n = F_N(X_{n-1})$, $S_1 = \min\{n: X_n(\omega) \in \{X_0(\omega),\ldots,X_{n-1}(\omega)\}\}$. Then (X_n) is i.i.d. uniform until time S_1, and we get the simple formula

$$P(S_1 \geq n) = \prod_{m=0}^{n-1} (1 - m/N) .$$

Of course this is just "the birthday problem." Calculus gives

$$N^{-1/2}S_1 \xrightarrow{\mathcal{D}} \hat{S}_1, \text{ where } \hat{S}_1 \text{ has density } f(s) = s\cdot\exp(-s^2/2) .$$

Now let $Y_0 = 2$, $Y_n = F_N(Y_{n-1})$, $S_2 = \min\{n: Y_n(\omega) \in \{X_0(\omega),\ldots,X_{S_1-1}(\omega), Y_0(\omega),\ldots,Y_{n-1}(\omega)\}\}$, and let A_N be the event $\{Y_{S_2} \in \{X_0,\ldots,X_{S_1-1}\}\} =$ {1 and 2 in same component of F_N}. Again (Y_n) is i.i.d. uniform until time S_2, there is a simple formula for $P(A_N, S_2 = n | S_1 = q)$, and calculus gives $(A_N, S_1, S_2) \xrightarrow{\mathcal{D}} (\hat{A}, \hat{S}_1, \hat{S}_2)$, where the limit has density

$$P(\hat{A},\ \hat{S}_2 \in (s_2, s_2+ds_2),\ \hat{S}_1 \in (s_1, s_1+ds_1)) = s_1^2 \cdot \exp(-\tfrac{1}{2}(s_1+s_2)^2)ds_1 ds_2 \ .$$

Integrating this density gives $P(\hat{A}) = 2/3$, and this is Lemma 11.23.

Call the process $(X_n,\ n < S_1;\ Y_n,\ n < S_2)$ above the N-process. Let B_N be the event "some element X_n or Y_n of the N-process equals N, and $F_N(N) = N$". Construct a process $(X_n^*,\ n < S_1^*;\ Y_n^*,\ n < S_2^*)$ by deleting any terms of the N-process which equal N. Then conditional on B_N^c the process (X_n^*, Y_n^*) is distributed as the (N-1)-process. So

$$|P(1 \text{ and } 2 \text{ in same component of } F_N) - P(1 \text{ and } 2 \text{ in same component of } F_{N-1})| \\ \leq P(B_N) \ .$$

But $P(B_N) \leq c_2 N^{-3/2}$, since S_1 and S_2 are of order $N^{1/2}$ and $P(F_N(N) = N) = N^{-1}$. A similar argument considering iterates of F_N starting from $1,2,\ldots,K$ shows

$$|P(R^{N,K} = A) - P(R^{N-1,K} = A)| \leq c_K N^{-3/2} \ ; \quad A \in S_K \ .$$

Since $\sum N^{-3/2} < \infty$ this gives Lemma 11.21.

(11.26) Random graphs. Another way to construct random graphs on N vertices is to have each edge present with probability λ/N, independent for different possible edges. In this case the component containing "1", C_1^N, satisfies

$$N^{-1}\ \#C_1^N \xrightarrow{\mathcal{D}} T \quad \text{as} \quad N \to \infty \ ,$$

where $T = 0$ for $\lambda \leq 1$; $P(T = c(\lambda)) = c(\lambda)$, $P(T = 0) = 1 - c(\lambda)$ for $\lambda > 1$, where $c(\lambda) > 0$ for $\lambda > 1$. Thus (c.f. 11.17) the partition into components of these random graphs cannot be fitted into the framework of Theorem 11.14. It would be interesting to know which classes of random graphs had components following the Poisson-Dirichlet distribution predicted by Theorem 11.14.

PART III

The class of exchangeable sequences can be viewed as the class of distributions invariant under certain transformations. So a natural generalization of exchangeability is to consider distributions invariant under families of transformations. Section 12 describes some abstract theory; Sections 13-16 some particular cases.

12. Abstract results

In the general setting of distributions invariant under a group of transformations, there is one classical result: that each invariant measure is a mixture of ergodic invariant measures. We shall discuss this result (Theorem 12.10) without giving detailed proof; we then specialize to the "partial exchangeability" setting, and pose some hard general problems.

Until further notice, we work in the following general setting:

(12.1) S is a Polish space; $\mathcal{T}$ is a countable group (under composition) of measurable maps $T: S \to S$.

Call a random element X of S invariant if

$$T(X) \overset{\mathcal{D}}{=} X \,; \quad T \in \mathcal{T} \,. \tag{12.2}$$

Call a distribution μ on S invariant if it is the distribution of an invariant random element, i.e. if

$$\tilde{T}(\mu) = \mu \,; \quad T \in \mathcal{T} \tag{12.3}$$

where $\tilde{T}$ is the induced map (7.3). Let M denote the set of invariant distributions, and suppose M is non-empty. Call a subset A of S invariant if

(12.4) $$T(A) = A\ ; \quad T \in \mathcal{T}\ .$$

The family of invariant subsets forms the invariant σ-field $\mathcal{J}$. Call an invariant distribution μ ergodic if

(12.5) $$\mu(A) = 0 \text{ or } 1; \quad \text{each } A \in \mathcal{J}\ .$$

We quote two straightforward results:

(12.6) If μ is invariant and if A is an invariant set with $\mu(A) > 0$ then the conditional distribution $\mu(\cdot|A)$ is invariant.

(12.7) If μ is ergodic, and if a subset B of S is almost invariant, in the sense that $T(B) = B$ μ-a.s., each $T \in \mathcal{T}$, then $\mu(B) = 0$ or 1.

The two obvious examples (to a probabilist!) are the classes of stationary and of exchangeable sequences. To obtain stationarity, take $S = \mathbb{R}^{\mathbb{Z}}$, $\mathcal{T} = (T_k;\ k \in \mathbb{Z})$, where T_k is the "shift by k" map taking (x_i) to (x_{i-k}). Then a sequence $X = (X_i\colon i \in \mathbb{Z})$ is stationary iff it is invariant under $\mathcal{T}$, and the definitions of "invariant σ-field", "ergodic" given above are just the definitions of these concepts for stationary sequences.

To obtain exchangeability, take $S = \mathbb{R}^{\infty}$, $\mathcal{T} = (T_\pi)$, where for a finite permutation π the map T_π takes (x_i) to $(x_{\pi(i)})$. Then a sequence $X = (X_i)$ is exchangeable iff it is invariant under $\mathcal{T}$. The invariant σ-field $\mathcal{J}$ here is the exchangeable σ-field of Section 3; the ergodic processes are those with trivial exchangeable σ-fields, which by Corollary 3.10 are precisely the i.i.d. sequences.

Returning to the abstract setting, the set $\mathcal{M}$ of invariant distributions is convex:

(12.8) $\mu_1, \mu_2 \in M$ implies $\mu = c\mu_1 + (1-c)\mu_2 \in M$; $0 < c < 1$.

So we can define an extreme point: μ is extreme if the only representation of μ as $c\mu_1 + (1-c)\mu_2$, $\mu_i \in M$, has $\mu_1 = \mu_2 = \mu$.

(12.9) Lemma. *An invariant distribution is extreme in M iff it is ergodic.*

Proof. Let μ be invariant but not ergodic. Then $0 < \mu(A) < 1$ for some invariant A. And $\mu = P(A)\mu(\cdot|A) + P(A^c)\mu(\cdot|A^c)$ represents μ as a linear combination of invariant distributions, by (12.6), so μ is not extreme.

Conversely, suppose μ is ergodic, and suppose $\mu = c\mu_1 + (1-c)\mu_2$ for invariant μ_i, $0 < c < 1$. Let f be the Radon-Nikodym derivative $d\mu_1/d\mu$. Then by invariance, $f = f\circ T$ μ-a.s., each $T \in T$. So sets of the form $\{f \geq a\}$, a constants, are almost invariant in the sense of (12.7), which implies f is μ-a.s. constant. This implies $\mu_1 = \mu$, and hence μ is extreme.

Write E for the set of ergodic (= extreme) distributions. For an invariant random element X, write J_X for the σ-field of sets $\{X \in A\}$, $A \in J$. We can now state the abstract result: the reader is referred to Dynkin (1978), Theorems 3.1 and 6.1 for a proof (in somewhat different notation) and further results.

(12.10) Theorem. *(a) E is a measurable subset of $P(S)$.*

(b) Let X be an invariant random element. Let α be a r.c.d. for X given J_X. Then $\alpha(\omega) \in E$ a.s.

(c) To each invariant distribution μ there corresponds a distribution Λ_μ on E such that $\mu(\cdot) = \int_E \nu(\cdot)\Lambda_\mu(d\nu)$.

(d) The distribution Λ_μ in (c) is unique.

(12.11) Remarks. (i) Assertions (b) and (c) are different ways of saying that an invariant distribution is a mixture of ergodic distributions, corresponding to the "strong" and "weak" notions of mixture discussed in Section 2.

(ii) Dynkin (1978) proves the theorem directly. Maitra (1977) gives another direct proof. Parts (a), (c), (d) may alternatively be deduced from general results on the representation of elements of convex sets as means of distributions on the extreme points. See Choquet (1969), Theorem 31.3, for a version of Theorem 12.10 under somewhat different hypotheses; see also Phelps (1966).

(iii) In the usual case where $\mathcal{T}$ consists of continuous maps T, the set $\mathcal{M}$ is closed (in the weak topology). But $\mathcal{E}$ need not be closed; for instance, in the case of stationary $\{0,1\}$-valued sequences $\mathcal{E}$ is a dense G_δ.

(iv) It is easy to see informally why (b) should be true. Fix X and α. Property (12.6) implies that α is also a r.c.d. for $T(X)$ given $\mathcal{J}_X$, and so

$$\tilde{T}(\alpha(\omega)) = \alpha(\omega) \text{ a.s.; each } T \in \mathcal{T} .$$

Since $\mathcal{T}$ is countable, this implies

$$\alpha(\omega) \in \mathcal{M} \text{ a.s.}$$

Now for $A \in \mathcal{J}$ we have $\alpha(\omega,A) = P(X\in A|\mathcal{J}_X) = 1_{(X\in A)}$ a.s. Then for a sub-σ-field $\mathcal{A}$ of $\mathcal{J}$ generated by a countable sequence

(12.11a) $$\mathcal{A} = \sigma(A_1,A_2,A_3,\ldots) \subset \mathcal{J}$$

we have

$$P(\omega: \alpha(\omega,A) = 0 \text{ or } 1 \text{ for each } A\in\mathcal{A}) = 1 .$$

Unfortunately we cannot conclude

$$P\big(\omega:\ \alpha(\omega,A) = 0 \text{ or } 1 \text{ for each } A \in \mathcal{J}\big) = 1$$

because in general the invariant σ-field $\mathcal{J}$ itself cannot be expressed in the form (12.11a); this technical difficulty forces proofs of Theorem 12.10 to take a less direct approach.

Proposition 3.8 generalizes to our abstract setting in the following way (implicit in the "sufficiency" discussion in Dynkin (1978)).

(12.12) Proposition. Let X, V be random elements of S, S' such that $(X,V) \stackrel{\mathcal{D}}{=} (T(X),V)$ for each $T \in \mathcal{T}$ (so in particular X is invariant). Then X and V are conditionally independent given $\mathcal{J}_X$.

Proof. Consider first the special case where X is ergodic. If $0 < P(V \in B) < 1$ then X is a mixture of the conditional distributions $P(X \in \cdot \mid V \in B)$ and $P(X \in \cdot \mid V \in B^c)$ which are invariant by hypothesis; so by extremality $\mathcal{L}(X) = \mathcal{L}(X \mid V \in B)$, and so X and V are independent.

For the general case, let $\mathcal{M}^2$ be the class of distributions of (X^*,V^*) on $S \times S'$ such that $(X^*,V^*) \stackrel{\mathcal{D}}{=} (T(X^*),V^*)$ for each T. Let ψ be a r.c.d. for (X,V) given $\mathcal{J}_X$.

We assert

$$\psi(\omega) \in \mathcal{M}^2 \text{ a.s.} \tag{12.13}$$

Observe first that by the countability of $\mathcal{T}$ and (7.11), there exists a countable subset (h_i) of $C(S \times S')$ such that

$$\theta \in \mathcal{M}^2 \text{ iff } \int h_i \, d\theta = 0, \quad i \geq 1 . \tag{12.14}$$

Next, for $A \in \mathcal{J}$ with $P(X \in A) > 0$ the hypothesis implies that the conditional distribution of (X,V) given $\{X \in A\}$ is in $\mathcal{M}^2$. Thus if ψ_n is a r.c.d. for (V,X) given a finite sub-σ-field F_n of J_X, then $\psi_n(\omega) \in M^2$ a.s. Because J_X is a sub-σ-field of the separable σ-field $\sigma(X)$, it is essentially separable, that is $J_X = F_\infty$ a.s. for some separable F_∞. Taking finite σ-fields $F_n \uparrow F_\infty$, Lemma 7.14(b) says $\psi_n \to \psi$ a.s. Since $\psi_n \in M^2$ a.s., (12.14) establishes (12.13).

For $\theta \in P(S \times S')$ let $\theta_1 \in P(S)$ be the marginal distribution. Then $\psi_1(\omega)$ is a r.c.d. for X given J_X, and so by Theorem 12.10 $\psi_1(\omega) \in E$ a.s. But now for each ω we can use the result in the special case to show that $\psi(\omega)$ is a product measure; this establishes the Proposition.

(12.15) Remarks. (i) In the case where T is a compact group (in particular, when T is finite), let T^* be a random element of T with Haar distribution (i.e. uniform, when T is finite). Then for fixed $s \in S$ the random element $T^*(s)$ is invariant; and it is not hard to show that the set of distributions of $T^*(s)$, as s varies, is the set of ergodic distributions. This is an abstract version of Lemma 5.4 for finite exchangeable sequences.

(ii) In the setting (12.1), call a distribution μ quasi-invariant if for each $T \in T$ the distributions μ and $\tilde{T}(\mu)$ are mutually absolutely continuous. Much of the general theory extends to this setting: any quasi-invariant distribution is a mixture of ergodic quasi-invariant distributions. See e.g. Blum (1982); Brown and Dooley (1983).

So far we have been working under assumptions (12.1). We now specialize. Suppose

(12.16) I is a countable set, Γ is a countable group (under convolution) of maps $\gamma: I \to I$, and S is a Polish space.

By a process X we mean a family $(X_i : i \in I)$ with X_i taking values in S.

The process is invariant if

(12.17) $$X \overset{\mathcal{D}}{=} (X_{\gamma(i)} : i \in I), \quad \text{each } \gamma \in \Gamma.$$

To see that this is a particular case of (12.1), take $S^* = S^I$, and take T to be the family $(\gamma^* : \gamma \in \Gamma)$ of maps $S^* \to S^*$, where γ^* maps (x_i) to $(x_{\gamma(i)})$. Note that γ^* is continuous (in the product topology on S^I), and so the class of invariant processes is closed under weak convergence.

Obviously any exchangeable process X is invariant. We use the phrase partially exchangeable rather loosely to indicate the invariant processes when (I,Γ) is given. Our main interest is in

(12.18) The Characterization Problem. Given (I,Γ), can we say explicitly what the partially exchangeable processes are?

Remarks. (a) To the reader who answers "the ones satisfying (12.17)" we offer the following analogy: the definition of a "simple" finite group tells us when a particular group is simple; the problem of saying explicitly what are the finite simple groups is harder!

(b) In view of Borel-isomorphism (Section 7), the nature of S is essentially unimportant for the characterization problem: one can assume $S = [0,1]$.

Theorem 12.10 gives some information: any partially exchangeable process is a mixture of ergodic partially exchangeable processes. This is all that is presently known in general. But there seem possibilities for better results, as we now describe.

Suppose we are given a collection $\mathcal{M}_0$ of invariant processes. How can we construct from these more invariant processes? In the general setting (12.1), the only way apparent is to take mixtures. Thus it is natural to ask

(12.19) What is the minimal class M_0 from which we can obtain all invariant distributions by taking mixtures?

Of course Theorem 12.10 tells us that the minimal class is E, the ergodic distributions. However, in the partial exchangeability setting (12.17) there are other ways to get new processes from old. Let X be an (ergodic) partially exchangeable process, and let $f: S \to S'$ be a function; then $Y_i = f(X_i)$ defines an (ergodic) partially exchangeable process Y. Let X' be another (ergodic) partially exchangeable process, independent of X; then $Z_i = (X_i, X'_i)$ defines an (ergodic) partially exchangeable process Z. Combining these ideas, we get

> Let M_0 be a class of (ergodic) partially exchangeable processes. Let $\hat{M}_0$ be the class of processes of the form $Y_i = f(X_i^1, X_i^2, X_i^3, \ldots)$, where each process X^k is in M_0, the family (X^k) are independent, and f is any function. Then $\hat{M}_0$ consists of (ergodic) partially exchangeable processes.

In view of this observation, it is natural to pose the problem analogous to (12.19).

(12.20) <u>Problem</u>. Is there a minimal class M_0 such that $\hat{M}_0 = E$, the class of all ergodic processes? Can this minimal class be specified abstractly?

Nothing is known about this problem in general. Let us discuss Problems 12.18 and 12.20 in connection with the examples already mentioned; in the next sections we will discuss further examples.

For the class of exchangeable sequences, de Finetti's theorem answers Problem 12.18, and the ergodic processes are the i.i.d. sequences. Now recall (7.6) that for any distribution μ on a Polish space S there exists a function $f: [0,1] \to S$ such that $f(\xi)$ has distribution for ξ with $U(0,1)$ distribution. Thus for exchangeable sequences we can take the class M_0 in (12.20) to consist of a single element, the i.i.d. $U(0,1)$ sequence.

For the simple generalizations of exchangeability described in Section 9, the results in Section 9 answer Problem 12.18. For each of those hypotheses it is not difficult to identify the ergodic processes and to describe a class M_0 in Problem 12.20 consisting of a single element. Let us describe the hardest case, and leave the others to the interested reader.

Hypotheses (9.1) and (9.3). As in Section 9 let $h(x,y) = (y,x)$, so that for a distribution μ on $S \times S$ the distribution $\tilde{h}(\mu)$ is the measure obtained by interchanging the coordinates. Let $Z_i = (X_i, Y_i)$ satisfy (A) and (C). Then Z is ergodic iff its directing random measure α is of the form

$$(12.21) \qquad P(\alpha = \mu) = P(\alpha = \tilde{h}(\mu)) = \tfrac{1}{2}; \quad \text{some } \mu \in P(S \times S) .$$

The existence of a single-element class M_0 follows from the next lemma, since we can use the process satisfying (12.21) for the particular distribution μ_0 below.

(12.22) Lemma. There exists a distribution μ_0 on $[0,1] \times [0,1]$ such that for any distribution μ on $S \times S$ there exists a function $f: [0,1] \to S$ such that $(f(U_1), f(U_2))$ has distribution μ when (U_1, U_2) has distribution μ_0.

Proof. Let $\Omega = [0,1]$. By considering the set of random elements $\underset{\sim}{V} = (V_1, V_2)\colon \Omega \to [0,1]^2$ as a subset of Hilbert space $L^2(\Omega; [0,1]^2)$, we see that there exists an L^2-dense sequence $\underset{\sim}{V}^1, \underset{\sim}{V}^2, \ldots$. Set

$V_j = (V_j^1, V_j^2, V_j^3, \ldots)$, $j = 1,2$. Now given μ, there exists a random element $\underset{\sim}{W}: \Omega \to [0,1]^2$ with distribution μ. By L^2-denseness, there exists a subsequence $V^{n_k} \to \underset{\sim}{W}$ in L^2, and by passing to a further subsequence we may take $\underset{\sim}{V}^{\tilde{n}_k} \to \underset{\sim}{W}$ a.s. Setting $g(x) = \limsup x_{n_k}$ gives $L(g(V_1), g(V_2)) = \mu$. Finally, let $\phi: [0,1]^\infty \to A \subset [0,1]$ be a Borel-isomorphism, and set $\mu_0 = L(\phi(V_1), \phi(V_2))$. Then $f = g \circ \phi^{-1}$ satisfies the assertion of the lemma.

For the class of stationary sequences, these types of problems are much harder. Informally, the class seems too big to allow any simple explicit description of the general stationary ergodic process. Let us just describe one result in this area. Let S be finite. Let $(\xi_i)_{i \in \mathbb{Z}}$ be i.i.d. $U(0,1)$. Let $f: [0,1]^{\mathbb{Z}} \to S$ and let

$$Y_k = f((\xi_{i+k})_{i \in \mathbb{Z}}); \quad \underset{\sim}{Y} = (Y_k)_{k \in \mathbb{Z}} \tag{12.23}$$

Then $\underset{\sim}{Y}$ is a stationary ergodic S-valued sequence. Ornstein showed that not all stationary ergodic processes can be expressed in this form: in fact, $\underset{\sim}{Y}$ can be represented in form (12.23) iff $\underset{\sim}{Y}$ satisfies a condition called "finitely determined." See Shields (1979) for an account of these results.

(12.24) Second-order structure. When a partially exchangeable process $X = (X_i : i \in I)$ is real-valued square-integrable, it has a correlation function $\rho(i,j) = \rho(X_i, X_j)$ which plainly satisfies

$$\begin{aligned} &\rho \text{ is non-negative definite;} \\ &\rho \text{ is invariant, that is } \rho(i,j) = \rho(\gamma(i), \gamma(j)), \quad \gamma \in \Gamma. \end{aligned} \tag{12.25}$$

Conversely, any ρ satisfying (12.25) is the correlation function of some (Gaussian) partially exchangeable process. Thus the study of the second-order structure of partially exchangeable processes reduces to the study of functions satisfying (12.25), and this can be regarded as a problem in abstract harmonic analysis. In the classical case of stationary sequences, ρ satisfies (12.25) iff $\rho(i,j) = \rho(j-i)$, where $\rho(n) = \int_0^{\pi} \cos(n\lambda)\mu(d\lambda)$ for some distribution μ on $[0,\pi]$. For exchangeable sequences, the possible correlations were described in Section 1. For the examples in the following sections, we shall obtain the correlation structure as a corollary of the (much deeper) probabilistic structure; and give references to alternative derivations using harmonic analysis. Letac (1981b) is a good introduction to the analytic methods.

(12.26) Topologies on $\mathcal{M}$. Finally, we mention some abstract topological questions. Consider the general setting (12.1), and suppose the maps T are continuous. The set $\mathcal{M}$ of invariant distributions inherits the weak topology from $\mathcal{P}(S)$, but there is another topology which can be defined on $\mathcal{M}$. Call a pair (X,Y) of S-valued random elements jointly invariant if $(X,Y) \stackrel{\mathcal{D}}{=} (T(X),T(Y))$, $T \in \mathcal{T}$. Let d be a bounded metric on S. Then

$$(12.27) \quad \bar{d}(\mu,\nu) = \inf\{Ed(X,Y): (X,Y) \text{ jointly invariant, } \mathcal{L}(X)=\mu,\ \mathcal{L}(Y)=\nu\}$$

defines a metric on $\mathcal{M}$ which is stronger than (or equivalent to) the weak topology. Now specialize to the partial exchangeability setting (12.16), and suppose there exists $i_0 \in I$ such that $\{\gamma(i_0): \gamma \in \Gamma\} = I$. Then processes (X,Y) are jointly invariant if $((X_i,Y_i): i \in I) \stackrel{\mathcal{D}}{=} ((X_{\gamma(i)},Y_{\gamma(i)}): i \in I)$ for $\gamma \in \Gamma$; and

$$(12.28) \quad \bar{d}(\mu,\nu) = \inf\{Ed(X_{i_0},Y_{i_0}): (X,Y) \text{ jointly invariant, } \mathcal{L}(X)=\mu,\ \mathcal{L}(Y)=\nu\}$$

defines a metric equivalent to that of (12.27). This $\bar{d}$-topology was introduced by Ornstein for stationary sequences; see Ornstein (1973) for a survey. A discussion of its possible uses in partial exchangeability is given in Aldous (1982a). Informally, the $\bar{d}$-topology seems more compatible with the characterization problem than is the weak topology (for instance, E is always $\bar{d}$-closed). A natural open problem is:

(12.29) Problem. Under what conditions are the $\bar{d}$-topology and the weak topology equivalent?

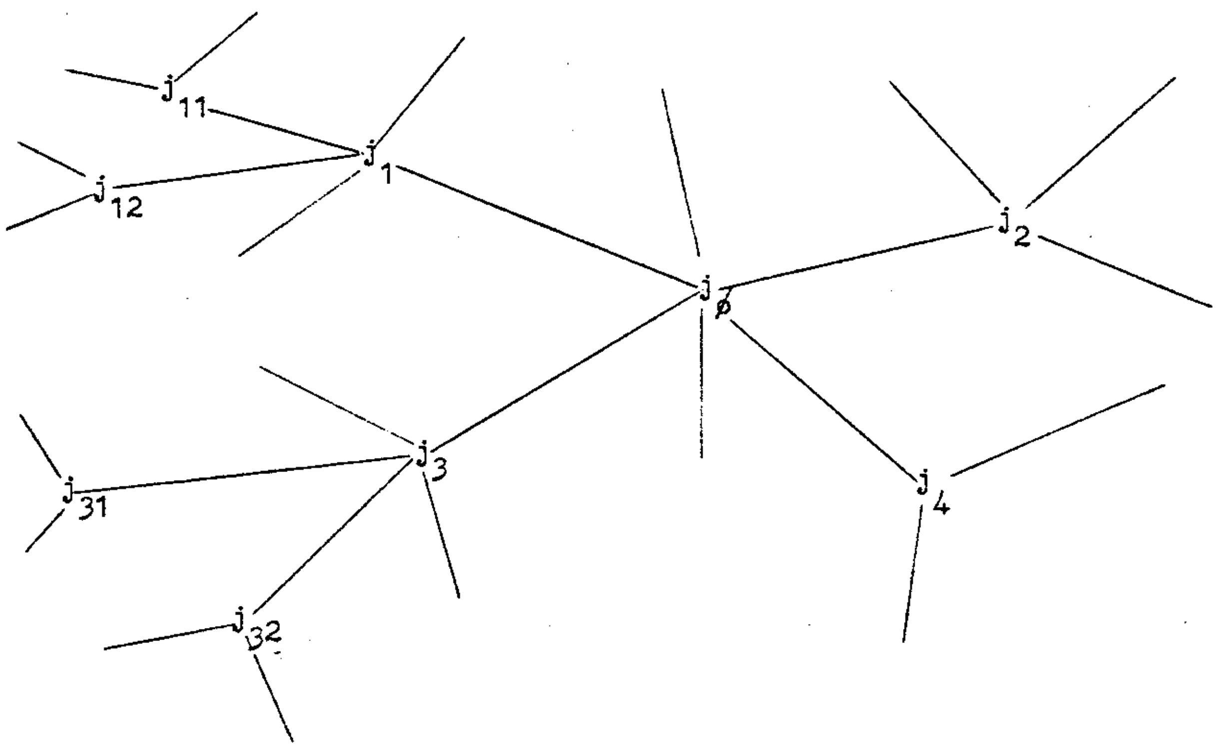
This is true for exchangeable sequences, and in the general setting when T is a compact group. It is not true for stationary sequences, or for the examples in Sections 13 and 14.

13. The infinitary tree

In this section we analyse a particular example of partial exchangeability. Though this example is somewhat artificial, we shall see some connections with naturally-arising problems.

A tree is a connected undirected graph without loops. Two vertices are neighbors if they are joined by an edge. For any two distinct vertices i, j there is a unique path of distinct vertices $i = i_1, i_2, \ldots, i_n = j$ such that i_r and i_{r+1} are neighbors, $1 \leq r < n$. A simple example is the infinite binary tree, in which there are a countable infinite number of vertices, and each vertex has exactly 3 neighbors. Similarly there is the infinite k-ary tree, where each vertex has exactly $k+1$ neighbors. We shall consider the infinite ∞-ary tree (the infinitary tree) where each vertex has infinitely many neighbors: for our purposes this is simpler than the finite case (see Remark 13.11). Let T denote the set of vertices of this tree. Let D be the set of finite sequences $d = (d_1, d_2, \ldots, d_n)$ of positive integers, and include in D the empty sequence $\emptyset$. Pick a vertex of T arbitrarily, and label it $j_\emptyset$. Then put the neighbors of $j_\emptyset$ in arbitrary order and label them $j_1, j_2, j_3, \ldots$. Then for each n put the neighbors of j_n (other than $j_\emptyset$) in arbitrary order and label them $j_{n1}, j_{n2}, j_{n3}, \ldots$. Continuing in the obvious manner, we obtain a labelling scheme $\{j_d : d \in D\}$ for all T.

An automorphism of T is a bijection $\gamma: T \to T$ such that $\gamma(i)$ and $\gamma(j)$ are neighbors iff i and j are neighbors. The easiest way to think of automorphisms is via labelling schemes. Given two labelling schemes $\{j_d : d \in D\}$, $\{\hat{j}_d : d \in D\}$, the map $\gamma(j_d) = \hat{j}_d$ is an automorphism; conversely, every automorphism is of this form.



A tree-process X is a family $(X_i : i \in T)$ of random variables indexed by T and taking values in some space S. Given a tree-process and a labelling scheme, we can define a process indexed by D:

(13.1) $$X_d^* = X_{j_d} .$$

A tree-process X is <u>invariant</u> if

(13.2) $$X \stackrel{\mathcal{D}}{=} (X_{\gamma(i)}: i \in T), \quad \text{each automorphism } \gamma.$$

Equivalently, X is invariant iff the distribution of $X^* = (X_d^*: d \in D)$ defined at (13.1) does not depend on the particular labelling scheme $(j_d: d \in D)$ used. Informally, this says the distribution of any $(X_{i_1}, \ldots, X_{i_n})$ depends only on the graph structure on $(i_1, \ldots, i_n)$. The set Γ of automorphisms is uncountable, but it is easy to see there exists a countable subset Γ_0 such that a process is Γ-invariant iff it is Γ_0-invariant. Thus invariant tree-processes do indeed fit into the "partial exchangeability" set-up of Section 12.

Our purpose is to describe explicitly the general invariant tree-process (i.e. to solve Problem 12.18 for this particular instance of partial exchangeability).

First we describe the well-known special case of Markov tree-processes. Informally, a tree-process Y is Markov if, conditional on the value Y_i at a vertex i, the values of the process along different branches from i are independent. To make this precise, for vertex i let N_i denote the set of vertices neighboring i. For $j \in N_i$ let the <u>branch</u> $B_{i,j}$ be the set of vertices k for which the path $i = i_1, i_2, \ldots, i_n = k$ of distinct neighboring vertices has $i_2 = j$. Let $Y(B_{i,j})$ be the array $(Y_k: k \in B_{i,j})$. Then Y is <u>Markov</u> if, for each i,

(13.3) $(Y(B_{i,j}): j \in N_i)$ are conditionally independent given Y_i.

<u>Warning</u>. There are two definitions of "Markov" for processes on trees. In the language of Spitzer (1975) we mean "Markov chain", not "Markov random

field"--see Remark 13.13.

If Y is an invariant tree-process then the distribution

(13.4) $\theta = L(Y_i, Y_j)$; i, j neighbors

is a symmetric distribution on $S \times S$ which does not depend on the particular pair i,j of neighbors. It is easy to verify (c.f. Spitzer (1975), Theorem 2) that for any symmetric θ there exists a unique (in distribution) invariant Markov tree-process satisfying (13.4). Call this the _invariant Markov tree-process associated with_ θ. Given such a process Y and a map $f: S \to S'$, the process $(f(Y_i): i \in T)$ is a tree-process with range space S'. This process inherits the invariance property from Y, but in general will not inherit the Markov property. The next result says that every invariant process can be obtained this way.

(13.5) _Theorem_. _Let_ X _be an invariant tree-process with range space_ S. _Then there exists an invariant Markov tree-process_ Y _with range space_ $[0,1]$ _and a function_ $f: [0,1] \to S$ _such that_ $X \overset{\mathcal{D}}{=} (f(Y_i): i \in T)$.

Proof. The idea of the proof is simple. The family of sub-processes on different branches from i is exchangeable; let Y_i be the directing random measure; then it turns out that (X_i, Y_i) is Markov.

For the formalities, we need some notation. For $n \geq 1$ and $d = (d_1, \ldots, d_m) \in D$ let $nd = (n, d_1, \ldots, d_m)$, and let $\emptyset d = d$. Fix a vertex i, and set up a labelling scheme $(j_d: d \in D)$ with $j_\emptyset = i$. For $n \geq 1$, let $C_n = (X_{nd}: d \in D)$ be the values of the process X on the branch B_{i,j_n}. So C_n takes values in S^D.

(13.6) Lemma. (a) $(C_1, C_2, C_3, \ldots)$ is exchangeable over X_i.

(b) The distribution of $(X_i; C_1, C_2, C_3, \ldots)$ does not depend on i or on the labelling scheme.

Proof. Assertion (b) is immediate from invariance. To prove (a), let π be a finite permutation of $\{1,2,3,\ldots\}$, and let $\gamma_\pi(i) = i$, $\gamma_\pi(j_{nd}) = j_{\pi(n)d}$. Then γ_π is an automorphism, and the invariance of X under γ_π implies $(X_i; C_1, C_2, \ldots) \overset{\mathcal{D}}{=} (X_i; C_{\pi(1)}, C_{\pi(2)}, \ldots)$.

Let Y_i be the directing random measure for (C_n). So Y_i takes values in $P(S^D)$. Keep i fixed. It is clear from (13.6)(b) that the distribution of Y_i does not depend on the labelling scheme used to define (C_n); what is crucial is that the actual random variable Y_i does not depend (a.s.) on the labelling scheme. To prove this, let Y_i, $\hat{Y}_i$ be derived from schemes (j_d), $(\hat{j}_d)$ with $j_\emptyset = \hat{j}_\emptyset = i$. Now the directing random measure for an infinite permutation $(Z_{\pi(i)})$ of an exchangeable sequence Z is a.s. equal to the directing measure for Z; so by re-ordering the neighbors of i we may suppose $j_n = \hat{j}_n$, $n \geq 1$. Now let A be a subset of $\{1,2,3,\ldots\}$, and let j^A be the labelling scheme $j^A_\emptyset = i$, $j^A_{nd} = j_{nd}$, $n \in A$, $j^A_{nd} = \hat{j}_{nd}$, $n \notin A$. By (13.6)(b) the distribution of $(X_{j^A_d} : d \in D)$ does not depend on A, and then Lemma 9.7 implies $Y_i = \hat{Y}_i$ a.s.

Now set $Y^*_i = (X_i, Y_i)$. We shall prove that the process $Y^* = (Y_i : i \in T)$ is invariant. Let γ be an automorphism. The invariance of X under γ implies

$$(X_i,\ C^i_n : n \geq 1,\ i \in T) \overset{\mathcal{D}}{=} (X_{\gamma(i)},\ \hat{C}^{\gamma(i)}_n : n \geq 1,\ i \in T)$$

where for each i the sequence $(\hat{C}^i_n : n \geq 1)$ describes the values of X on the branches $B_{i,j}$, $j \in N_i$, using some labelling scheme perhaps different from that used to define (C_n^i). Now the fact that Y_i does not depend on the labelling scheme shows

$$(X_i, Y_i : i \in T) \stackrel{\mathcal{D}}{=} (X_{\gamma(i)}, Y_{\gamma(i)} : i \in T) .$$

This gives invariance of Y^*.

Since X_i is a function of Y_i^*, and since the range space of Y_i^* is Borel-isomorphic to some subset of $[0,1]$, the proof of Theorem 13.5 will be complete when we show Y^* is Markov. Write $C_n^i = (X_{j_{nd}} : d \in D)$, for some labelling scheme (j_d) with $j_\emptyset = i$. So $\sigma(C_n^i) = \sigma(X_v : v \in B_{i,j_n}) = F_j^i$, say, where $j = j_n$.

(13.7) Lemma. Let i,j be neighbors, and let $k \in B_{i,j}$. Then $Y_k \in F_j^i$ a.s.

Proof. Let $i = i_1, i_2, \ldots, i_{r+1} = k$ be the path of distinct vertices linking i with k, so $i_2 = j$. Clearly for any neighbor v of k, $v \neq i_r$, we have

$$B_{k,v} \subset B_{i,j} . \tag{13.8}$$

Consider a labelling scheme $(\hat{j}_d)$ with $\hat{j}_\emptyset = k$, $\hat{j}_1 = i_r$. Now Y_k is the canonical random measure for the exchangeable family $(C_n^k : n \geq 1)$ defined using $(\hat{j}_d)$. So Y_k is some function of $(C_n^k : n \geq 2)$. But by (13.8) this array is contained in the array $(X_q : q \in B_{i,j})$, which generates the σ-field F_j^i.

To prove Y^* is Markov, fix i. In the notation of (13.3), we must prove $(X(B_{i,j}), Y(B_{i,j}))$, $j \in N_i$ are conditionally independent given (X_i, Y_i). Using Lemma 13.7, this reduces to proving

(13.9) $(F_j^i : j \in N_i)$ are conditionally independent given (X_i, Y_i) .

But Lemma 13.6(a) and Proposition 3.8 show

$$\{X_i, C_1^i, C_2^i, \ldots\} \text{ are conditionally independent given } Y_i,$$

and this implies (13.9).

The remainder of this section is devoted to various remarks suggested by Theorem 13.5.

Theorem 13.5 answers the "characterization problem" (12.18). In view of the discussion in Section 12 it is natural to ask which are the ergodic processes. It is easy to see that an invariant Markov tree-process X is ergodic iff there do not exist sets A_1, A_2 such that for neighbors i, j

$$\{X_i \in A_1\} = \{X_j \in A_2\} \text{ a.s.}, \quad 0 < P(X_i \in A_1) < 1.$$

And a general invariant tree-process X is ergodic iff it has a representation $X_i = f(Y_i)$ for some ergodic Markov Y.

Problem 12.20 and the subsequent discussion suggest

(13.10) Problem. Does there exist an invariant Markov tree-process Y which is "universal," in that for any invariant X there exists f such that $X \overset{\mathcal{D}}{=} (f(Y_i))$?

In connection with this problem, we remark that there exists (c.f. 12.22) a symmetric distribution θ on $[0,1]^2$ which is universal, in that for $\theta = L(V_1, V_2)$ the distributions $L(f(V_1), f(V_2))$ as f varies give all symmetric distributions. But this is not sufficient to answer Problem 13.10.

(13.11) The finite case. The conclusion of Theorem 13.5 is false for T_k, the infinite k-ary tree. For it is easy to construct an invariant process X on T_k such that $\{X_i, X_j : j \in N_i\}$ is distributed as a random permutation of $\{1,\dots,k+2\}$. Then $\{X_j: j \in N_i\}$ cannot be extended to an infinite exchangeable sequence, so the conclusion of Theorem 13.5 fails.

One way to construct invariant processes on T_k is to take $(\xi_i: i \in T_k)$ i.i.d. $U(0,1)$, and take $g: [0,1]^{k+2} \to S$ such that $g(x,y_1,\dots,y_{k+1})$ is symmetric in (y_i) for each x. Then the process

$$(13.12) \qquad X_i = g(\xi_i,\xi_{j_1},\dots,\xi_{j_{k+1}}); \quad j_r \in N_i$$

is invariant. More generally, we can take $X_i = g(\xi_{j_d}: d \in D)$ for a labelling scheme (j_d) with $j_\emptyset = i$ and for g satisfying the natural symmetry conditions. Presumably, as in the case (12.23) of ordinary stationary sequences, this construction gives all "finitely determined" invariant processes--but this looks hard to prove.

(13.13) <u>Markov random fields</u>. A tree-process X is a <u>Markov random field</u> if, for each vertex i,

$$\sigma(X_j: j \neq i) \text{ and } X_i \text{ are conditionally independent given } \sigma(X_j: j \in N_i).$$

This is weaker than our definition (13.3) of "Markov": see Kindermann and Snell (1980). It is natural to ask

(13.14) <u>Problem</u>. Is there a simple explicit description of the invariant Markov random fields on the infinitary tree T?

In the notation of Theorem 13.5, the Markov random field property for X is equivalent to some messy conditional independence properties for the triple $(Y_i, Y_j, f(Y_j))$, i,j neighbors, but it seems hard to say when these properties hold.

Spitzer (1975) (see also Zachary (1983)) discusses {0,1}-valued Markov random fields on the tree T_k. For such a field, let $V_i = \sum_{j \in N_i} X_j$. Then

(a) there is an explicit description of the possible conditional distribution of X_i given V_i in an invariant Markov random field;

(b) any such conditional distribution is attained by some Markov tree-process.

For the tree T we can set $V_i = \lim_n n^{-1} \sum_1^n X_{j_r}$, where $N_i = \{j_1, j_2, \ldots\}$, and ask the same questions. It can be shown that (b) is not true on T; the analog of (a) is unclear.

(13.15) Stationary reversible processes. A stationary process $X = (X_n : n \in \mathbb{Z})$ is called reversible if $X \overset{\mathcal{D}}{=} (X_{-n} : n \in Z)$. So a stationary Markov process Y is reversible iff the distribution $L(Y_0, Y_1)$ is symmetric. Reversible Markov processes occur often in applied probability models--see e.g. Kelly (1979)--and since we are frequently interested in functions of the underlying process, the following theoretical definition and problem are suggested.

(13.16) Definition. Given a space S, let H_0 be the class of S-valued processes of the form $(f(Y_n) : n \in \mathbb{Z})$, where Y is a stationary reversible Markov process on some space S' (= [0,1], without loss of generality) and $f: S' \to S$ is some function.

(13.17) Problem. Give intrinsic conditions for a stationary reversible process X to belong to H_0.

We shall see later (13.24) that not every stationary reversible sequence is in H_0.

To see the connection between Problem 13.17 and tree-processes, consider

(13.18) Definitions. A line in the infinitary tree T is a sequence $(i_n : n \in \mathbb{Z})$ of distinct vertices with i_n, i_{n+1} neighbors for each n.

Given a space S, let H_1 be the class of processes $(X_{i_n} : n \in \mathbb{Z})$, where X is some S-valued invariant tree-process, and (i_n) is a line in T.

Now Theorem 13.5 has the immediate

(13.19) Corollary. $H_1 = H_0$.

Though this hardly solves Problem 13.17, it does yield some information. For instance, H_1 is easily seen to be closed under weak convergence, so

(13.20) Corollary. H_0 is closed under weak convergence.

This fact is not apparent from the definition of H_0.

(13.21) Correlation structure. On the tree T there is the natural distance d(i,j), the number of edges on the minimal path from i to j. For an invariant tree-process X, where X_i is real-valued and square-integrable, the correlations must be of the form $\rho(X_i, X_j) = \rho_{d(i,j)}$ for some correlation function $(\rho_n : n \geq 0)$.

(13.22) Proposition. A sequence $(\rho_n : n \geq 0)$ is the correlation function of some invariant tree-process iff $\rho_n = \int x^n \lambda(dx)$ for some probability measure λ on $[-1,1]$.

Remark. This result can be deduced from harmonic analysis results for the trees T_k. Cartier (1973) shows that correlation functions on T_k are mixtures of functions of the form

$$\rho(n) = \frac{k(\lambda^{n+1} - \lambda^{-n-1}) - (\lambda^{n-1} - \lambda^{-n+1})}{k^{n/2}(k+1)(\lambda - \lambda^{-1})}$$

where λ is either real with $|\lambda| \leq k^{1/2}$, or complex with $|\lambda| = 1$. And ρ is a correlation function on T iff it is a correlation function on each T_k. See also Arnaud (1980).

Instead, we shall see how to deduce Proposition 13.22 from Theorem 13.5.

<u>Proof</u>. Let C denote the set of sequences of the form $c_n = \int x^n \lambda(dx)$, $\lambda(\cdot)$ some probability distribution on $[-1,1]$.

Let Y be the Markov tree-process associated with θ, where

$$\theta(0,0) = \theta(1,1) = \tfrac{1}{4}(1+\lambda); \quad \theta(0,1) = \theta(1,0) = \tfrac{1}{4}(1-\lambda)$$

for some $-1 \le \lambda \le 1$. Then Y has correlation function $\rho(n) = \lambda^n$. By taking mixtures over λ, we see that every sequence in C is indeed the correlation function of some invariant process.

By Theorem 13.5 the converse reduces to determining the correlation function for sequences $Y_n = f(X_n)$, where $X = (X_n: n \ge 0)$ is stationary reversible Markov. Consider first the case where X has a <u>finite</u> state space S. Let P be the matrix of transition probabilities, and let π be the vector $(\pi(s))$ of the stationary distribution. Then $\rho_n = c_n/c_0$, where

$$c_n = \sum_s \sum_t f(s)\pi(s)P^n(s,t)f(t) \ . \tag{13.23}$$

Reversibility implies $\pi(s)P(s,t) = \pi(t)P(t,s)$, and so the matrix $Q(s,t) = \pi^{1/2}(s)P(s,t)\pi^{-1/2}(t)$ is symmetric. The spectral theorem for symmetric matrices says we can write $Q = U\Lambda U^T$, where U is orthogonal and Λ is diagonal with real eigenvalues (λ_i); since these are also the eigenvalues of P we have $|\lambda_i| \le 1$. Substituting into (13.23), putting $v(s) = f(s)\pi^{1/2}(s)$,

$$c_n = vU\Lambda^n U^T v^T = \sum a_i^2 \lambda_i^n \ , \quad \text{say.}$$

So ρ_n is indeed in C.

The general case, where X has state space $[0,1]$, can presumably be obtained by appealing to a more sophisticated spectral theorem. Let us instead give a probabilistic argument. For $N \geq 1$ let $\phi_N(x) = i/2^N$ on $i/2^N \leq x < (i+1)/2^N$. Let $(X_0^N, X_1^N, X_2^N, \ldots)$ be the stationary reversible Markov chain for which $(X_0^N, X_1^N) \stackrel{\mathcal{D}}{=} (\phi_N(X_0), \phi_N(X_1))$. Obviously $(X_0^N, X_1^N) \xrightarrow{\mathcal{D}} (X_0, X_1)$, and it can be shown (we leave this to the reader as a hard exercise) that

$$(13.24) \quad (X_0^N, X_1^N, \ldots, X_k^N) \xrightarrow{\mathcal{D}} (X_0, X_1, \ldots, X_k) \quad \text{as} \quad N \to \infty, \quad k \text{ fixed.}$$

Observe also that by weak compactness of $\mathcal{P}[-1,1]$,

$$(13.25) \qquad \mathcal{C} \text{ is closed under pointwise convergence.}$$

Now consider $Y_n = f(X_n)$, where f is bounded continuous, and let (ρ_n) be the correlation function for Y. Using the result for finite state spaces, (13.24) and (13.25), we deduce that (ρ_n) is in $\mathcal{C}$. Measure theory extends this to general f.

(13.26) Remarks. Take X_0 and Y independent, X_0 uniform on $\{0,1,2,3\}$, Y uniform on $\{-1,1\}$, and define $X_n = X_0 + nY$ modulo 4. Then X is a stationary process which is reversible, and $\rho(X_0, X_2) = -1$, so by Proposition 13.22 X cannot be represented as a function of any stationary reversible Markov process.

We also remark that the analog of Problem 13.17 without reversibility (which stationary processes are functions of stationary Markov processes?) is uninteresting as it stands. For given a stationary process (X_n), let (Y_n) be the stationary Markov process given by $Y_n = (X_n, X_{n+1}, X_{n+2}, \ldots)$, and then $X_n = f(Y_n)$ for $f(x_0, x_1, x_2, \ldots) = x_0$. However, the problem of which finite state stationary processes are functions of finite state stationary Markov processes is non-trivial--see Heller (1965).

(13.27) A curious argument. Readers unwilling to consider the possibility of picking uniformly at random from a countable infinite set should skip this section. For the others, we present a curious argument for de Finetti's theorem and for Corollary 13.19 (that a line in an invariant tree-process yields a function of a stationary reversible Markov process).

Argument for de Finetti's theorem. Let $Z = (Z_n : n \in \mathbb{N})$ be exchangeable, so

(a) $Z \overset{\mathcal{D}}{=} (Z_{i_1}, Z_{i_2}, \ldots)$, any distinct $(i_1, i_2, \ldots)$.

Take $I_1, I_2, \ldots$ independent of each other and Z, uniform on $\mathbb{N}$. Then

(b) the values $I_1, I_2, \ldots$ are distinct;

(c) for any function z defined on $\mathbb{N}$, the sequence $(z(I_1), z(I_2), \ldots)$ is i.i.d.

Now consider the sequence Z_{I_n}. By conditioning on $(I_1, I_2, \ldots)$ and using (a) and (b), we see

(d) $Z \overset{\mathcal{D}}{=} (Z_{I_n})$.

For any sequence $z = (z_n)$ of constants, the distribution of (Z_{I_n}) given $Z = z$ is i.i.d., by (c). So

(e) (Z_{I_n}) is a mixture of i.i.d. sequences.

And (d) and (e) give de Finetti's theorem.

Argument for Corollary 13.19. Let X be an invariant tree-process, let $(i_n : n \geq 0)$ be a line in T, so

(a) $(X_{i_n}) \stackrel{\mathcal{D}}{=} (X_{j_n})$ for any line (j_n).

Take I_0 uniform on T, I_1 uniform on the neighbors of I_0, I_2 uniform on the neighbors of I_1, and so on. Then

(b) the values $I_0, I_1, I_2, \ldots$ form a line in T;

(c) (I_n) is a stationary reversible Markov process on T.

Now consider the sequence (X_{I_n}). By conditioning on $(I_1, I_2, \ldots)$ and using (a) and (b),

(d) $(X_{i_n}) \stackrel{\mathcal{D}}{=} (X_{I_n})$.

Recall definition (13.16) of H_0. For any function x on T, the sequence (x_{I_n}) is in H_0, by (c). So by conditioning on X,

(e) (X_{I_n}) is a mixture of processes in H_0.

But it is not hard to see H_0 is closed under mixtures, so (d) and (e) give Corollary 13.19.

(13.28) Problem. Is it possible to formalize the arguments above (e.g. by using finite additivity)?

14. Partial exchangeability for arrays: the basic structure results

Several closely related concepts of partial exchangeability for arrays of random variables have been studied recently. In this section we describe the analog of de Finetti's theorem for arrays.

Consider an array $X = (X_{i,j}: i,j \geq 1)$ of S-valued random variables such that

(14.1) $$X \overset{\mathcal{D}}{=} (X_{\pi_1(i),\pi_2(j)})$$

for all finite permutations π_1, π_2 of $\mathbb{N}$. In terms of the general description (12.16) of partial exchangeability, $I = \mathbb{N}^2$ and Γ is the product of the finite permutation groups. Writing $R_i = (X_{i,j}: j \geq 1)$ for the i^{th} row of the array X, and $C_j = (X_{i,j}: i \geq 1)$ for the j^{th} column, condition (14.1) is equivalent to the conditions

(14.2)(a) $(R_1, R_2, R_3, \ldots)$ is exchangeable;

(b) $(C_1, C_2, C_3, \ldots)$ is exchangeable.

If these conditions hold, call X a row-and-column-exchangeable (RCE) array. Here are three obvious examples of such arrays.

(14.3)(a) $(X_{i,j})$ i.i.d.

(b) $X_{i,j} = \xi_i$, where $(\xi_1, \xi_2, \ldots)$ are i.i.d. Here the entries within a column are i.i.d., but different columns are identical; or equivalently, entries within a row are identical, but different rows are i.i.d.

(c) $X_{i,j} = \eta_j$, where $(\eta_1, \eta_2, \ldots)$ are i.i.d. Similar to (b), interchanging rows with columns.

Here is a more interesting class of examples. Let $\phi: [0,1]^2 \to [0,1]$ be an arbitrary measurable function. Let $(\xi_1, \xi_2, \ldots; \eta_1, \eta_2, \ldots)$ be i.i.d. uniform on $[0,1]$. Then we can define a $\{0,1\}$-valued array X by

(14.4) conditional on $(\xi_1, \xi_2, \ldots; \eta_1, \eta_2, \ldots)$, $P(X_{i,j} = 1) = \phi(\xi_i, \eta_j)$, $P(X_{i,j} = 0) = 1 - \phi(\xi_i, \eta_j)$, and the $X_{i,j}$ are independent.

Such processes, called <u>ϕ-processes</u>, can be simulated on a computer and the realizations drawn as a pattern of black and white squares. Diaconis and Freedman (1981a) present such simulations and use them to discuss hypotheses about human visual perception.

What is the analog of de Finetti's theorem for RCE arrays? There are several easy but rather superficial results which we describe below; the harder and deeper result is Theorem 14.11, and the reader may well skip to the statement of that theorem.

One method of analysing RCE arrays is to consider the rows $(R_1,R_2,\ldots)$ as an exchangeable sequence such that $(R_1,R_2,\ldots) \stackrel{\mathcal{D}}{=} (h_\pi(R_1),h_\pi(R_2),\ldots)$ for each π, where $h_\pi((x_i)) = (x_{\pi(i)})$. Then Lemma 9.6(i) says that the possible directing random measures α for (R_i) from a RCE array are precisely those with certain invariance properties--see Lynch (1982a). But this approach does not lead to any explicit construction of the general RCE array.

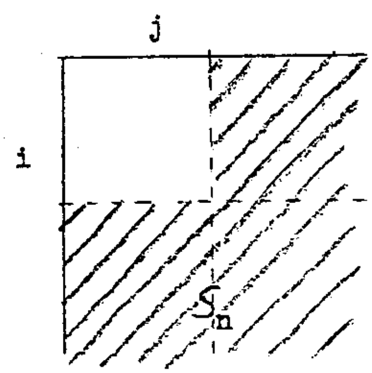
Recall that one version of de Finetti's theorem is: conditional on the tail σ-field of an exchangeable (Z_i), the variables $(Z_1,Z_2,\ldots)$ are i.i.d. So it is natural to study what happens to RCE arrays when we condition on some suitable "remote" σ-field. Define the <u>tail</u> σ-field $\mathcal{T}$ and the <u>shell</u> σ-field $\mathcal{S}$ as follows.

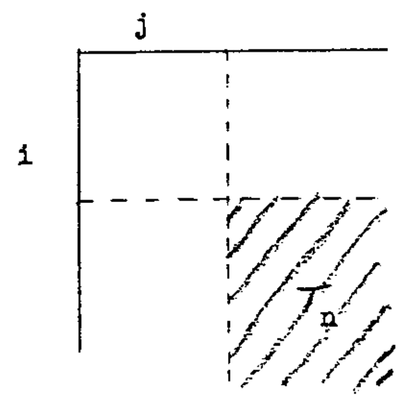
$\mathcal{T} = \bigcap_n \mathcal{T}_n$, where

$\mathcal{T}_n = \sigma\big(X_{i,j} : \min(i,j) > n\big)$.

$\mathcal{S} = \bigcap_n \mathcal{S}_n$, where

$\mathcal{S}_n = \sigma\big(X_{i,j} : \max(i,j) > n\big)$.





We need another definition. Call an array X dissociated if

(14.5) $(X_{i,j}: \max(i,j) \leq n)$ independent of $(X_{i,j}: \min(i,j) > n)$ for each n.

The next results describe what happens when we condition on these σ-fields.

(14.6) Proposition. *Let X be a RCE array. Conditional on $\mathcal{T}$, the array X is RCE and dissociated.*

(14.7) Proposition. *Let X be a RCE array. Conditional on $\mathcal{S}$, the variables $X_{i,j}$ are independent (but in general not identically distributed).*

Proof of Proposition 14.6. Let π_1, π_2 be permutations, and take n so large that π_1 and π_2 do not alter $n+1, n+2, \ldots$. Since $X \overset{\mathcal{D}}{=} (X_{\pi_1(i),\pi_2(j)})$, we have

conditional on $\mathcal{T}_n$, the distributions of $(X_{i,j}: i,j \leq n)$ and $(X_{\pi_1(i),\pi_2(j)}: i,j \leq n)$ are identical.

Letting $n \to \infty$, we see that conditional on $\mathcal{T}$, X is RCE.

Now fix M, and consider the diagonal squares $S_k = (X_{i,j}: (k-1)M+1 \leq i,j \leq kM)$. Given k and a permutation π of $\{1,\ldots,k\}$, there exists a permutation ρ of $\{1,\ldots,kM\}$ such that, setting $\hat{X}_{i,j} = X_{\rho(i),\rho(j)}$, we have $(\hat{S}_1,\ldots,\hat{S}_k) = (S_{\pi(1)},\ldots,S_{\pi(k)})$. It follows that $(S_1,\ldots,S_k)$ is exchangeable over $\mathcal{T}_n$ $(n > kM)$. Letting $n \to \infty$,

$$(S_1, S_2, \ldots) \text{ is exchangeable over } \mathcal{T}.$$

Moreover the tail of (S_i) is contained in $\mathcal{T}$. Using Proposition 3.8 we see that (S_k) is conditionally i.i.d. given $\mathcal{T}$, so in particular

$(X_{i,j}: i,j \le M)$ and $(X_{i,j}: M < i,j \le 2M)$ are conditionally independent given $\mathcal{T}$.

This holds for all M; it is easy to deduce X is conditionally dissociated.

Proof of Proposition 14.7. Let $Y_1, Y_2, \ldots$ be an arbitrary enumeration of $(X_{i,j}: i,j \ge 1)$. It is not hard to verify (see Aldous (1982b)) that (Y_i) satisfies the hypotheses of Proposition 6.4. Hence $Y_1, Y_2, \ldots$ are conditionally independent given their tail σ-field $\mathcal{T}_Y$. But $\mathcal{T}_Y = \mathcal{S}$.

Finally, in example (14.3b) we have $\mathcal{S} = \sigma(\xi_1, \xi_2, \ldots) = \sigma(X)$, so that conditional on $\mathcal{S}$ the array X is deterministic, and so the entries $X_{i,j}$ are not conditionally identically distributed (except when ξ is degenerate).

Although Propositions 14.6 and 14.7 give useful information, they do not provide a complete description of RCE arrays. Another approach is to use the general theory outlined in Section 12. The general theory says that each RCE array is a mixture of ergodic RCE arrays; the next result identifies the ergodic arrays as the dissociated arrays.

(14.8) Proposition. *For a RCE array X the following are equivalent:*

(a) X *is extreme* (= *ergodic*) *in the class of RCE arrays.*

(b) $\mathcal{T}$ *is trivial.*

(c) X *is dissociated.*

Proof. (a)⇒(b). If X is ergodic then the ergodic σ-field $\mathcal{E}$ is trivial. But $\mathcal{E} \supset \mathcal{T}$, just as in the 1-parameter case.

(b)⇒(c). Proposition 14.6.

(c)⇒(a). Suppose $\mathcal{L}(X) = \frac{1}{2}\mathcal{L}(Y) + \frac{1}{2}\mathcal{L}(Z)$ for RCE arrays Y, Z. We must show $\mathcal{L}(X) = \mathcal{L}(Y)$. Fix M, and consider the diagonal squares

$S_k^X = (X_{i,j}: (k-1)M+1 \le i,j \le kM)$. Then $L(S_1^X, S_2^X, \ldots) = \frac{1}{2}L(S_1^Y, S_2^Y, \ldots) + \frac{1}{2}L(S_1^Z, S_2^Z, \ldots)$. Each of these sequences is exchangeable, and (S_k^X) is i.i.d. by hypothesis; but de Finetti's theorem implies that an i.i.d. sequence is extreme in the class of exchangeable sequences, and hence we must have $S_1^X = S_1^Y$. But this says $(X_{i,j}: i,j \le M) \overset{\mathcal{D}}{=} (Y_{i,j}: i,j \le M)$, and since M is arbitrary $X \overset{\mathcal{D}}{=} Y$.

However, the net effect of Proposition 14.8 and the general theory is to show that each RCE array is a mixture of dissociated RCE arrays--and this was already given by Proposition 14.6.

The results so far have been fairly direct consequences of the circle of ideas around de Finetti's theorem. We now come to the fundamental "characterization theorem," which seems somewhat deeper.

(14.9) <u>Convention</u>. Let $\{\alpha;\ \xi_i,\ i \ge 1;\ \eta_j,\ j \ge 1;\ \lambda_{i,j},\ i,j \ge 1\}$ denote independent $U(0,1)$ random variables.

Now given any $f: (0,1)^4 \to S$ we can define

$$X^*_{i,j} = f(\alpha, \xi_i, \eta_j, \lambda_{i,j}) \tag{14.10}$$

and this yields an RCE array X^*. Say f <u>represents</u> X^*. This class of arrays is the class obtained from the examples in (14.3), together with the trivial array $X_{i,j} = \alpha$, by the general methods of Section 12. Note that a ϕ-process is represented by $f(a,b,c,d) = 1_{(d \le \phi(b,c))}$.

(14.11) <u>Theorem</u>. <u>Let</u> X <u>be a RCE array</u>. <u>Then there exists</u> $f: [0,1]^4 \to S$ <u>such that</u> $X \overset{\mathcal{D}}{=} X^*$, <u>where</u> X^* <u>is represented by</u> f.

This result was obtained independently by Aldous (1979, 1981a) and Hoover (1979, 1982); the more general results of Hoover will be described later.

The proof which appears in Aldous (1981a) is a formalized version of an argument due to Kingman (personal communication), which we present essentially verbatim below.

By a <u>coding</u> ξ of a random variable Y we mean a presentation $Y \overset{\mathcal{D}}{=} g(\xi)$ for ξ with $U(0,1)$ distribution.

<u>Kingman's proof of Theorem 14.11</u>. There is no loss of generality in supposing that X is a quadrant of an array $(X_{i,j}: i,j \in \mathbf{Z})$. Write

$$A = (X_{i,j}: i,j \le 0)$$
$$B_i = (X_{i,j}: j \le 0) \qquad B = (B_i: i \ge 1)$$
$$C_j = (X_{i,j}: i \le 0) \qquad C = (C_j: j \ge 1) \,.$$

de Finetti's theorem implies that for an exchangeable sequence $(Y_i: i \in \mathbf{Z})$ the variables $(Y_i: i \ge 1)$ are conditionally i.i.d. given $(Y_i: i \le 0)$. Applying this to the three sequences

$$Y_i = (X_{i,j}: j \in \mathbf{Z}) = (B_i: X_{i,j},\ j \ge 1)$$
$$Y_i = (B_i, X_{i,j}) \quad (j \text{ fixed})$$
and
$$Y_i = B_i \,,$$

we see that

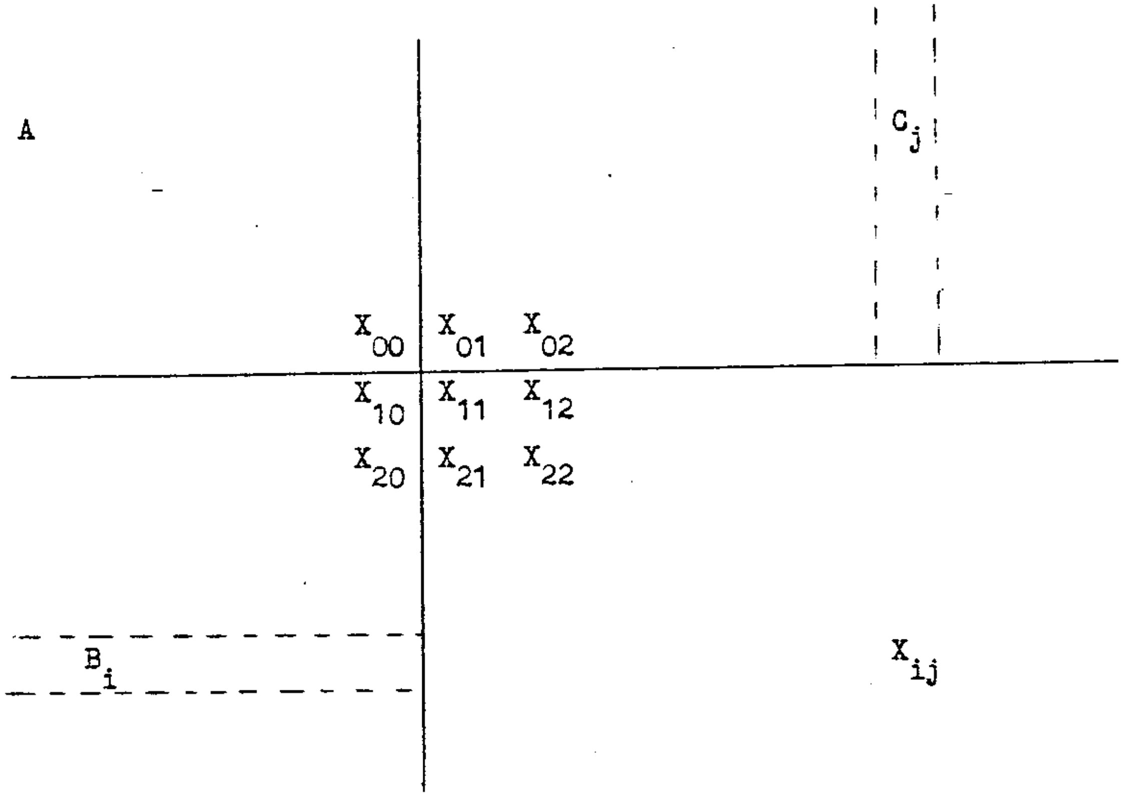
(i) the variables $(B_i, X_{i,1}, X_{i,2}, \ldots)$ for $i \ge 1$ are conditionally i.i.d. given (A,C);

(ii) for fixed $j \ge 1$, the variables $(B_i, X_{i,j})$ for $i \ge 1$ are conditionally i.i.d. given (A, C_j);

(iii) the variables B_i for $i \ge 1$ are conditionally i.i.d. given A.

Moreover, the conditional distribution of B_i given (A,C) is expressible as a function of A, so

(iv) B_i is conditionally independent of C given A, for each i.



Corresponding results hold when the roles of i and j are reversed, and standard conditional probability manipulations then show that

(v) $B_1, B_2, \ldots, C_1, C_2, \ldots$ are conditionally independent given A, the conditional distributions of B_i and C_j not varying with i and j respectively;

(vi) the variables $(X_{i,j}: i,j \geq 1)$ are conditionally independent given (A,B,C), the conditional distribution of $X_{i,j}$ depending only on (A,B_i,C_j).

Now take a coding α of A, and condition everything on α. Choose a coding ξ_1 for B_1, and let ξ_i be the corresponding coding for B_i; similarly choose codings η_j for C_j (all this conditional on α). Then $\alpha, \xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots$ are independent $U(0,1)$, and there is a function $g:(0,1)^3 \to P(S)$ such that $g(\alpha, \xi_i, \xi_j)$ is the conditional distribution of

$X_{i,j}$ given (A,B,C). The usual construction then yields a function $f: (0,1)^4 \to S$ such that the array X has the same distribution as the array $f(\alpha,\xi_i,\eta_j,\lambda_{i,j})$, where $(\lambda_{i,j})$ are more independent $U(0,1)$ variables.

Remarks. Representation (14.10) has a natural statistical interpretation: $X^*_{i,j}$ is determined by a "row effect" ξ_i, a "column effect" η_j, an "individual random effect" λ_{ij} and an "overall effect" α. Theorem 14.11 may be regarded as a natural extension of de Finetti's theorem if we formulate the latter using (2.5) as: (Z_i) is exchangeable iff $(Z_i) \overset{\mathcal{D}}{=} (f(\alpha,\xi_i), i \geq 1)$ for some $f: (0,1)^2 \to S$.

Theorem 14.11 builds the general RCE array from four basic components. The arrays which can be built from only some of these components are the arrays with certain extra properties, as the next few results show.

For $f: [0,1]^3 \to S$ define

$$X^*_{i,j} = f(\xi_i,\eta_j,\lambda_{i,j}) \,. \tag{14.12}$$

Then X^* is a dissociated RCE array. Conversely, we have

(14.13) Corollary. Let X be a dissociated RCE array. Then there exists a function $f: [0,1]^3 \to S$ such that $X = X^*$, for X^* defined at (14.12).

Proof. Theorem 14.11 says X can be represented by some $f: [0,1]^4 \to S$. For each $a \in [0,1]$, let $f_a(b,c,d) = f(a,b,c,d)$. By conditioning on α in the representation (14.10), we see that X is a mixture (over a) of arrays X^a, where $X^a_{i,j} = f_a(\xi_i,\eta_j,\lambda_{i,j})$. But X is dissociated, so by Proposition 14.8 it is extreme in the class of RCE arrays, so $X \overset{\mathcal{D}}{=} X^a$ for almost all a.

(14.14) Corollary. Let X be a dissociated {0,1}-valued RCE array. Then X is distributed as a ϕ-process, for some $\phi: [0,1]^2 \to [0,1]$.

Proof. Let f be as in Corollary 14.13, and set $\phi(x,y) = P(f(x,y,\lambda_{i,j}) = 1)$.

It is natural to ask which arrays are of the form $f(\xi_i,\eta_j)$ --note the general ϕ-process is not of this form. This result is somewhat deeper. Different proofs appear in Aldous (1981a) and Hoover (1979), and will not be repeated here. See also Lynch (1982b).

(14.15) Corollary. For a dissociated RCE array X, the following are equivalent:

(a) $X_{1,1} \in \mathcal{S}$ a.s.,

(b) $X \underset{\mathcal{D}}{=} X^*$, where $X^*_{i,j} = f(\xi_i,\eta_j)$ a.s. for some $f: (0,1)^2 \to S$.

An alternative characterization of such arrays, based upon entropy ideas, will be given in (15.28). We remark that although it is intuitively obvious that a non-trivial array of the form $f(\xi_i,\eta_j)$ cannot have i.i.d. entries, there seems no simple proof of this fact. But it is a consequence of Corollary 14.15, since for an i.i.d. array $\mathcal{S}$ is trivial.

The next result completes the list of characterizations of arrays representable by functions of fewer than four components.

(14.16) Corollary. For a dissociated RCE array X, the following are equivalent:

(a) $X \overset{\mathcal{D}}{=} (X_{i,\pi_i(j)}: i,j \geq 1)$ for all finite permutations $\pi_1,\pi_2,\dots$.

(b) $X \overset{\mathcal{D}}{=} X^*$, where $X^*_{i,j} = f(\xi_i,\lambda_{i,j})$ for some $f: (0,1)^2 \to S$.

Proof. Let α_i be the directing random measure for $(X_{i,j}: j \geq 1)$. Corollary 3.9 implies that for each (i,j), (14.17a) α_i is a r.c.d. for $X_{i,j}$ given $\sigma\{X_{i',j'}: (i',j') \neq (i,j)\}$.

Let N_i be disjoint infinite subsets of $\{1,2,\ldots\}$. Dissociation implies $\sigma(X_{i,j}: j \in N_i)$, $i \geq 1$, are independent, and since $\alpha_i \in \sigma(X_{i,j}: j \in N_i)$ we get

(14.17b) $(\alpha_i: i \geq 1)$ are independent.

Set $X^*_{i,j} = F^{-1}(\alpha_i, \lambda_{i,j})$, where $F^{-1}(\theta,\cdot)$ is the inverse distribution function of θ. Then (14.17a) and (14.17b) imply $X^* \overset{\mathcal{D}}{=} X$. Finally, code α_i as $g(\xi_i)$.

Another question suggested by Theorem 14.11 concerns uniqueness of the representing function. Suppose T_i $(1 \leq i \leq 4)$ are measure-preserving functions $[0,1] \to [0,1]$. Then f and $f^*(a,b,c,d) = f(T_1(a),T_2(b),T_3(c),T_4(d))$ represent arrays X and X^* which have the same distribution. It is natural to conjecture that if X and X^* have the same distribution then any representing functions f, f^* must be "equivalent" in the sense above. Hoover (1979) gives a precise statement and proof of this fact.

Finally, let us mention a different type of exchangeability property for arrays. This is motivated by the concept of U-statistics, that is to say sequences (U_n) of the form

(14.18) $$U_n = \frac{1}{\binom{n}{2}} \sum_{1<i<j<n} g(V_i,V_j)$$

where (V_i) is i.i.d. and $g(\cdot,\cdot)$ is symmetric. (Many natural statistical estimators are of this form--see Serfling (1980), Chapter 5.) Now we can regard (U_n) as the partial averages

$$U_n = \frac{1}{\binom{n}{2}} \sum_{\{i,j\} \subset \{1,\ldots,n\}} X_{\{i,j\}}$$

of an array $X = (X_{\{i,j\}})$ indexed by unordered pairs: $X_{\{i,j\}} = g(V_i, V_j)$. Since (V_i) is exchangeable, X has the property

(14.19) $$X \stackrel{\mathcal{D}}{=} (X_{\{\pi(i),\pi(j)\}}), \text{ each finite permutation } \pi.$$

This property has been called <u>weak exchangeability</u>. In the spirit of (14.10) we can construct a more general weakly exchangeable array by

(14.20) $$X^*_{\{i,j\}} = g(\alpha, \xi_i, \xi_j, \lambda_{\{i,j\}}) ,$$

where $g(a,\cdot,\cdot,d)$ is symmetric for each (a,d). It is possible to modify the proof of Theorem 14.11 to prove

(14.21) <u>Theorem</u>. <u>Let</u> X <u>be a weakly exchangeable array</u>. <u>Then</u> $X \stackrel{\mathcal{D}}{=} X^*$ <u>for some array</u> X^* <u>of the form</u> (14.20).

And all the other results for RCE arrays have natural analogs for weakly exchangeable arrays.

So far we have considered 2-dimensional arrays. The definitions of RCE and weak exchangeability have natural extensions for k-dimensional arrays, and it is not hard to guess what the analogs of Theorems 14.11 and 14.21 should be. <u>Proving</u> these along the lines of the proof of Theorem 14.11 seems hard. Hoover (1979) uses quite different techniques to establish a general result encompassing Theorems 14.11, 14.21 and their k-dimensional versions. The statement and proof of these results involve ideas from logic which we shall not attempt to present here--we refer the reader to the expository account in Hoover (1982).

15. Partial exchangeability for arrays: complements

Most of the results in Parts 1 and 2 about exchangeable sequences suggest conjectures for similar results for arrays. Rather than attempt any systematic program of extension, we shall present merely a selection of results and open problems which look interesting. Studying the open problems would perhaps make a good Ph.D. thesis.

(15.1) Correlation structure. For a RCE array X with real square-integrable entries, the correlation structure is determined by the three numbers

$$\rho = \rho(X_{1,1},X_{2,2}); \quad \rho_c = \rho(X_{1,1},X_{1,2}); \quad \rho_r = \rho(X_{1,1},X_{2,1}) .$$

(15.2) Proposition. *The possible correlations (ρ,ρ_c,ρ_r) for a RCE array are precisely those satisfying*

$$0 \le \rho \le \min(\rho_r,\rho_c); \quad \rho_r + \rho_c \le 1 + \rho .$$

Proof. Let $\hat{\alpha}$, $(\hat{\xi}_i)$, $(\hat{\eta}_j)$, $(\hat{\lambda}_{i,j})$ be independent $N(0,1)$ and define

$$X_{i,j} = a\hat{\alpha} + b\hat{\xi}_i + c\hat{\eta}_j + d\hat{\lambda}_{i,j} + e . \tag{15.3}$$

Then, setting $\sigma^2 = a^2+b^2+c^2+d^2$, we have

$$\rho = a^2/\sigma^2; \quad \rho_c = (a^2+c^2)/\sigma^2; \quad \rho_r = (a^2+b^2)/\sigma^2 .$$

And we can choose (a,b,c,d) to attain any correlations in the range specified.

Conversely, let X be a RCE array. Suppose first that X is dissociated. Then $\rho = 0$. Of course $\min(\rho_r,\rho_c) \ge 0$ by (1.8). We shall prove

$$\rho_r + \rho_c \le 1 . \tag{15.4}$$

We may suppose $EX_{1,1} = 0$, $EX_{1,1}^2 = 1$ and, by Corollary 14.13, that $X_{i,j} = f(\xi_i,\eta_j,\lambda_{i,j})$. Define

$$g(x,y) = Ef(x,y,\lambda_{1,1})$$
$$g_1(x) = Eg(x,\eta_1); \quad g_2(y) = Eg(\xi_1,y) .$$

Then
$$\begin{aligned}\rho_c &= EX_{1,1}X_{1,2} \\ &= Eg(\xi_1,\eta_1)g(\xi_1,\eta_2) \\ &= Eg_1^2(\xi_1)\end{aligned}$$

and similarly $\rho_r = Eg_2^2(\eta_1)$. So

$$\begin{aligned}1 &= EX_{1,1}^2 \geq Eg^2(\xi_1,\eta_1) \\ &= E(g(\xi_1,\eta_1) - g_1(\xi_1))^2 + Eg_1^2(\xi_1) \\ &\quad \text{by conditioning on } \xi_1 \\ &\geq Eg_2^2(\eta_1) + Eg_1^2(\xi_1) \quad \text{by conditioning on } \eta_1 \\ &= \rho_c + \rho_r, \quad \text{establishing (15.4).}\end{aligned}$$

For general X with $EX_{1,1} = 0$ and $EX_{1,1}^2 = 1$, condition on α in the representation (14.10) and let ρ_r^*, ρ_c^*, ρ^* be the conditional correlations. Then

$$\begin{aligned}\rho^* &= E^2(X_{1,1}|\alpha) \\ \rho_r^* &= \{E(X_{1,1}X_{1,2}|\alpha) - E^2(X_{1,1}|\alpha)\}/\text{var}(X_{1,1}|\alpha) \\ &= \{E(X_{1,1}X_{1,2}|\alpha) - \rho^*\}/\text{var}(X_{1,1}|\alpha)\end{aligned}$$

and similarly for ρ_c^*, replacing $X_{1,2}$ with $X_{2,1}$. And

$$\begin{aligned}\rho = EX_{1,1}X_{2,2} &= E\, E(X_{1,1}X_{2,2}|\alpha) \\ &= E\{E(X_{1,1}|\alpha)E(X_{2,2}|\alpha)\} \\ &= E\rho^* .\end{aligned}$$

So
$$\begin{aligned}\rho_c = EX_{1,1}X_{1,2} &= E\, E(X_{1,1}X_{1,2}|\alpha) \\ &\geq E\{E(X_{1,1}|\alpha)E(X_{1,2}|\alpha)\} \\ &\quad \text{by (1.8) for the first row of the} \\ &\quad \text{conditioned array}\end{aligned}$$

$$= E\rho^* = \rho,$$

and similarly for ρ_r. Finally, $\rho_r^* + \rho_c^* \leq 1$ by (15.4), so

$$\begin{aligned} E(X_{1,1}X_{1,2}|\alpha) + E(X_{1,1}X_{2,1}|\alpha) &\leq E\ \mathrm{var}(X_{1,1}|\alpha) + 2\rho^* \\ &= \mathrm{var}(X_{1,1}) - \mathrm{var}\ E(X_{1,1}|\alpha) + 2\rho^* \\ &= 1 + \rho^*, \end{aligned}$$

and taking expectations gives $\rho_c + \rho_r \leq 1 + \rho$.

Remarks. (a) Proposition 15.1 could alternatively be derived by analytic methods, without using Theorem 14.11.

(b) Proposition 15.5 implies that the general Gaussian RCE array is of the form (15.3).

(c) One could consider "second order" RCE arrays, in which only the correlation structure is assumed invariant. Such arrays are discussed in Bailey et al. (1984) as part of an abstract treatment of analysis of variance. Invariance of higher moments is discussed by Speed (1982).

(15.5) Estimating the representing function. Corollary 14.14 says that a dissociated {0,1}-valued RCE array is a ϕ-process. In other words, the family of dissociated {0,1}-valued RCE arrays can be regarded as a parametric family, parametrized by the set Φ of measurable functions $\phi: [0,1]^2 \to [0,1]$. Can one consistently estimate the parameter? More precisely,

(15.6) Problem. Do there exist functionals $\Lambda_N: \{0,1\}^{\{1,\dots,N\}\times\{1,\dots,N\}} \to \Phi$ such that for any ϕ-process X,

$$\Lambda_N(X_{i,j}: i,j \leq N) \to \phi^* \text{ a.s. in } L^1([0,1]^2),$$

for some ϕ^* such that X is a ϕ^*-process?

In view of the non-uniqueness of ϕ representing X, one should really expect a somewhat weaker conclusion: but no results are known.

(15.7) Spherical matrices. Schoenberg's theorem (3.6) asserts that every spherically symmetric infinite sequence is a mixture of i.i.d. $N(0,v)$ sequences. Dawid (1977) introduced the analogous concept for matrices. Call an infinite array $Y = (Y_{i,j}: i,j \geq 1)$ spherical if for each $n \geq 1$,

$$U_1 Y_n U_2 \stackrel{\mathcal{D}}{=} Y_n \quad \text{for all orthogonal } n \times n \text{ matrices } U_1, U_2 ,$$

where Y_n denotes the $n \times n$ matrix $(Y_{i,j}: 1 \leq i,j \leq n)$. Here are two examples of spherical arrays:

(i) a normal array Y, where $(Y_{i,j}: i,j \geq 1)$ are i.i.d. $N(0,1)$;

(ii) a product-normal array Y, where $Y_{i,j} = V_i W_j$ for $(V_1, V_2, \ldots, W_1, W_2, \ldots)$ i.i.d. $N(0,1)$.

Now a spherical array is RCE, by considering permutation matrices. Then Theorem 14.11 can be applied to obtain (with some effort) the following result, conjectured in Dawid (1978) and proved in Aldous (1981a).

(15.8) Corollary. *For an array $Y = (Y_{i,j}: i,j \geq 1)$ the following are equivalent:*

(a) *Y is spherical and dissociated, and $EY^2_{1,1} < \infty$.*

(b) *$Y = a_0 Y^0 + \sum a_m Y^m$; where $\sum a_m^2 < \infty$, Y^0 is Normal, Y^m $(m \geq 1)$ are product-Normal, and $(Y^m: m \geq 0)$ are independent.*

It is clear that (b) implies (a): let us say a few words about the implication (a)$\Rightarrow$(b). Writing $Y = E(Y|\mathcal{S}) + \hat{Y}$, where $\mathcal{S}$ is the shell σ-field of Y, it can be shown that $E(Y|\mathcal{S})$ and $\hat{Y}$ are independent; that $E(Y|\mathcal{S})$ is of the form $a_0 Y^0$; and that $\hat{Y}$ is spherical, dissociated and $\mathcal{S}$-measurable. Corollary 14.15 gives a representation $\hat{Y}_{i,j} = f(\xi_i, \eta_j)$, and the constants

(a_i) appear as the eigenvalues of the integral operator $h(\cdot) \to \int f(\cdot,y)h(y)dy$ associated with f.

(15.9) Finite arrays. Proposition 5.6 gave bounds on how far the initial portion of a finite exchangeable sequence could differ from the initial part of an infinite exchangeable sequence. Analogously, for $m \le n$ let $c_{m,n}$ be the smallest number such that for any $n \times n$ RCE array X taking values $\{0,1\}$ only there exists an infinite RCE array Y such that

$$\| L(X_{i,j}: 1 \le i,j \le m) - L(Y_{i,j}: 1 \le i,j \le m) \| \le c_{m,n} .$$

Weak convergence arguments show that $\lim_{n\to\infty} c_{m,n} = 0$ for each m, but still open is

(15.10) Problem. Give explicit upper bounds for $c_{m,n}$.

(15.11) Continuous-parameter processes. For 2-parameter processes $X_{s,t}$, $0 \le s,t < \infty$, $X_{0,0} = 0$, we can define analogs of the 1-parameter "processes with interchangeable increments" discussed in Section 10. For a rectangle $B = (s_1,s_2] \times (t_1,t_2]$ and a function $f(s,t)$ let $f(B)$ be the increment of f over B:

$$f(B) = f(s_2,t_2) + f(s_1,t_1) - f(s_1,t_2) - f(s_2,t_1) .$$

For fixed $\delta > 0$ let $B_{i,j} = ((i-1)\delta, i\delta] \times ((j-1), j\delta]$. Say X has separately interchangeable increments if

(15.12) the array $X(B_{i,j})$, $i,j \ge 1$ is RCE (for each δ).

Say X has simultaneously interchangeable increments if

(15.13a) $$X_{s,t} = X_{t,s}$$

(15.13b) $(X(B_{i,j}), i,j \geq 1) = (X(B_{\pi(i),\pi(j)}), i,j \geq 1)$ for each permutation π (for each δ).

(Condition (15.13b) is slightly more than weak exchangeability.)

Using Theorems 14.11 and 14.21 to describe all processes with these invariance properties is perhaps a feasible project, but looks much harder than the 1-parameter result, Proposition 10.5. Rather than tackle the general case, let us look at two special cases, corresponding to the special case in Corollary 10.6.

(15.14) <u>2-parameter counting processes with separately interchangeable increments</u>

Here are four methods of constructing such processes:

(i) Let $\lambda > 0$. Take a Poisson process on $R^+ \times R^+$ of rate λ.

(ii) Let $\lambda_1, \lambda_2, \ldots > 0$; let $(\xi_i^k: k \geq 1)$ be i.i.d. $U(0,1)$; and let $f: \mathbb{N} \times [0,1] \to [0,\infty)$ be such that $\sum_k \lambda_k f(k, \xi_1^1) < \infty$ a.s. For each k take a Poisson process of horizontal lines of rate λ_k and attach labels $(\xi_i^k: i \geq 1)$ to these lines; independently for different k. On the line labelled ξ_i^k place a Poisson process of points of rate $f(k, \xi_i^k)$, independently for different lines.

(iii) The analog of (ii) with vertical lines, using constants $(\hat{\lambda}_m)$, say, and labels (η_j^m), say.

(iv) Construct both a process of horizontal lines as in (ii), and a process of vertical lines as in (iii). Let $g: \mathbb{N} \times [0,1] \times \mathbb{N} \times [0,1] \to [0,1]$ be such that $\sum_{k,m} \lambda_k \hat{\lambda}_m g(k, \xi_1^1, m, \eta_1^1) < \infty$ a.s. At the intersection of the lines labelled ξ_i^k and η_j^m put a point with probability $g(k, \xi_i^k, m, \eta_j^m)$.

Since the superposition (i.e. sum) of independent invariant processes is invariant, it is natural to make

(15.15) Conjecture. The general ergodic process of type (15.14) is a sum of independent processes of types (i)-(iv) above.

Presumably this can be proved by applying Theorem 14.11 to the arrays of point counts in small squares, and letting the sizes of the squares decrease to zero; but the details look messy.

(15.16) 2-parameter continuous-path processes with simultaneously interchangeable increments. Here are two examples of such processes:

(15.17) $Y_{s,t} = B_s B_t$; where $(B_t : t \geq 0)$ is Brownian motion;

(15.18) On $\nabla = \{0 \leq s \leq t < \infty\}$ let X be 2-parameter Brownian motion; that is, for rectangles B with Lebesgue measure $|B|$, the variable $X(B)$ has Normal $N(0,|B|)$ distribution, and variables $X(B_i)$ are independent for disjoint B_i. For $t < s$ let $X_{s,t} = X_{t,s}$.

Also, the deterministic processes $X_{s,t} = st$ and $X_{s,t} = \min(s,t)$ have the required properties. From these examples we can construct more processes of the form

(15.19) $$Z_{s,t} = \alpha X_{s,t} + \sum_{j=1}^{\infty} \beta_j Y_{s,t}^{(j)} + \gamma st + \delta \min(s,t)$$

where X has distribution (15.18)

$Y^{(j)}$ has distribution (15.17), for each $j \geq 1$

$(X, Y^{(1)}, Y^{(2)}, \ldots)$ are independent

$(\alpha, \beta_1, \beta_2, \ldots, \gamma, \delta)$ are random variables, and $\sum \beta_j^2 < \infty$ a.s.

It is natural to make the

(15.20) Conjecture. Any 2-parameter continuous-path process $Z_{s,t}$ with simultaneously interchangeable increments has a representation of the form (15.19).

These types of processes arise naturally in the study of the asymptotic distributions of U-statistics. Let us describe the simplest cases. Let $(V_1, V_2, \ldots)$ be i.i.d. and let $g(x,y)$ be a symmetric function such that

$$Eg(V_1,V_2) = 0; \quad Eg^2(V_1,V_2) < \infty .$$

As at (14.18), define the U-statistics

$$U_n = \frac{1}{\binom{n}{2}} \sum\sum_{1 \le i < j \le n} g(V_i, V_j)$$

Also define

$$\sigma^2 = E\big(E(g(V_1,V_2)|V_1)\big)^2 \ge 0 .$$

Then we get the following fundamental theorem for the asymptotic behavior of U-statistics. See Serfling (1980), Chapter 5, for proof and a detailed discussion of U-statistics.

(15.21) Theorem. (i) *Suppose* $\sigma^2 > 0$. *Then* $n^{1/2} U_n \xrightarrow{\mathcal{D}} N(0, 4\sigma^2)$.

(ii) *Suppose* $\sigma^2 = 0$. *Then* $nU_n \xrightarrow{\mathcal{D}} \sum \lambda_j (W_j^2 - 1)$, *where* $(W_1, W_2, \ldots)$ *are independent* $N(0,1)$ *and* $(\lambda_1, \lambda_2, \ldots)$ *are the eigenvalues of the operator* A *defined by* $(Af)(x) = Eg(x, V_2) f(V_2)$.

These results are analogous to the ordinary central limit theorem; what is the corresponding "process" result? Define processes

$$\begin{aligned} Z^{(n)}_{s,t} &= \frac{1}{\binom{n}{2}} \sum\sum_{\substack{1 \le i \le [ns] \\ 1 \le j \le [nt] \\ i<j}} g(V_i, V_j) \; ; \quad s \le t \\ &= Z^{(n)}_{t,s} \; ; \quad s \ge t. \end{aligned}$$

Then $Z^{(n)}$ has simultaneously interchangeable increments, and $U_n = Z^{(n)}_{1,1}$. Thus we would anticipate that any process arising as a limit of the normalized

$Z^{(n)}$ should be of the form (15.19). This is indeed true; the arguments of Mandelbaum and Taqqu (1983) yield

(15.22) Theorem. (i) *Suppose* $\sigma^2 > 0$. *Then* $n^{1/2}Z^{(n)} \xrightarrow{\mathcal{D}} 4\sigma^2 X$, *where* X *has distribution* (15.18).

(ii) *Suppose* $\sigma^2 = 0$. *Then* $nZ^{(n)} \xrightarrow{\mathcal{D}} \sum\lambda_j(Y^{(j)}_{s,t} - \min(s,t))$, *where* $(Y^{(1)}, Y^{(2)}, \ldots)$ *are independent with distribution* (15.17), *and* (λ_j) *are as in Theorem 15.21*.

(15.23) Exchangeable random partitions. Consider a weakly exchangeable family $(R_{i,j})$ of events. By the analog of Proposition 14.6 for weak exchangeability, the family is a mixture of dissociated families. As in (14.14), a dissociated family of events can be described as a ϕ-process:

$$P(R_{i,j} | \xi_1, \xi_2, \ldots) = \phi(\xi_i, \xi_j)$$

where $\phi\colon [0,1]^2 \to [0,1]$ is now symmetric. This leads to an alternative argument for Proposition 11.9. An exchangeable partition (11.4) is a weakly exchangeable array $(R_{i,j})$ with the special property

$$P(R_{i,j} \cap R_{j,k} \cap R^c_{i,k}) = 0; \quad i,j,k \text{ distinct.}$$

This translates into the following property for ϕ:

$$E\,\phi(\xi_1,\xi_2)\phi(\xi_2,\xi_3)(1 - \phi(\xi_1,\xi_3)) = 0 .$$

It is not hard to deduce from this that

$$\phi(x,y) = \sum_n 1_{B_n \times B_n}(x,y) \text{ a.e. for some disjoint } (B_n) .$$

This in turn implies that $(R_{i,j})$ has the "paintbox" distribution ψ_p of Section 11, where $(p_1, p_2, \ldots)$ is the decreasing rearrangement of $(|B_n|)$.

Thus any <u>dissociated</u> exchangeable partition has a "paintbox" distribution, and the general case is a mixture.

(15.24) <u>Ergodic theory techniques</u>. One characterization of RCE arrays of the form $f(\xi_i,\eta_j)$ was given in Corollary 14.15. Here we show how ergodic theory concepts lead to another characterization.

The $\bar{d}$ metric (12.28) on the set of distributions of RCE arrays is

$$\bar{d}(\mu,\nu) = \inf\{E\min(1,|X_{1,1}-Y_{1,1}|)\}$$

where the infimum is taken over bivariate RCE arrays $Z_{i,j} = (X_{i,j},Y_{i,j})$ such that $L(X) = \mu$, $L(Y) = \nu$. It turns out (Aldous (1982a)) that the $\bar{d}$ topology is strictly stronger than the weak topology; and the $\bar{d}$ topology fits in nicely with the characterization theorem. Consider for simplicity only dissociated arrays. Corollary 14.13 says that any RCE distribution μ has a <u>representation</u>

$$\mu = L(X); \quad X_{i,j} = f(\xi_i,\eta_j,\lambda_{i,j}) ,$$

for some f in the space L^0 of functions $f\colon (0,1)^3 \to \mathbb{R}$. Give L^0 the topology of convergence in measure.

(15.25) <u>Lemma</u>. $\mu_k \xrightarrow{\bar{d}} \mu_\infty$ <u>iff there exist functions</u> f_k <u>representing</u> μ_k <u>with</u> $f_k \to f_\infty$.

<u>Proof</u>. The "if" is straightforward, considering $(X_{i,j},Y_{i,j}) = (f_k(\xi_i,\eta_j,\lambda_{i,j}),f_\infty(\xi_i,\eta_j,\lambda_{i,j}))$. Conversely, suppose $\mu_k \xrightarrow{\bar{d}} \mu_\infty$. For each k there exists a bivariate RCE array $(X^k_{i,j},X^\infty_{i,j})$ such that $L(X^k) = \mu_k$, $L(X^\infty) = \mu_\infty$ and

$$E\min(1,|X^k_{1,1}-X^\infty_{1,1}|) \to 0 . \tag{15.26}$$

By taking $(X^k: 1 \le k < \infty)$ conditionally independent given X^∞, we can form a process $(X^\infty; X^k, 1 \le k < \infty)$ with the bivariate distributions (X^k, X^∞) above and such that $\chi_{i,j} = (X^\infty_{i,j}; X^k_{i,j}, k \ge 1)$ is a RCE $\mathbb{R}^\infty$-valued array. By Corollary 14.13 we can take $\chi_{i,j} = g(\xi_i, \eta_j, \lambda_{i,j})$ for some $g: (0,1)^3 \to \mathbb{R}^\infty$. Now set $g_k = g \circ \pi_k$, where $\pi_k((x_r)) = x_k$. Then g_k represents μ_k, and (15.26) implies

$$E \min(1, |g_k(\xi_1, \eta_1, \lambda_{1,1}) - g_\infty(\xi_1, \eta_1, \lambda_{1,1})|) \to 0 \ ,$$

so $g_k \to g_\infty$ in L^0.

Next, recall the definition and elementary properties of entropy. A random variable Y with finite range (y_i) has entropy $\mathcal{E}(Y) = -\sum P(Y = y_i) \log P(Y = y_i)$. And

(15.27)(a) $\mathcal{E}(h(Y)) \le \mathcal{E}(Y)$; any function h.

(b) $\mathcal{E}(X,Y) = \mathcal{E}(X) + \mathcal{E}(Y)$ for independent X, Y.

(c) $\mathcal{E}(Y) \ge E\, \mathcal{E}(Y|F)$ for any σ-field F, where $\mathcal{E}(Y|F)(\omega)$ is the entropy of the conditional distribution $\alpha(\omega, \cdot)$ of Y given F.

For a dissociated RCE array Y such that $Y_{i,j}$ takes values in a finite set, let $\mathcal{E}^Y_n = \mathcal{E}(Y_{i,j}: 1 \le i,j \le n)$. Say Y has <u>linear entropy</u> if $\limsup n^{-1} \mathcal{E}^Y_n < \infty$.

(15.28) <u>Proposition</u>. <u>A dissociated RCE array has a representation as $(f(\xi_i, \eta_j): i,j \ge 1)$ for some f iff it is in the $\bar{d}$-closure of the set of linear entropy arrays</u>.

<u>Proof</u>. Suppose μ is the distribution of an array $(f(\xi_i, \eta_j))$. Let F_k be the set of functions $g: (0,1)^2 \to R$ which are constant on each square of the form $(r2^{-k}, (r+1)2^{-k}) \times (s2^{-k}, (s+1)2^{-k})$. Martingale convergence says

there exist $f_k \in F_k$ such that $f_k \to f$ in measure. Let Y^k be the array $(f_k(\xi_i,\eta_j))$. Then $L(Y^k) \xrightarrow{\bar{d}} \mu$ by Lemma 15.25. Now fix k, and set $\hat{\xi}_i = r2^{-k}$ on $\{r2^{-k} < \xi_i < (r+1)2^{-k}\}$, and similarly for $\hat{\eta}_j$. Then $(Y_{i,j}: i,j \le n)$ is a function of $(\hat{\xi}_i,\hat{\eta}_j: i,j \le n)$. So

$$
\begin{aligned}
E_n^Y &\le E(\hat{\xi}_i,\hat{\eta}_j: i,j \le n) && \text{by (15.27)(a)} \\
&= 2n \log(2^k) && \text{by (15.27)(b).}
\end{aligned}
$$

So Y^k has linear entropy.

For the converse we need

(15.29) Lemma. For a finite-valued dissociated RCE array Y, either

(a) there exists $b > 0$ such that $E_n^Y \ge bn^2$, $n \ge 1$; or

(b) each representation f for Y has $f(\xi_1,\eta_1,\lambda_{1,1}) = \bar{f}(\xi_1,\eta_1)$ a.s. for some $\bar{f}$.

Proof. If (b) fails for some representation f, then there exists a subset $B \subset (0,1)^2$ with measure $|B| > 0$ and there exists $\delta > 0$ such that

$$E(f(x,y,\lambda_{1,1})) > \delta, \quad (x,y) \in B.$$

Define
$$
\begin{aligned}
F_n &= \sigma(\xi_i,\eta_j: i,j \le n) \\
C_n &= \#\{(i,j): i,j \le n,\ (\xi_i,\eta_j) \in B\}\ .
\end{aligned}
$$

Then $E(Y_{i,j}: i,j \le n | F_n) \ge \delta C_n$ by (15.27)(b), and then using (15.27)(c) $E_n^Y \ge \delta E C_n = \delta |B| n^2$.

For the converse part of Proposition 15.28, let $X = (f(\xi_i,\eta_j,\lambda_{i,j}))$ be in the $\bar{d}$-closure of the set of linear entropy arrays. By Lemma 15.26 there exist $f_k: (0,1)^3 \to \mathbb{R}$ such that $f_k \to f$ in measure and f_k represents a linear entropy array. But by Lemma 15.29 $f_k(\xi_1,\eta_1,\lambda_{1,1}) = \bar{f}_k(\xi_1,\eta_1)$ a.s. for some $\bar{f}_k$, and this implies $f(\xi_1,\eta_1,\lambda_{1,1}) = \bar{f}(\xi_1,\eta_1)$ a.s. for some $\bar{f}$.

Remarks. With somewhat more work, one can show that for any dissociated finite-valued RCE array Y represented by f

$$n^{-2}E_n^Y \to \int_0^1\int_0^1 E(f(x,y,\lambda_{1,1}))dxdy .$$

This leads to an alternative characterization in Proposition 15.28. In particular, consider Y of the form $g(\xi_i,\eta_j)$ for some finite-valued g. The assertion above implies E_n^Y is $o(n^2)$:

(15.30) Problem. What is the exact growth rate of E_n^Y in terms of g?

16. The infinite-dimensional cube

Here we present a final example of partial exchangeability where the characterization problem has not been solved--perhaps the examples given here will encourage the reader to tackle the problem.

Let I be the set of infinite sequences $i = (i_1,i_2,\ldots)$ of 0's and 1's such that $\#\{n: i_n = 1\} < \infty$; let I_d be the subset of sequences i such that $i_n = 0$ for all $n > d$. Think of I_d as the set of vertices of the d-dimensional unit cube; think of I as the set of vertices of the infinite-dimensional cube. For a permutation π of $\mathbb{N}$ leaving $\{d+1,d+2,\ldots\}$ fixed, define $\tilde{\pi}: I \to I$ by

$$(16.1) \qquad (\tilde{\pi} i)_n = i_{\pi(n)} .$$

Geometrically, $\tilde{\pi}$ acts on the cube I_d as a rotation about the origin $\underset{\sim}{0}$. For $1 \le s \le d$ define $r_s: I \to I$ by

$$(16.2) \qquad \begin{aligned} (r_s i)_n &= i_n \quad , \quad n \ne s \\ &= 1 - i_n , \quad n = s . \end{aligned}$$

Geometrically, r_s acts on the cube I_d as a reflection in the hyperplane

$\{\underset{\sim}{x}: x_s = \frac{1}{2}\}$. The group Γ_d of isometries of the cube I_d is generated by $\{r_s,\ 1 \leq s \leq d;\ \tilde{\pi},\ \pi$ acting on $\{1,2,\ldots,d\}\}$. And we can regard $\Gamma = \cup \Gamma_d$ as the group of isometries of the infinite-dimensional cube I. Note that I and Γ are both countable.

The pair (I,Γ) fits into the general partial exchangeability setting of Section 12. We are concerned with processes $X = (X_i: i \in I)$, where X_i takes values in some space S, which are invariant in the usual sense

$$X \overset{\mathcal{D}}{=} (X_{\gamma(i)}: i \in I);\quad \text{each } \gamma \in \Gamma$$

For such a process, the processes $X^d = (X_i^d: i \in I_d)$ are invariant processes on the finite-dimensional cubes I_d with the natural consistency property; conversely, any consistent family of invariant processes on the finite-dimensional cubes yields a process on the infinite-dimensional cube.

Here is some more notation. For $i \in I$ let $C_i = \{n: i_n = 1\}$. For $i,j \in I$ let $d(i,j) = \#(C_i \Delta C_j)$, so $d(i,j)$ is the number of edges on the minimal path of edges from i to j. A <u>path</u> in I is a sequence $i^1, i^2, i^3, \ldots$ of vertices such that the sets $C_{i^k} \Delta C_{i^{k+1}}$ are distinct singletons.

As well as the obvious example of i.i.d. processes, there is a related class of invariant processes which involve the "period 2" character of the cube. Given two distributions μ, ν on S let $\theta^0_{\mu,\nu}$ be the distribution of the process (X_i) consisting of independent random variables such that $L(X_i) = \mu$ when $\#C_i$ is even, $L(X_i) = \nu$ when $\#C_i$ is odd. Then the mixture $\theta_{\mu,\nu} = \frac{1}{2}\theta^0_{\mu,\nu} + \frac{1}{2}\theta^0_{\nu,\mu}$ is invariant.

Before proceding further, the reader may like to attempt to construct other examples of invariant processes.

It is interesting to note that an invariant process on the infinite-dimensional cube contains, as subprocesses, examples of other partially exchangeable structures we have described. Let X be invariant.

(16.3) The variables at distance 1 from $\underset{\sim}{0}$, that is $\{X_i: \#C_i = 1\}$, are exchangeable (in fact, exchangeable over $X_{\underset{\sim}{0}}$).

(16.4) The variables at distance 2, that is $\{X_i: \#C_i = 2\}$, form a weakly exchangeable array.

The next result is less obvious. Regard I as a graph. Let $\bar{\Gamma} \supset \Gamma$ be the set of graph-automorphisms of I, that is the set of bijections $\gamma: I \to I$ such that (i,j) is an edge iff $(\gamma(i),\gamma(j))$ is an edge. It is not hard to see that any Γ-invariant process is $\bar{\Gamma}$-invariant.

(16.5) Lemma. There exists a subset $T \subset I$ which is an infinitary tree, in the sense of Section 13, and such that every tree-automorphism γ of T extends to a graph-automorphism θ of I. Hence if $(X_i: i \in I)$ is an invariant process on the cube I then the restriction $(X_i: i \in T)$ is an invariant process on the infinitary tree T.

Proof. As in Section 13 let D be the set of finite sequences $d = (d_1,\ldots,d_m)$ of strictly positive integers. Let $f: D \to \mathbb{N}$ be the prime factorization map $f(d_1,\ldots,d_m) = 2^{d_1}\cdot 3^{d_2}\cdot\ldots$. Now define $\psi: D \to I$ by $C_{\psi(d)} = \{f(d_1),f(d_1,d_2),\ldots,f(d_1,\ldots,d_m)\}$. Then $T = \psi(D)$ is an infinitary tree and $\{\psi(d): d \in D\}$ is a labelling scheme for T.

Now fix a tree-automorphism $\gamma: T \to T$. The map γ induces a map $\tilde{\gamma}: f(D) \to f(D)$ in the following way: if γ maps the edge $(\psi(d),\psi(dq))$ to the edge $(\psi(\hat{d}),\psi(\hat{d}\hat{q}))$ then let $\tilde{\gamma}$ map $f(dq)$ to $f(\hat{d}\hat{q})$. Now define $\theta: I \to I$ as follows. Let $\theta(\underset{\sim}{0}) = \gamma(\underset{\sim}{0}) = \gamma(\psi(\phi))$. For $i \neq \underset{\sim}{0} \in I$ write $C_i = A_i \cup B_i$, where $A_i = C_i \cap f(D)$ and $B_i = C_i \setminus f(D)$. Define $\theta(i)$ by

$$C_{\theta(i)} \,\Delta\, C_{\theta(\underset{\sim}{0})} = \tilde{\gamma}(A_i) \cup B_i \ .$$

By construction θ is an extension of γ. And θ is a graph-automorphism because (i,j) is an edge iff $\#(C_i \Delta C_j) = 1$ iff $\#(C_{\theta(i)} \Delta C_{\theta(j)}) = 1$.

Lemma 16.5 has one noteworthy consequence. For an invariant process on the infinite-dimensional cube with square-integrable real entries, the correlations $\rho(X_i, X_j)$ equal $\rho(d(i,j))$ for some correlation function $\rho(n)$. By (16.5), $\rho(n)$ must be of the form described in Proposition 13.22. Example 16.9 later shows that for each $\lambda \in [-1,1]$ there exists an invariant process with $\rho(n) = \lambda^n$, so by taking mixtures we get

(16.6) Corollary. *A sequence* $(\rho(n): n \geq 0)$ *is the correlation function of some invariant process on the infinite-dimensional cube iff* $\rho(n) = \int x^n \lambda(dx)$ *for some probability measure* λ *on* $[-1,1]$.

This result can be proved by harmonic analysis--see Mansour (1981), who also describes the correlation functions of invariant processes on finite-dimensional cubes. Kingman (personal communication) also has a direct proof of Corollary 16.6.

We now describe a sequence of examples of invariant processes, which we shall loosely refer to as "symmetric random walk models." Here is the basic example, suggested by Kingman.

(16.7) Example. *The basic random walk.* Let $(S,+)$ be a compact Abelian group. Let ξ be a random element of S whose distribution is symmetric, in the sense $\xi \stackrel{\mathcal{D}}{=} -\xi$. Let U be a random element of S whose distribution is Haar measure (i.e. uniform), independent of ξ. Then

$$(\xi+U, U) \stackrel{\mathcal{D}}{=} (U, \xi+U) . \tag{16.8}$$

For $(U,\xi+U) = (-\xi+(\xi+U),\xi+U)$

$\overset{\mathcal{D}}{=} (-\xi+U,U)$ because $\xi+U$ is uniform and independent of ξ

$\overset{\mathcal{D}}{=} (\xi+U,U)$ by symmetry.

Now let $\xi_1,\xi_2,\xi_3,\dots$ be independent copies of ξ, independent of U. For $i \in I$ define

$$X_i = \xi_{n_1}+\xi_{n_2}+\dots+\xi_{n_m}+U, \quad \text{where } \{n_1,\dots,n_m\} = C_i = \{n: i_n = 1\} .$$

Then X is invariant: for invariance under the maps $\tilde{\pi}$ of (16.1) is immediate, and invariance under the maps r_s of (16.2) follows from (16.8).

As a particular case of Example 16.7, suppose

$$\text{(16.9)} \quad S = \{1,0\}; \quad P(\xi=1) = \tfrac{1}{2}(1-\lambda), \quad P(\xi=0) = \tfrac{1}{2}(1+\lambda);$$
$$P(U=1) = P(U=0) = \tfrac{1}{2} .$$

This process has correlation function $\rho(n) = \lambda^n$; indeed, the tree-process which is embedded in X by (16.5) is precisely the tree-process exhibited in the proof of Proposition 13.22.

(16.10) <u>Example</u>. <u>A generalized random walk</u>. Let $(G,\circ)$ be an Abelian group acting on a space S; that is, G consists of functions $g: S \to S$ which form a group under convolution. Let ξ and U be independent random elements of G and S respectively, and suppose

$$\text{(16.11)} \qquad (\xi(U),U) \overset{\mathcal{D}}{=} (U,\xi(U)) .$$

Now let $\xi_1,\xi_2,\dots$ be independent copies of ξ, independent of U, and let

$$X_i = \xi_{n_1}\circ\xi_{n_2}\circ\dots\circ\xi_{n_m}(U), \quad \text{where } \{n_1,\dots,n_m\} = C_i .$$

Then X is invariant, by the same argument as in the previous example.

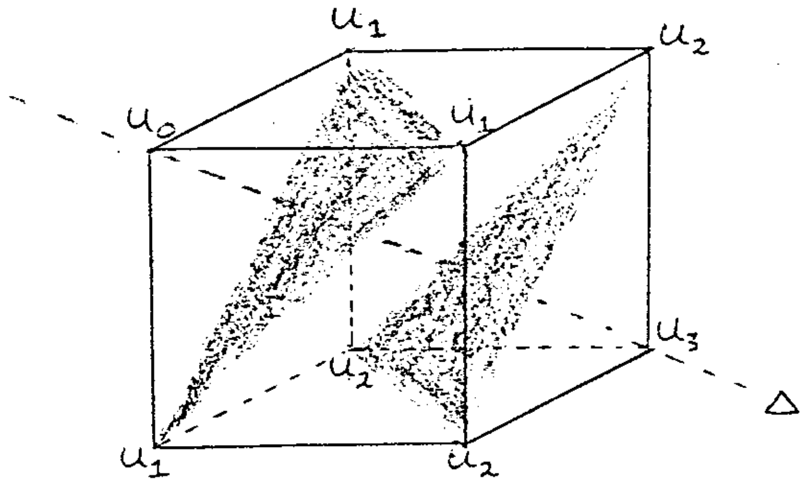
This construction can yield processes rather more general than is suggested by the phrase "random walk," as the next example shows.

Remark. We call this a "random walk" model because the values $X_{i^1}, X_{i^2}, \ldots$ along a path $i^1, i^2, \ldots$ in I are a random walk on S, in the usual sense.

(16.12) Example. Randomly-oriented stationary process. Let $U = (\ldots, U_{-1}, U_0, U_1, \ldots)$ be an arbitrary stationary sequence. On the d-dimensional cube I_d choose a diagonal Δ at random (uniformly); each vertex lies in one of $d+1$ hyperplanes $H_0, H_1, \ldots, H_d$ orthogonal to Δ; set $X_i = U_m$ for $i \in H_m$. This describes an invariant process indexed by I_d. As d varies, these are consistent, and so determine an invariant process on I. The process on I has the following alternative description. Regard U as a random element of $S = \mathbb{R}^{\mathbf{Z}}$. Let θ^n be the shift on S; $\theta^n((x_m)) = (x_{m+n})$. Let $G = \{\theta^n : n \in \mathbf{Z}\}$ and let ξ be the random element of G such that $P(\xi = \theta^1) = P(\xi = \theta^{-1}) = \frac{1}{2}$. Then (16.11) holds because

$$
\begin{aligned}
L(U, \xi(U)) &= \tfrac{1}{2} L(U, \theta^1(U)) + \tfrac{1}{2} L(U, \theta^{-1}(U)) \\
&= \tfrac{1}{2} L(\theta^{-1}(U), U) + \tfrac{1}{2} L(\theta^1(U), U) \quad \text{by stationarity} \\
&= L(\xi(U), U) .
\end{aligned}
$$

So as in Example 16.10 we can construct an invariant process $\hat{X}$ from U and ξ. Let $g((x_m)) = x_0$. Then the process $X_i = g(\hat{X}_i)$ is the randomly-oriented stationary process described originally.



Here is a different generalization of the basic random walk model.

(16.13) Transient random walk. Let $(G,+)$ be a countable Abelian group. Let ξ be a random element of G and let π be a σ-finite measure on G. Suppose

$$(16.14) \qquad \pi(g_1)P(g_1+\xi = g_2) = \pi(g_2)P(g_2+\xi = g_1); \quad \text{all } g_1, g_2 \in G .$$

This is analogous to (16.8) and (16.11); π is a σ-finite invariant measure for the random walk generated by ξ. This random walk may be transient; consider the particular case

$$(16.15) \qquad G = \mathbb{Z}; \quad P(\xi = -1) = \alpha, \quad P(\xi = 1) = 1-\alpha; \quad \pi(n) = c\left(\frac{1-\alpha}{\alpha}\right)^n .$$

Though the random walk has no stationary distribution in the usual sense, there is a different interpretation. Suppose that at time 0 we place a random number Y_g^0 of particles at each g, where (Y_g^0) are independent and Y_g^0 has distribution Poisson$(\pi(g))$. Then let each particle move independently as a random walk with step distribution ξ. Let Y_g^n be the number of particles at position g at time n, and let $Y^n = (Y_g^n)$, a random element of $S = (\mathbb{Z}^+)^G$. Then it is easy to see, using (16.14),

(16.16) $Y^0, Y^1, Y^2, \ldots$ is a stationary reversible Markov chain.

By adding more detail to the description above, we shall produce a process indexed by the infinite-dimensional cube. Suppose that particle u is initially placed at point $g_0(u)$ and has written on it an i.i.d. sequence $(\xi_1^u, \xi_2^u, \ldots)$ of copies of ξ, representing the successive steps to be made by the particle. So $Y_g^n = \#\{u: g_0(u) + \xi_1^u + \cdots + \xi_n^u = g\}$. Now for $i \in I$ define

(16.17) $X_g^i = \#\{u: g_0(u) + \xi_{j_1}^u + \cdots + \xi_{j_m}^u = g\}$, where $\{j_1, \ldots, j_m\} = C_i$.

So $X^i = (X_g^i)$ describes the configuration of particles when only the jumps at times in C_i are allowed. It is easy to check that $(X^i: i \in I)$ is invariant.

Here is a more concrete example which turns out to be a special case of the construction above.

(16.18) Example. On the d-dimensional cube I_d pick $k = k(d)$ vertices $V_1, \ldots, V_k$ uniformly at random and define

$$X_i^n = \#\{m: d(i, V_m) = n + c\}; \quad i \in I_d, \quad n \in \mathbb{Z},$$

for some given $c = c(d)$. Let $X_i = (X_i^n: n \in \mathbb{Z})$, taking values in $S = (\mathbb{Z}^+)^{\mathbb{Z}}$. Plainly $X^{(d)} = (X_i: i \in I_d)$ is invariant. It can be shown that it is possible to pick $k(d)$ and $c(d)$ such that the processes $X^{(d)}$ converge weakly to some process $\hat{X}$ on the infinite-dimensional cube, and such that

$$d^{-1}\log(k) \to \log(2) + (1-\alpha)\log(1-\alpha) + \alpha \log(\alpha)$$

for any prescribed $0 < \alpha < \frac{1}{2}$. And the limit process $\hat{X}$ is just the particular case (16.15) of the general construction (16.13).

(16.19) Remarks. These "random walk" constructions for invariant processes on the infinite-dimensional cube seem analogous to the constructions $X_{i,j} = f(\xi_i, \eta_j)$ for RCE arrays. Perhaps there is an analog of Corollary 14.15 (resp. Proposition 15.28) which says that an ergodic invariant process on the cube can be represented as a function of some random walk model iff a certain "remote" σ-field contains all the information about the process (resp. iff some "linear entropy" condition holds). On the other hand, it looks plausible that the characterization problem on the cube is rather harder than for RCE arrays, in that the next examples suggest that the general process cannot be obtained from random walk models and independent models.

(16.20) Example. For $1 < k \leq d$ a k-face of I_d is a set of vertices isometric (in I_d) to I_k. Let $X = (X_i : i \in I_d)$ be i.i.d. with $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$. Let Y $(= Y^{k,d})$ be the process X conditioned on the event

$$\sum_{i \in F} X_i = 0 \bmod 2 \quad \text{for each k-face } F.$$

For fixed k, the processes Y are consistent as d increases, and hence determine a process Y^k on the infinite cube. For $k = 2$ this process Y^k is just Example 16.9 with $\lambda = -1$; for $k \geq 3$ the processes Y^k do not seem to have "random walk" descriptions.

Finally, we can construct invariant processes by borrowing an idea from statistical mechanics (see e.g. Kindermann and Snell (1980)).

(16.21) Example. Ising models. Fix $\alpha \in \mathbb{R}$, $d \geq 1$. For a configuration $\underset{\sim}{x} = (x_i : i \in I_d)$ of 0's and 1's on the d-dimensional cube, define

$$V(\underset{\sim}{x}) = \sum_{\text{edges }(i,j)} 1_{(x_i = x_j)} .$$

The function V is invariant under the isometries of the cube, so we can define an invariant distribution by

$$P(X = \underset{\sim}{x}) = c_\alpha \exp(\alpha V(\underset{\sim}{x}))$$

where c_α is a normalization constant. By symmetry, $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$. Let $\rho_{d,\alpha}$ be the correlation $\rho(X_i, X_j)$ for neighbors i, j. For fixed d, $\rho_{d,\alpha}$ increases continuously from -1 to $+1$ as α increases from $-\infty$ to $+\infty$. There are heuristic arguments which suggest

(16.22) $$\rho_{d,\alpha} \to (e^\alpha - 1)/(e^\alpha + 1) \quad \text{as} \quad d \to \infty; \quad \alpha \text{ fixed.}$$

It this is true, then by fixing α, letting $d \to \infty$ and taking (subsequential, if necessary) weak limits we can construct invariant processes on the infinite-dimensional cube with correlation $(e^\alpha - 1)/(e^\alpha + 1)$ between neighbors (and even without (16.22), this holds for some $\alpha(d)$). It would be interesting to get more information about these limit processes; heuristic arguments suggest they are not of the "random walk" types described earlier.

PART IV

17. Exchangeable random sets

In this section we discuss exchangeability concepts for certain types of random subsets M of $[0,1)$ or $[0,\infty)$. Let us start by giving some examples of random subsets M.

(17.1) The zeros of Brownian motion: $M = \{t: W_t = 0\}$.

(17.2) The range of a subordinator: $M = \{X_t(\omega): 0 \le t < \infty\}$, where X_t is a subordinator, that is a Lévy process with $X_0 = 0$ and increasing sample paths.

(17.3) The zeros of Brownian bridge: $M = \{t: W_t^0 = 0\} \subset [0,1]$.

(17.4) An exchangeable interval partition. Take an infinite sequence of constants $c_1 \ge c_2 \ge \cdots > 0$ with $\sum c_i = 1$; take (ξ_i) i.i.d. $U(0,1)$; set

$$L_i = \sum_j c_j 1_{(\xi_j < \xi_i)}, \quad R_i = L_i + c_i .$$

So the intervals (L_i, R_i) have lengths c_i and occur in random order. Let M be the complement of $\bigcup_i (L_i, R_i)$.

These examples all have an exchangeability property we shall specify below. The first three examples are probabilistically natural; the fourth arose in game theory, and attracted interest because certain "intuitively obvious" properties are hard to prove, e.g. the fact (Berbee (1981))

(17.5) $$P(x \in M) = 0 \quad \text{for each } 0 < x < 1.$$

The characterization results for exchangeable sets are roughly similar to those in Section 10 for interchangeable increments processes, but are interesting in that stopping time methods seem the natural tool. Our account closely follows Kallenberg (1982a,b), which the reader should consult for proofs and further results.

Formally, we consider random subsets M of $[0,1]$ or $[0,\infty)$ satisfying

(17.6) $\quad M$ is closed; M has Lebesgue measure zero.

So the complement M^c is a union of disjoint open intervals (L_α, R_α). For each $\varepsilon > 0$ let N_ε be the number of intervals (L_α, R_α) of length at least ε; call these intervals $(L_1^\varepsilon, R_1^\varepsilon), (L_2^\varepsilon, R_2^\varepsilon), \ldots$. Call M _exchangeable_ if for each ε and each $1 \leq n \leq \infty$ the lengths $(R_i - L_i)$ are, conditional on $\{N_\varepsilon = n\}$, an n-exchangeable sequence.

Consider now the case where M is the closed range of a subordinator (i.e. the closure of M in (17.2)). Set $M^t = \{x-t: x \in M, x \geq t\}$. The strong Markov property of the subordinator X_t implies that for any stopping time T taking values in $M' = M \setminus (\{L_\alpha\} \setminus \{R_\alpha\})$ we have

(17.7) $\quad M^T$ is independent of $M \cap [0,T]$; $M^T \overset{\mathcal{D}}{=} M$.

Call random subsets satisfying (17.7) _regenerative sets_. Horowitz (1972) shows a converse: all regenerative sets arise as the closed range of some subordinator. By analogy with (6.1B) and (10.7) consider the condition

(17.8) $\quad M^T \overset{\mathcal{D}}{=} M$; each stopping time $T \in M'$.

Kallenberg calls this _strong homogeneity_. Kallenberg (1982a), Theorem 4.1, proves

(17.9) _Theorem_. _For unbounded random subsets_ $M \subset [0,\infty)$ _satisfying_ (17.6), _the following are equivalent_:

(a) M _is exchangeable_.

(b) M *is strongly homogeneous.*

(c) M *is a mixture of regenerative sets.*

For finite intervals we get a weaker result: Kallenberg (1982a), Theorem 4.2 implies

(17.10) *Proposition. For random subsets* $M \subset [0,1]$ *satisfying* (17.6) *and with a.s. infinitely many points, the following are equivalent:*

(a) M *is exchangeable.*

(b) M *is a mixture of exchangeable interval partitions.*

Finally, we remark that the classical theory of local time at the zeros of Brownian motion extends to a theory of local time for regenerative sets, and hence for exchangeable subsets of $[0,\infty)$. For exchangeable interval partitions there is an elementary definition of "local time": in (17.4) set

$$Q_t = \xi_i \quad \text{on} \quad (L_i, R_i) .$$

This concept appears useful for tackling problems like (17.5)--see Kallenberg (1983).

18. Sufficient statistics and mixtures

Recall the classical notion of sufficiency. Let $(P_\theta : \theta \in \Theta)$ be a family of distributions on a space S. For notational convenience, let $X: S \to S$ denote the identity map. Then a map $T: S \to \hat{S}$ is a *sufficient statistic* for the family (P_θ) if the P_θ-conditional distribution of X given $T(X)$ does not depend on θ. More precisely, T is sufficient if there exists a kernel $Q(t,A)$, $t \in \hat{S}$, $A \subset S$, such that for each θ

(18.1) $\quad Q(T(X),\cdot)$ is a P_θ-r.c.d. for X given $T(X)$.

For instance, if (P_θ) is the family of distributions of i.i.d. Normal sequences $X = (X_1,\dots,X_n)$ on $S = R^n$, then

(18.2)(a) $T_n(x) = (T_{n,1}(x), T_{n,2}(x)) = (\sum x_i, (\sum x_i^2)^{1/2})$ is sufficient, with

(b) $Q_n((t_1,t_2),\cdot)$ the uniform distribution on the surface of the sphere $\{x: T_n(x) = (t_1,t_2)\}$.

The classical interest in sufficiency has been in the context of inference: if $X_1,\dots,X_n$ are assumed to be observations from a known parametric family, then for inference about the unknown parameter one need consider only statistics which are functions of sufficient statistics.

Our interests are rather different. Consider the following general program. Let T_n, Q_n, $n \geq 1$, be a given sequence of maps and kernels. Then study the set M of distributions of sequences $(X_1,X_2,\dots)$ such that for each n

(18.3) $Q_n(T_n(X_1,\dots,X_n),\cdot)$ is a r.c.d. for $(X_1,\dots,X_n)$ given $T_n(X_1,\dots,X_n)$.

For instance, if T_n, Q_n, are the natural sufficient statistics and kernels associated with an exponential family of distributions (P_θ), then by definition M contains the distributions of i.i.d. P_θ sequences. But M is closed under taking mixtures, so M contains the class M_0 of mixtures of i.i.d. P_θ sequences. It generally turns out that $M = M_0$, and so this program leads to a systematic method for characterizing those exchangeable sequences which are mixtures of i.i.d. sequences with distributions from a specified family.

The general program has a much wider scope than the preceding discussion might suggest. First, observe that the class of exchangeable sequences can be defined in this way. For as at (5.2) let $\Lambda_n: R^n \to P(R)$ be the empirical distribution map, and $\Phi_n^{-1}(\mu,\cdot) = L(x_{\pi^*(1)},\ldots,x_{\pi^*(n)})$, where $\mu = \Lambda_n(x)$ and π^* is the uniform random permutation. Then Lemma 5.4 says $(X_1,\ldots,X_n)$ is exchangeable iff $\Phi_n^{-1}(\Lambda_n(X_1,\ldots,X_n),\cdot)$ is a r.c.d. for $(X_1,\ldots,X_n)$ given $\Lambda_n(X_1,\ldots,X_n)$. Thus the class M associated with the sufficient statistics Λ_n and kernels Φ_n^{-1} is precisely the class of infinite exchangeable sequences. Similarly, the other partially exchangeable models in Part III can be fitted into this setting.

Further afield, the study of Markov random fields (as a probabilistic formulation of statistical mechanics problems--Kindermann and Snell (1980)) involves the same ideas: one studies the class of processes $(X_i: i \in \Gamma)$ on a graph Γ such that the conditional distribution of X_i given the distribution at neighboring vertices $(X_j: j \in N_i)$ has a specified form. Yet another subject which can be fitted into the general program is the study of entrance and exit laws for Markov processes.

This general program has been developed recently by several authors, from somewhat different viewpoints: Dynkin (1978), Lauritzen (1982), Diaconis and Freedman (1982), Accardi and Pistone (1982), Dawid (1982). A main theoretical result is a generalization of Theorem 12.10, describing the general distribution in M as a mixture of "extreme" distributions. Our account closely follows that of Diaconis and Freedman (1982): we now state their hypotheses and their version of this main theoretical result.

Let S_i, W_i, $i \geq 1$, be Polish spaces. Let $X_i: \prod_j S_j \to S_i$ be the coordinate map. Let $T_n: \prod_{i=1}^n S_i \to W_n$, and let Q_n be a kernel $Q_n(w,A)$, $w \in W_n$, $A \subset \prod_{i=1}^n S_i$. Suppose

(18.4)(i) $Q_n(w,\{T_n = w\}) = 1$; $w \in W_n$.

(ii) if $T_n(x) = T_n(x')$ then $T_{n+1}(x,y) = T_{n+1}(x',y)$; $y \in S_{n+1}$.

(iii) for each $w \in W_{n+1}$, $Q_n(T_n(X_1,\ldots,X_n),\cdot)$ is a $Q_{n+1}(w,\cdot)$ r.c.d. for $(X_1,\ldots,X_n)$ given $\sigma(T_n(X_1,\ldots,X_n),X_{n+1})$.

Then let M be the set of distributions P on $\prod_{i\geq 1} S_i$ such that for each n

(18.5) $\quad Q_n(T_n(X_1,\ldots,X_n),\cdot)$ is a P-r.c.d. for $(X_1,\ldots,X_n)$ given $T_n(X_1,\ldots,X_n)$.

Conditions (i) and (ii) are natural: here is an interpretation for (iii). Take the Bayesian viewpoint that (X_i) is an i.i.d. (θ) sequence, where θ has been picked at random from some family. Saying T_n is sufficient is saying that $(X_1,\ldots,X_n)$ and X_{n+1} are conditionally independent given $T_n = T_n(X_1,\ldots,X_n)$. Consider now the conditional distribution of $(X_1,\ldots,X_n)$ given (T_n,X_{n+1},T_{n+1}). By (ii), T_{n+1} is a function of (T_n,X_{n+1}). This and the conditional independence shows that the conditional distribution of $(X_1,\ldots,X_n)$ given (T_n,X_{n+1},T_{n+1}) is the same as the conditional distribution given T_n, which is the kernel distribution $Q_n(T_n,\cdot)$; this is the assertion of (iii). Lauritzen (1982), II.2,3 gives a more detailed discussion.

Next set $S = \bigcap_n \sigma(T_n(X_1,\ldots,X_n),X_{n+1},X_{n+2},\ldots)$, so S is a σ-field on $\prod_{i\geq 1} S_i$. In the context of exchangeable sequences described earlier, S is the exchangeable σ-field. Diaconis and Freedman (1982) prove

(18.6) Theorem. There is a set $S_0 \subset \prod S_i$, $S_0 \in S$, with the following properties:

(i) $P(S_0) = 1$; each $P \in M$.

(ii) $Q(s,\cdot) = \text{weak-limit}_{n\to\infty}\, Q_n(T_n(s),\cdot)$ exists as a distribution on $\prod S_i$; each $s \in S_0$.

(iii) <u>The set of distributions</u> $\{Q(s,\cdot): s \in S_0\}$ <u>is precisely the set of extreme points of the convex set</u> M.

(iv) <u>For each</u> $P \in M$ <u>we have</u> $P(\cdot) = \int_{S_0} Q(s,\cdot)\hat{P}(ds)$, <u>where</u> $\hat{P}$ <u>denotes the restriction of</u> P <u>to</u> S. <u>Thus</u> $Q((X_1,X_2,\ldots),\cdot)$ <u>is a P-r.c.d. for</u> $(X_1,X_2,\ldots)$ <u>given</u> S.

(v) $P \in M$ <u>is extreme iff</u> S <u>is P-trivial</u>.

In the context of exchangeable sequences, S_0 is the set of sequences s for which the limiting empirical distribution $\Lambda(s) = \text{weak-limit } \Lambda_n(s_1,\ldots,s_n)$ exists, and $Q(s,\cdot)$ is the i.i.d. $(\Lambda(s))$ distribution. Thus (iv) recovers a standard form of de Finetti's theorem.

The idea in the proof of Theorem 18.6 is that, if $Q(s,\cdot) = \text{weak-limit } Q_n(T_n(s),\cdot)$ exists, then $Q(s,\cdot)$ defines a distribution in M. Reversed martingale convergence arguments in the spirit of the first proof of de Finetti's theorem show that $Q(s,\cdot)$ exists P-a.s., each $P \in M$. The family of all limiting distributions $Q(s,\cdot)$ is sometimes called the family of <u>Boltzmann laws</u>; this family may contain non-extreme elements of M.

One nice example, outside the context of exchangeability, is the study of mixtures of Markov chains by Diaconis and Freedman (1980b). Let S be a countable set of states. For a sequence $\sigma = (\sigma_1,\ldots,\sigma_n)$ of states and a pair s,t of states let $T_{s,t}(\sigma) = \#\{i: (\sigma_i,\sigma_{i+1}) = (s,t)\}$ be the number of transitions from s to t in the sequence σ. Let $T_n(\sigma) = (\sigma_1; T_{s,t}(\sigma), s,t \in S)$. So $T_n(\sigma) = T_n(\sigma')$ iff σ and σ' have the same initial state and the same transition counts. Now consider a homogenous Markov chain (X_i) on S. Plainly

(18.7) $$P((X_1,\ldots,X_n)=\sigma) = P((X_1,\ldots,X_n)=\sigma') \quad \text{whenever} \quad T_n(\sigma) = T_n(\sigma').$$

Diaconis and Freedman (1980b) prove

(18.8) Proposition. Suppose $X = (X_0, X_1, X_2, \ldots)$ is a process taking values in S which is recurrent, i.e. $P(X_n = X_0$ for infinitely many $n) = 1$. Then X is a mixture of homogenous Markov chains iff X satisfies (18.7).

This fits into the general set-up by making $Q_n(t,\cdot)$ the distribution uniform on the set of sequences σ such that $T_n(\sigma) = t$. Then the set of processes satisfying (18.7) is the set M defined by (18.5); and Proposition 18.8 says that the extreme points of $M \cap \{$recurrent processes$\}$ are precisely the recurrent homogenous Markov chains. (A different characterization of such mixtures is in Kallenberg (1982a).)

Another interesting example is the conditional Rasch model discussed by Lauritzen (1982), II.9.7.

We now turn to characterizations of mixtures of i.i.d. sequences. We have already seen one such result, Schoenberg's Theorem 3.6. To fit this into the present context, take $T_n(x_1,\ldots,x_n) = (\sum x_i^2)^{1/2}$, and let $Q_n(t,\cdot)$ be uniform on the surface of the sphere with center 0 and radius t in R^n. Then the set M defined by (18.5) is the set of spherically symmetric sequences. Schoenberg's theorem asserts that each element of M is a mixture of i.i.d. $N(0,\sigma^2)$ sequences; thus the extreme points of M are the i.i.d. $N(0,\sigma^2)$ sequences. There is a related result for general mixtures of i.i.d. Normal sequences. Take T_n, Q_n as at (18.2); then M can be described as the set of sequences (X_i) such that for each n the random vector $(X_1,\ldots,X_n)$ is invariant under the action of all orthogonal $n \times n$ matrices U which preserve the vector $(1,\ldots,1)$. It can be shown (Dawid (1977a); Smith (1981)) that each process in M is a mixture (over μ, σ) of i.i.d. $N(\mu,\sigma^2)$ sequences. These results can in fact be deduced fairly directly from Theorem 18.6; see Diaconis and Freedman (1982); Dawid (1982) for outlines of the argument.

Consider now discrete distributions. For the family of i.i.d. Poisson (λ) sequences, the sufficient statistics are $T_n(x_1,\ldots,x_n) = \sum x_i$ and the kernels are $Q_n(t,(i_1,\ldots,i_n)) = n^{-t}t!/(i_1!\cdots i_n!)$, $\sum i_j = t$. It is natural to hope that M, defined by (18.5), is the class of mixtures of i.i.d. Poisson sequences. This result, and the corresponding results for Binomial and Negative Binomial sequences, are proved in Freedman (1962b). Lauritzen (1982), Section III, gives an abstract treatment of general exponential families.

There are several variations on this theme. One is to consider mixtures of independent non-identically distributed sequences with distributions in a specified family. For example, fix constants $c_i > 0$. For each $\lambda > 0$ let P_λ be the distribution of the independent sequence (X_i), where X_i has Poisson (λ^{c_i}) distribution. Then $T_n(x_1,\ldots,x_n) = \sum c_i x_i$ is sufficient for (P_λ), with kernel $Q_n(t,\cdot)$ being the multinomial distribution of t balls into n equiprobable boxes. Alternatively, for $\mu > 0$ let P_μ be the distribution of the independent sequence (X_i), where X_i has Poisson (μc_i) distribution. Then $T_n(x_1,\ldots,x_n) = \sum x_i$ is sufficient, and the kernel $Q_n(t,\cdot)$ is the multinomial distribution of t balls into n boxes where box i has chance $c_i/\sum c_j$ of being chosen. The structure of M and its extreme points in these examples is discussed in Lauritzen (1982), II.9.20 and in Diaconis and Freedman (1982), Examples 2.5 and 2.6.

So far, we have assumed that both T_n and Q_n are prescribed. Another variant is to prescribe only T_n, and ask what processes are in M for some sequence of kernels Q_n. For instance, it is natural to ask for what classes of exchangeable sequences (X_i) do the partial sums $T_n(x_1,\ldots,x_n) = \sum x_i$ form sufficient statistics; this problem, in the integer-valued case, is discussed in detail in Diaconis and Freedman (1982).

A very recent preprint of Ressel (1983) uses techniques from harmonic analysis on semigroups to obtain characterizations of mixtures of i.i.d. sequences from specific families of distributions. For an infinite sequence $X = (X_j)$ let $\phi_n(\underset{\sim}{t}) = E \exp(\sum_{j=1}^{n} t_j X_j)$. Schoenberg's theorem 3.6 can be stated as

(18.9) If $\phi_n(\underset{\sim}{t}) = f(\sum t_j^2)$ for some function f, then X is a mixture of i.i.d. $N(0,\sigma^2)$ sequences.

Similarly, one can prove the following.

(18.10) If $\phi_n(\underset{\sim}{t}) = f(\sum |t_j|^\alpha)$ then X is a scale mixture of i.i.d. symmetric stable (α) sequences.

(18.11) If $\phi_n(\underset{\sim}{t}) = f\big(\Pi(1 + t_j)\big)$ then X is a mixture of i.i.d. Gamma$(\lambda,1)$ sequences.

Ressel (1983) gives an abstract result which yields these and other characterizations.

19. Exchangeability in population genetics

Perhaps the most remarkable applications of exchangeability are those to mathematical population genetics developed recently by Kingman and others. Our brief account is abstracted from the monograph of Kingman (1980), which the reader should consult for more complete discussion and references.

Consider the distribution of relative frequencies of alleles (i.e. types of gene) at a single locus in a population which is diploid (i.e. with chromosome-pairs, as for humans). Here is the basic Wright-Fisher model for mutation which is neutral (i.e. the genetic differences do not affect fitnesses of individuals).

(19.1) Model. (a) The population contains a fixed number N of individuals (and hence $2N$ genes at the locus under consideration) in each generation.

(b) Each gene is one of a finite number s of allelic types $(A_1,\dots,A_s)$.

(c) Each gene in the $(n+1)^{st}$ generation can be considered as a copy of a uniformly randomly chosen gene from the n^{th} generation, different choices being independent; except

(d) there is a (small) chance $u_{i,j}$ that a gene of type A_i is mistakenly copied as type A_j (mutation).

Let $X_i^N(n)$ be the proportion of type A_i alleles in the n^{th} generation. Then the vector $(X_1^N(n),\dots,X_s^N(n))$ evolves as a Markov chain on a finite state space, and converges in distribution as $n\to\infty$ to some stationary distribution

$$(X_1^N,\dots,X_s^N) \ . \tag{19.2}$$

We shall consider only the special case where all mutations are equally likely:

(19.3) $u_{i,j} = v/s \quad (i \neq j)$, for some $0 < v$.

Then by symmetry $(X_1^N,\dots,X_s^N)$ is exchangeable, so $EX_i^N = s^{-1}$. Consider how this distribution varies with the mutation rate v. In the absence of mutation the frequencies $X_i(n)$ evolve as martingales and so eventually get absorbed at 0 or 1; thus $(X_1^N,\dots,X_s^N) \approx (1_{(U=1)},\dots,1_{(U=s)})$, U uniform on $\{1,\dots,s\}$, as $v \to 0$. On the other hand for large v the mutation effect dominates the random sampling effect, so the allele distribution becomes like the multinomial distribution of $2N$ objects into s classes, so for large v we have $(X_1^N,\dots,X_s^N) \approx (1/s,\dots,1/s)$ + order $N^{-1/2}$. To obtain more quantitative information, observe that the proportion $X_1^N(n)$ of type 1 alleles evolves as a Markov chain. It is not difficult to get an expression for the variance of the stationary distribution which simplifies to

(19.4) $$\operatorname{var}(X_1^N) \approx \frac{s^{-1} - s^{-2}}{1 + 4Nv/(s-1)}, \quad N \text{ large}, \quad v \text{ small}.$$

Of course the biologically interesting case is N large, v small, and we can approximate this by taking the limit as

(19.5) $$N \to \infty, \quad v \to 0, \quad 4Nv \to \theta, \quad \text{say}.$$

Then (19.4) suggests we should get some non-trivial limit

(19.6a) $$(X_1^N,\dots,X_s^N) \xrightarrow{\mathcal{D}} (X_1,\dots,X_s)$$

where X_i represents the relative frequency of allele A_i in a large population with small mutation rate, when the population is in (time-) equilibrium. This is indeed true, and (Watterson (1976))

(19.6b) $(X_1,\dots,X_s)$ has the exchangeable Dirichlet distribution (10.22), for $(a,k) = (\theta,s)$.

The infinite-allele model. The s-allele model above describes recurrent mutation, where the effects of one mutation can be undone by subsequent mutation. An opposite assumption, perhaps biologically more accurate, is to suppose that each mutation produces a new allele, different from all other alleles. So consider model (19.1) with this modification, and let v be the probability of mutation. Fix the population size N. It is clear that any given allele will eventually become extinct. So instead of looking at proportions of alleles in prespecified order, look at them in order of frequency; let $Y_1^N(n)$ be the proportion of genes in generation n which are of the most numerous allelic type: $Y_2^N(n)$ the proportion of the second most numerous type, and so on. Again $(Y_1^N(n), Y_2^N(n), \ldots)$ evolves as a finite Markov chain and so converges to a stationary distribution $(Y_1^N, Y_2^N, \ldots)$ with $\sum_i Y_i^N = 1$. Again it is easy to see how this distribution depends on the mutation probability v: as $v \to 0$ we have $Y_1^N \xrightarrow[p]{} 1$; as $v \to 1$ we have each Y_i^N of order (N^{-1}).

What happens as $N \to \infty$? At first sight one might argue that the number of different allelic types in existence simultaneously would increase to infinity, and so the proportions of each type would decrease to zero. But this reasoning is false. In fact, under the assumptions $N \to \infty$, $v \to 0$, $4Nv \to \theta$ used before, we have (see Kingman (1980), p. 40)

(19.7) $(Y_1^N, Y_2^N, \ldots) \xrightarrow{\mathcal{D}} (D_1, D_2, \ldots)$ where (D_i) has the Poisson-Dirichlet(θ) distribution.

Thus for a large population subject to slow, non-recurrent neutral mutation, the proportions of the different alleles present at a particular time, arranged in decreasing order, should follow a Poisson-Dirichlet distribution.

Now consider sampling K genes from such a population. Let a_r be the number of allelic types for which there are exactly r genes of that type in the sample. Then Theorem 11.14 shows that the chance of obtaining a specified $(a_1, a_2, \ldots)$ is given by formula (11.16), the Ewens sampling formula. Indeed, if we consider the partition R^K into allelic types of a sample of size K from a hypothetical limiting infinite population, these random partitions satisfy the consistency conditions of Theorem 11.14.

Let us outline a method for deriving the infinite-allele result (19.7) from the s-allele result (19.6). Fix the population size N. Imagine that each new allele created by mutation is named by a random variable ξ distributed uniformly on $(0,1)$. So each gene g has a label ξ_g which indicates its allelic type. Thus the genetic composition of generation n can be described by a process $(W_n^N(u): 0 \le u \le 1)$, where $W_n^N(u)$ is the proportion of genes g for which $\xi_g \le u$. As $n \to \infty$ this converges to a process $(W^N(u): 0 \le u \le 1)$, where the jump sizes $(W^N(u) - W^N(u-))$, rearranged in decreasing order, are the variables $(Y_1^N, Y_2^N, \ldots)$ above, and the jump positions are independent uniform. Now fix s, and call an allele "type j", $1 \le j \le s$, if its name ξ is in the interval $((j-1)/s, j/s)$. If we only take notice of the "type" of alleles, then the infinite-allele model evolves in precisely the same way as the s-allele model. The convergence result (19.6) translates to

$$(19.8) \qquad (W^N(0), W^N(1/s), \ldots, W^N(1)) \xrightarrow{\mathcal{D}} (Z(0), Z(1/s), \ldots, Z(1)) ,$$

where Z is the Dirichlet(θ) process. But then

$$(19.9) \qquad (W^N(u): 0 \le u \le 1) \xrightarrow{\mathcal{D}} (Z(u): 0 \le u \le 1) \quad \text{in} \quad D(0,1),$$

since (19.8) gives convergence of finite-dimensional distributions, and establishing tightness is an exercise in technicalities. But convergence in $D(0,1)$ implies convergence of jump sizes, and this gives (19.7).

Other applications. There are other, quite different, applications of exchangeability to genetics. Suppose the "fitness" of an individual does depend on his genetic type, an individual with gene-pair (A_i,A_j) having fitness $w_{i,j}$. Imagine alleles labelled $A_1,A_2,\ldots$ in order of their creation by mutation. Mutation is a random process, so the $w_{i,j}$ should be regarded as random variables. It is not a priori apparent how to model the distribution $(w_{i,j})$, but it is natural to argue that $(w_{i,j})$ should be weakly exchangeable in the sense of (14.19), and then Theorem 14.21 can be brought to bear. See Kingman (1980), Section 2.5.

Another application is to the gene genealogy of haploid (i.e. single sex) populations. Suppose we sample K individuals from the current generation. For each $n \geq 0$ we can define an exchangeable random partition $R^K(n)$ of $\{1,\ldots,K\}$, where the components are the families of individuals with a common ancestor in the n^{th} previous generation. Letting the population size increase, K increase, and rescaling time, the process $(R^K(n): n \geq 0)$ approximates a certain continuous-time partition-valued process $(R(t): t \geq 0)$, the coalescent. See Kingman (1982a,b).

Finally, Dawson and Hochberg (1982) involve exchangeability ideas in a diffusion analysis of infinite-allele models more complicated than that described here.

20. Sampling processes and weak convergence

Given a finite sequence $x_1,\ldots,x_M$ of real constants, recall that the urn process is the sequence of random draws without replacement:

$$X_i = x_{\pi^*(i)} \quad \text{where } \pi^* \text{ is the uniform random permutation on } \{1,\ldots,M\}.$$

By the sampling process we mean the process of partial sums:

$$S_n = \sum_{i=1}^{n} X_i \ .$$

We shall often consider sampling processes drawn from normalized urns, where

$$\sum x_i = 0 \ , \quad \sum x_i^2 = 1 \ .$$

There is of course a vast literature on sampling: we shall merely mention a few results which relate to other ideas in exchangeability. We can distinguish two types of results: "universal" results true for all (normalized) urns, and "asymptotic" results as the individual elements of the urn become negligible. The main asymptotic result, Theorem 20.7, leads naturally to questions about weak convergence of general finite exchangeable sequences.

The most basic universal results are the elementary formulas for moments.

(20.1)
$$ES_n = n\mu/M$$
$$\mathrm{var}(S_n) = \frac{n(M-n)(\sigma^2 - \mu^2/M)}{M(M-1)} \quad \text{where} \quad \mu = \sum x_i, \quad \sigma^2 = \sum x_i^2 \ .$$

Restricting to normalized urns, we have also

(20.2)
$$ES_n^4 = \frac{n(M-n)}{M(M-1)} \sum x_i^4 + \frac{3n(n-1)(M-n)(M-n-1)}{M(M-1)(M-2)(M-3)}(1 - 2\sum x_i^4) \ .$$

A more abstract universal result involves rescaling the sampling process to make it a continuous-parameter process

$$S_t = S_{[Mt]} \ , \quad 0 \leq t \leq 1 \ .$$

Then we can think of S as a random element of the function space $D(0,1)$ with its usual topology (Billingsley (1968)). In this setting, we have

(20.3) Proposition. The family of processes S_t obtained from all normalized urns is a tight family in $D(0,1)$.

This is implicit in Billingsley (1968), (24.11) and Theorem 15.6. An alternative proof can be obtained from the tightness criteria in Aldous (1978).

In particular, Proposition 20.3 implies that there are bounds on the maxima of sampling processes which are uniform over the family of normalized urns. In other words, there exists a function ϕ with $\phi(\lambda) \to 0$ as $\lambda \to \infty$ and

$$(20.4) \qquad P\left(\max_n |S_n| > \lambda\right) \le \phi(\lambda) \; ; \quad \text{all normalized urns.}$$

I do not know what the best possible function ϕ is; here is a crude bound.

(20.5) Lemma. $\phi(\lambda) = 8/\lambda^2$ satisfies (20.4).

Proof. Let $F_k = \sigma(X_1,\dots,X_k)$, let $T = \min\{i: S_i > \lambda\}$. For $k \le m = [M/2]$,

$$E(S_m|F_k) = \frac{(M-m)}{(M-k)} S_k$$

and so

$$E(S_m|F_{T\wedge m}) = \frac{M-m}{M-T\wedge m} S_{T\wedge m} \ge \frac{1}{2}\lambda \quad \text{on} \quad \{T \le m\} ,$$

and so $E(S_m^2|F_{T\wedge m}) \ge \frac{1}{4}\lambda^2$ on $\{T \le m\}$. So

$$P(\max_{i\le m} S_i > \lambda) = P(T \le m) \le 4\lambda^{-2} ES_m^2 \le 2\lambda^{-2} \quad \text{using (20.2).}$$

Using the same inequality on $(-S_i)$, and using the symmetry $(S_i) \overset{\mathcal{D}}{=} (S_{M-i})$, we obtain the desired result.

Another type of universal result relates the sampling process S_n arising from sampling without replacement to the process $S_n^* = \sum_{i=1}^{n} X_i^*$ arising from random draws (X_i^*) from the same urn made with replacement, that is with (X_i^*) an i.i.d. sequence. Proposition 5.6 shows that when n is small compared to $M^{1/2}$ then the total variation distance between S_n and S_n^* is small, and this total variation result cannot be improved. However, there are results which compare the distributions of S_n and S_n^*. Given random variables U, V on R, say U is a <u>dilation</u> of V if there exist random variables $\hat{U}, \hat{V}$ such that

$$\hat{U} \overset{\mathcal{D}}{=} U, \quad \hat{V} \overset{\mathcal{D}}{=} V, \quad E(\hat{U}|\hat{V}) = \hat{V} .$$

This implies (and in fact is equivalent to--see e.g. Strassen (1965))

$$E\phi(U) \geq E\phi(V), \quad \text{all continuous convex } \phi: R \to R .$$

Informally, the distribution of U is "more spread out" than that of V. The next result extends the familiar result that $\mathrm{var}(S_n) \leq \mathrm{var}(S_n^*)$. See Kemperman (1973) for further related results.

(20.6) <u>Proposition</u>. <u>S_n^* is a dilation of S_n, for each urn and each $n \geq 1$.</u>

<u>Proof</u>. Without essential loss of generality, suppose the urn $\{x_1,\dots,x_M\}$ contains distinct elements. Fix n and let $(X_1,\dots,X_n)$ be draws without replacement. For distinct $(y_1,\dots,y_n) \subset \{x_i\}$ and not necessarily distinct $\{z_1,\dots,z_n\} \subset \{y_i\}$ define the conditional distribution

$$P(X_1^* = z_1,\dots,X_n^* = z_n | X_1 = y_1,\dots,X_n = y_n) = \frac{\binom{M}{n}}{n^M n! \binom{M-L}{n-L}} \quad \text{where } L = \#\{z_i\} .$$

Then it can be verified that the unconditional distribution of (X_i^*) is the distribution of sampling with replacement. And by symmetry, for $N(y) = \#\{i \leq n: X_i^* = y\}$ we have

$$E\big(N(y)\,|\,X_1 = y_1,\dots,X_n = y_n\big) = \begin{cases} 1, & y \in \{y_i\} \\ 0 & \text{otherwise} \end{cases}$$

So $E(S_n^*|X_1 = y_1,\dots,X_n = y_n) = \sum y_i$, which gives $E(S_n^*|S_n) = S_n$ as required.

We turn now to asymptotic results. For each $q \geq 1$ let $(x_1^q,\dots,x_{M_q}^q)$ be a normalized urn. Let S^q be the scaled sampling process

$$S_t^q = S_{[tM_q]}^q = \sum_{i=1}^{[tM_q]} X_i \;; \quad 0 \leq t \leq 1$$

and think of S^q as a random element of $D(0,1)$.

(20.7) Theorem. If

(20.8) $$\max_i |x_i^q| \to 0 \quad \text{as} \quad q \to \infty,$$

then $S^q \xrightarrow{\mathcal{D}} W^0$ in $D(0,1)$ as $q \to \infty$, where W^0 is Brownian bridge.

This is Theorem 24.1 of Billingsley (1968). Let us sketch a slightly different proof. By Proposition 20.3 we need only show that any subsequential weak limit process Z_t is Brownian bridge. By (20.8) Z has continuous paths, and clearly Z must have interchangeable increments, so by Theorem 10.11 we can take $Z_t = \alpha W_t^0$ for some $\alpha \geq 0$ independent of W^0. But using the moment formulas (20.1, 20.2) we can verify

$$E(S_t^q)^2 \to E(W_t^0)^2; \quad E(S_t^q)^4 \to E(W_t^0)^4 \quad \text{as} \quad q \to \infty .$$

So $E\alpha^2 = 1$ and $E\alpha^4 = 1$, which implies $\alpha = 1$, so Z is indeed Brownian bridge.

Theorem 20.7 leads to results for general finite exchangeable sequences. Consider a triangular array $(Z_{q,i}: 1 \leq i \leq M_q,\ 1 \leq q)$ such that

(20.9)(a) for each q the sequence $(Z_{q,1},\dots,Z_{q,M_q})$ is exchangeable

(b) $\max_i |Z_{q,i}| \xrightarrow[p]{} 0$ as $q \to \infty$.

As before let $S_t^q = \sum_{i=1}^{tM_q} Z_{n,i}$, let W^0 be Brownian bridge and let W be Brownian motion.

(20.10) Corollary. Suppose $(Z_{q,i})$ satisfies (20.9). Define random variables $\mu_q = \sum_i Z_{q,i}$, $\sigma_q^2 = \sum_i (Z_{q,i} - \mu_q/M_q)^2$. Let $X_t = \sigma W_t^0 + \mu t$, where (σ,μ) is independent of W^0. Then

(a) $S^q \xrightarrow{\mathcal{D}} X$ in $D(0,1)$ iff $(\mu_q,\sigma_q) \xrightarrow{\mathcal{D}} (\mu,\sigma)$.

In particular

(b) $S^q \xrightarrow{\mathcal{D}} W^0$ iff $\sum_i Z_{q,i} \xrightarrow[p]{} 0$ and $\sum_i Z_{q,i}^2 \xrightarrow[p]{} 1$,

(c) $S^q \xrightarrow{\mathcal{D}} W$ iff $\sum_i Z_{q,i} \xrightarrow{\mathcal{D}} N(0,1)$ and $\sum_i Z_{q,i}^2 \xrightarrow[p]{} 1$.

Sketch of proof. Suppose $(\hat{Z}_{q,i})$ satisfies (20.9) and

(20.11)
$$\sum_i \hat{Z}_{q,i} = 0; \quad \sum_i \hat{Z}_{q,i}^2 = 1 .$$

Then for each q the process $(\hat{Z}_{q,i})$ is a mixture of normalized urn rpocesses, and Theorem 20.7 implies $\hat{S}^q \xrightarrow[\mathcal{D}]{} W^0$. Now given $(Z_{q,i})$ define $\hat{Z}_{q,i} = (Z_{q,i} - \mu_q/M_q)/\sigma_q$. Then, in the case

(20.12)
$$\sigma > 0 \text{ a.s.}$$

the array $(\hat{Z}_{q,i})$ satisfies (20.9) and (20.11), and so $\hat{S}^q \xrightarrow{\mathcal{D}} W^0$. Furthermore, for events of the form $A_q = \{(\mu_q,\sigma_q) \in B_q\}$ with $\lim P(A_q) > 0$, the conditional distributions $(\mathcal{L}(\hat{Z}_{q,i} | A_q))$ also satisfy (20.9) and (20.11), so these conditional distributions also converge weakly to W^0. When $(\mu_q,\sigma_q) \xrightarrow{\mathcal{D}} (\mu,\sigma)$ this implies $(S^q,\mu_q,\sigma_q) \xrightarrow{\mathcal{D}} (W^0,\mu,\sigma)$. Since

(20.13) $$S_t^q = \sigma_q \hat{S}_t^q + \mu_q t$$

we finally obtain $S^q \xrightarrow{\mathcal{D}} X$.

In the case $\sigma = 0$ a.s., the maximal inequality (20.5) implies $\max_t |\sigma_q \hat{S}_t^q| \xrightarrow[p]{} 0$ as $q \to \infty$, and then (20.13) implies $S^q \xrightarrow{\mathcal{D}} X$. The general case can be reduced to these two cases, and this gives the "if" assertion of (a). The "only if" assertion follows from the "if" assertion and the facts

(i) tightness of (S^q) implies tightness of (μ_q, σ_q)

(ii) the distribution of $X_t = \sigma W_t^0 + \mu t$ determines the distribution of (σ, μ).

Assertions (b) and (c) are special cases of (a).

There is of course another result on weak convergence which is more celebrated than Theorem 20.7; it is interesting to note that this is essentially just a special case of Theorem 20.7.

(20.14) Corollary. *Let* (ξ_i) *be i.i.d. uniform on* $[0,1]$. *Let* $F^q(\omega,t)$, $0 \le t \le 1$, *be the empirical distribution function of* $(\xi_1(\omega),\ldots,\xi_q(\omega))$. *Let* $Y_t^q(\omega) = q^{1/2}(F_t^q(\omega) - t)$. *Then* $Y^q \xrightarrow{\mathcal{D}} W^0$ *as* $q \to \infty$.

Proof. Let $M_q = q^3$,

$$X_{q,i} = \#\{j: (i-1)/M_q < \xi_j \le i/M_q\} - q/M_q ,$$

so that

(20.15) $$\max_i X_{q,i} \xrightarrow[p]{} 1 \quad \text{as} \quad q \to \infty .$$

Then the array $Z_{q,i} = q^{-1/2} X_{q,i}$ satisfies the conditions of Corollary 20.10(b), so $S^q \xrightarrow{\mathcal{D}} W^0$. And (20.15) also gives $\sup_t |S_t^q - Y_t^q| \xrightarrow[p]{} 0$, so that $Y^q \xrightarrow{\mathcal{D}} W^0$.

Remark. The weak convergence form of Corollary 20.14 is nowadays rather obsolete, in that much more precise "strong approximation" results are known --see Csörgö and Revesz (1981). Similarly, there must be more precise strong approximation forms of Theorem 20.7, but I do not know any references in the literature.

The previous results do not really tackle the problem of when the sum of a finite exchangeable sequence can be approximated by a Normal distribution. To see that there can be no easy answer to this problem, recall that from an arbitrary sequence $X_1,\ldots,X_n$ we can produce an exchangeable sequence $Z_1,\ldots,Z_n$ by random permutation, and $\sum Z_i = \sum X_i$. Thus exchangeability imposes no restriction on the possible distribution of $\sum Z_i$.

One way to obtain weak convergence results for exchangeable sequences with little effort is to appeal to the general weak convergence results for martingales. A typical martingale result (Hall and Heyde (1980) Theorem 4.4; Helland (1982) Theorem 3.2) is the following.

(20.16) Theorem. *For each* $q \geq 1$ *let* $(X_{q,i}, F_{q,i}, 1 \leq i \leq M_q)$ *be a martingale difference sequence. If*

$$\sum_{i=1}^{[tM_q]} E(X_{q,i}^2 | F_{q,i-1}) \xrightarrow[p]{} t \; ; \quad \text{each } 0 \leq t \leq 1$$

$$\sum_{i=1}^{M_q} E(X_{q,i}^2 \, 1_{(|X_{q,i}|<\varepsilon)} | F_{q,i-1}) \xrightarrow[p]{} 0 \; ; \quad \text{each } \varepsilon > 0$$

then $S^q \xrightarrow{\mathcal{D}} W$, *where* $S_t^q = \sum_{i=1}^{[tM_q]} X_{q,i}$.

It is not difficult to specialize this to the exchangeable case and obtain the following sufficient conditions for convergence.

(20.17) Corollary. Let $(Z_{q,i})$ satisfy (20.9). Let $F_{q,j} = \sigma(Z_{q,i}: i \leq j)$, and let $Y_q = Z_{q,M_q} 1(|Z_{q,M_q}| \leq 1)$. Suppose that whenever $L_q/M_q \to t < 1$ we have

$$M_q E(Y_q | F_{q,L_q}) \underset{p}{\longrightarrow} 0$$

$$M_q E(Y_q^2 | F_{q,L_q}) \underset{p}{\longrightarrow} 1 .$$

Then $S^q \xrightarrow{\mathcal{D}} W$.

Of course, by using different forms of the martingale theorem we can obtain different sufficient conditions in Corollary 20.17. We remark also that, under moment conditions, conditions like those of Theorem 20.16 are necessary for convergence of martingales; but there can be no such necessity in Corollary 20.17 for exchangeable arrays, because for each q we could take $Z_{q,1}, \ldots, Z_{q,M_q}$ as a finite mixture of urn processes where the urns contained distinct values, and then $E(Z_{q,M_q} | F_{q,1}) = (S_1^q - Z_{q,1})/(M_q - 1)$.

The situation simplifies somewhat if we assume that the finite exchangeable sequences can be embedded into longer exchangeable sequences. The next result is due to Weber (1980).

(20.18) Proposition. Let $(Z_{q,i})$ satisfy (20.9). Suppose for each q that $(Z_{q,i}: 1 \leq i \leq M_q)$ extends to an exchangeable sequence $(Z_{q,i}: 1 \leq q \leq N_q)$, where $M_q/N_q \to 0$ as $q \to \infty$. Suppose also

$$EZ_{q,1} Z_{q,2} \to 0 \quad \text{as} \quad q \to \infty ;$$

$$\sum_{i=1}^{M_q} Z_{q,i}^2 \underset{p}{\longrightarrow} 1 \quad \text{as} \quad q \to \infty .$$

Then $S^q \xrightarrow{\mathcal{D}} W$.

However, this extendibility condition is very restrictive. The overall picture of central limit theorems for finite exchangeable sequences remains unsatisfactory. It turns out that Poisson limit theorems are somewhat more tractable: Eagleson (1982) gives a survey of such theorems for exchangeable sequences and partially exchangeable arrays. Of course there are many results which deal with special cases of finite exchangeable sequences, e.g. in the theory of random allocations--see Chow and Teicher (1978) Sections 3.2 and 9.2; Kolchin and Sevast'yanov (1978); Quine (1979); Johnson and Kotz (1977).

21. Other results and open problems

In this final section we give references to topics not previously mentioned and speculate on future lines of research.

(21.1) Non-standard versions of probability. Most mathematical probabilists (including the author) work within the "standard" model: (Ω,F,P), Radon measures and all that. There are however numerous alternative formalizations of probability theory, and exchangeability is such a basic concept that it makes sense within any version. Even if the reader is not interested in these alternative versions for their own sake, analysis within an alternative version can sometimes lead to results within the standard version.

Cylinder measures. On infinite-dimensional Hilbert space there are no non-trivial rotationally invariant σ-additive probability measures. Instead, one can consider cylinder measures, and obtain the analog of Schoenberg's Theorem 3.6: every rotationally invariant cylinder measure is a mixture of Gaussian cylinder measures (see e.g. Choquet (1969) Vol. 3). These measures play a fundamental role in certain analytical treatments of Brownian motion (Hida (1980)}.

Non-standard analysis. The treatment of partially exchangeable arrays by Hoover (1979, 1982) uses non-standard analysis; perhaps this approach would be useful for other exchangeability problems such as those of Section 16.

Finitely additive measures. At (13.27) we argued informally that the ability to pick an integer uniformly at random, in the finitely additive setting, led to plausibility arguments for several of our characterization results. Can finitely additive measures be employed to give rigorous results in the standard version more simply than the known proofs? See Dubins (1982) for one approach.

Quantum probability. Quantum theorists regard random variables as operators on Hilbert space. There is a version of de Finetti's theorem in this setting: see Accardi and Pistone (1982), Section 7, for an outline.

Function measures. Physicists define Brownian motion as the process on continuous functions $f \in C[0,1]$ with "density" $\psi(f)$ satisfying

$$\log \psi(f) = -\frac{1}{2}\int_0^1 (f'(x))^2 dx \ .$$

In a similar spirit, the Dirichlet(a) process (Section 10) on increasing functions f with $f(0) = 0$ and $f(1) = 1$ has "density"

$$\log \psi(f) = a\int_0^1 \log(f'(x))dx \ .$$

The form of these densities makes the interchangeable increments property plain. Does this approach yield other interesting processes with interchangeable increments?

(21.2) Other forms of invariance. There are several naturally-occurring classes of processes defined by invariance properties which we have not yet mentioned.

Self-similar processes. Fix $0 < H < \infty$. A process $(X_t: t \geq 0)$ satisfying

$$(X_t: t \geq 0) \overset{\mathcal{D}}{=} (X_{t_0+t} - X_{t_0}: t \geq 0); \quad \text{each } t_0 > 0,$$

$$(X_t: t \geq 0) \overset{\mathcal{D}}{=} (c^{-H} X_{ct}: t \geq 0); \quad \text{each } c > 0,$$

is self-similar with stationary increments. Taqqu (1982) surveys this area. See also O'Brien and Vewvaat (1983). As with stationary processes, this class seems too large for any explicit characterization of the ergodic processes to be possible; but perhaps some subclasses are more tractable?

Invariant point processes. Kallenberg (1976, 1976-81) and others have investigated point processes with invariance properties. Here is one natural problem, discussed in Kallenberg (1982c). Consider a group T of transformations of a space S, and suppose there is a unique (up to constant multiples) σ-finite T-invariant measure λ on S. Then a Poisson point process of intensity $c\lambda(\cdot)$ is T-invariant, and hence mixtures (over c) of such processes are also σ-invariant.

Problem. Under what circumstances are all T-invariant point processes mixtures of Poisson processes?

The most famous example, essentially due to Davidson (1974), concerns processes of random lines in $\mathbb{R}^2$ invariant under Euclidean motions of $\mathbb{R}^2$: if such a process has a.s. no parallel lines, then it is a mixture of Poisson line processes.

Further results on invariance of point processes are in Kallenberg (1982a), Section 13.

Sign-invariant processes. Berman (1965) studies processes with the property

$$(X_1,\ldots,X_n) \stackrel{\mathcal{D}}{=} (\varepsilon_1 X_1,\ldots,\varepsilon_n X_n); \quad \text{all} \quad \varepsilon_i \in \{-1,+1\},$$

and their continuous-time analogs.

(21.3) Special exchangeable sequences. One way to characterize exchangeable sequences which are mixtures of i.i.d. sequences with specific distributions is via sufficient statistics, as in Section 18. But several other types of characterizations exist. One type, natural from the Bayesian viewpoint, involves assumptions of regularity of posterior distributions: see e.g. Diaconis and Ylvisaker (1979); Zabell (1982). Another idea is to consider exchangeable renewal processes and impose regularity conditions on mean residual lifetimes; see Sigalotti (1982).

List of open problems posed in previous sections

1.11	1.12	2.29	3.14	5.13	7.21	9.12	11.26	12.20	12.29
13.10	13.14	13.17	13.28	15.6	15.10	15.15	15.20	15.30	Section 16

APPENDIX

Properties of Conditional Independence

Listed below are the properties of conditional independence we have used. Verification of these properties is a straightforward exercise in the use of properties of conditional expectations (with which we suppose the reader is familiar). Understanding these properties requires good intuition. An excellent elementary account is given by Pfeiffer (1979); a measure-theoretic account is given in Chow and Teicher (1978); more complete discussions and lists of properties are in Dawid (1979), Lauritzen (1982), Dohler (1980).

In what follows, sets are measurable, and functions ϕ are bounded measurable real-valued.

Say X, Y are conditionally independent given F if

(A1) $$P(X\in A,\ Y\in B|F) = P(X\in A|F)P(Y\in B|F);\quad \text{all } A, B.$$

Each of the following is equivalent to (A1):

(A2) $$E(\phi_1(X)\phi_2(Y)|F) = E(\phi_1(X)|F)E(\phi_2(Y)|F);\quad \text{all } \phi_1, \phi_2.$$

(A3) $$P(X\in A|F, Y) = P(X\in A|F);\quad \text{all } A.$$

(A4) $$E(\phi(X)|F, Y) = E(\phi(X)|F);\quad \text{all } \phi.$$

In these definitions we can replace a random variable X by a σ-field G, by replacing events $\{X\in A\}$ with events G $(G \in G)$, and replacing functions $\phi(X)$ with bounded random variables $V \in G$.

A family $\{X_i: i\in I\}$ is conditionally independent given F if the product formula in (A1) holds for each finite subset of (X_i).

Here are some other properties.

(A5) Suppose that for each $j \geq 1$, X_j and $\sigma(X_i: i > j)$ are conditionally independent given F. Then $(X_i: i \geq 1)$ are conditionally independent given F.

(A6) If X and X are conditionally independent given F then $X \in F$ a.s.

(A7) Suppose X and F are conditionally independent given G, and suppose X and G are conditionally independent given $H \subset G$. Then X and F are conditionally independent given H.

NOTATION

"Positive", "increasing" are used in the weak sense.

$\mathbb{R}$	set of real numbers
$\mathbb{Z}$; $\mathbb{N}$	set of all integers; set of natural numbers
$\#A$	cardinality of set A
1_A	indicator function/random variable: $1_A(x) = 1$ for $x \in A$; $= 0$ else.
$\delta_a(\cdot)$	probability measure degenerate at a: $\delta_a(A) = 1_A(a)$

$\mathcal{F} \subset \mathcal{G}$ a.s. means: for each $G \in \mathcal{G}$ there exists $F \in \mathcal{F}$ such that $P(F \Delta G) = 0$.

$\mathcal{F} = \mathcal{G}$ a.s. means $\mathcal{F} \subset \mathcal{G}$ a.s. and $\mathcal{G} \subset \mathcal{F}$ a.s.

$\mathcal{F}$ is trivial means $\mathcal{F} = \{\phi, \Omega\}$ a.s.

$\mathcal{L}(X)$	distribution of random variable X
$\sigma(X)$	σ-field generated by X
$\xrightarrow[p]{}$	convergence in probability
$\xrightarrow[\mathcal{D}]{}$	convergence in distribution
$N(\mu,\sigma^2)$	Normal distribution
$U(0,1)$	Uniform distribution on (0,1)

REFERENCES

Accardi, L. and Pistone, G. (1982). de Finetti's theorem, sufficiency, and Dobrushin's theory. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 125-156.

Ahmad, R. (1982). On the structure and application of restricted exchangeability. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.), North-Holland, Amsterdam, 157-164.

Aldous, D. J. (1977). Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete **40**, 59-82.

Aldous, D. J. (1978). Stopping times and tightness. Ann. Probability **6**, 335-340.

Aldous, D. J. (1979). Symmetry and independence for arrays of random variables. Preprint.

Aldous, D. J. (1981a). Representations for partially exchangeable arrays of random variables. J. Multivariate Anal. **11**, 581-598.

Aldous, D. J. (1981b). Subspaces of L^1, via random measures. Trans. Amer. Math. Soc. **267**, 445-463.

Aldous, D. J. (1982a). Partial exchangeability and $\bar{d}$-topologies. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 23-38.

Aldous, D. J. (1982b). On exchangeability and conditional independence. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 165-170.

Aldous, D. J. and Eagleson, G. K. (1978). On mixing and stability of limit theorems. Ann. Probability **6**, 325-331.

Aldous, D. J. and Pitman, J. (1979). On the zero-one law for exchangeable events. Ann. Probability 7, 704-723.

Arnaud, J.-P. (1980). Fonctions sphériques et fonctions définies positives sur l'arbre homogène. C. R. Acad. Sci. A 290, 99-101.

Bailey, R., Praeger, C., Speed, T. P. and Taylor, D. (1984). Analysis of Variance. Springer-Verlag, to appear.

Barbour, A. D. and Eagleson, G. K. (1983). Poisson approximations for some statistics based on exchangeable trials. Adv. Appl. Prob. 15, 585-600.

Berbee, H. (1981). On covering single points by randomly ordered intervals. Ann. Probability 9, 520-528.

Berkes, I. and Péter, E. (1983). Exchangeable r.v.'s and the subsequence principle. Preprint, Math. Inst. Hungarian Acad. Sci.

Berkes, I. and Rosenthal, H. P. (1983). Almost exchangeable sequences of random variables. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete (to appear).

Berman, S. (1965). Sign-invariant random variables and stochastic processes with sign-invariant increments. Trans. Amer. Math. Soc. 119, 216-243.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Blackwell, D. and Freedman, D. (1964). The tail σ-field of a Markov chain and a theorem of Orey. Ann. Math. Statist. 35, 1921-1925.

Blum, J. R. (1982). Exchangeability and quasi-exchangeability. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 171-176.

Blum, J., Chernoff, J., Rosenblatt, M., Teicher, H. (1958). Central limit theorems for interchangeable processes. Canad. J. Math. 10, 222-229.

Breiman, L. (1968). Probability. Addison-Wesley, Reading, Mass.

Brown, G. and Dooley, A. H. (1983). Ergodic measures are of weak product type. Preprint, School of Mathematics, University of New South Wales.

Brown, T. C. (1982). Poisson approximations and exchangeable random variables. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 177-184.

Bru, B., Heinich, H., Lootgitier, J.-C. (1981). Lois de grands nombres pour les variables échangeables. Comptes Rendus Acad. Sci. **293**, 485-488.

Bühlmann, H. (1958). Le problème "limite centrale" pour les variables aléatoires échangeables. C. R. Acad. Sci. Paris **246**, 534-536.

Bühlmann, H. (1960). Austauschbare stochastische Variabeln und ihre Grenzwertsätze. Univ. Calif. Publ. Statist. **3**, 1-35.

Carnal, E. (1980). Indépendence conditionelle permutable. Ann. Inst. Henri Poincaré B **16**, 39-47.

Cartier, P. (1973). Harmonic analysis on trees. In Harmonic Analysis on Homogenous Spaces. Amer. Math. Soc. (Proc. Sympos. Pure Math. **26**). Providence.

Chatterji, S. D. (1974). A subsequence principle in probability theory. Bull. Amer. Math. Soc. **80**, 495-497.

Choquet, G. (1969). Lectures on Analysis. Benjamin, Reading, Mass.

Chow, Y. S. and Teicher, H. (1978). Probability Theory. Springer-Verlag, New York.

Crisma, L. (1982). Quantitative analysis of exchangeability in alternative processes. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 207-216.

Csörgö, M. and Revesz, P. (1981). Strong Approximations in Probability and Statistics. Academic Press, New York.

Daboni, L. (1982). Exchangeability and completely monotone functions. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 39-45.

Dacunha-Castelle, D. (1974). Indiscernability and exchangeability in L^p spaces. Seminar on Random Series, Convex Sets and Geometry of Banach Spaces. Aarhus.

Dacunha-Castelle, D. (1982). A survey of exchangeable random variables in normed spaces. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 47-60.

Dacunha-Castelle, D. and Schreiber, M. (1974). Techniques probabilistes pour l'étude de problèmes d'isomorphismes entre espaces de Banach. Annales Inst. Henri Poincaré **10**, 229-277.

Davidson, R. (1974). Construction of line processes. In Stochastic Geometry (E. F. Harding and D. G. Kendall, eds.). Wiley, New York.

Dawid, A. P. (1977a). Invariant distributions and analysis of variance models. Biometrika **64**, 291-297.

Dawid, A. P. (1977b). Spherical matrix distributions and a multivariate model. J. Roy. Statist. Soc. B **39**, 254-261.

Dawid, A. P. (1978). Extendibility of spherical matrix distributions. J. Multivariate Anal. **8**, 567-572.

Dawid, A. P. (1979). Conditional independence in statistical theory. J. Roy. Statist. Soc. B **41**, 1-31.

Dawid, A. P. (1982). Intersubjective statistical models. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 217-232.

Dawson, D. A. and Hochberg, K. J. (1982). Wandering random measures in the Fleming-Viot model. Ann. Probability **10**, 554-580.

Dellacherie, C. and Meyer, P.-A. (1975, 1980). Probabilités et Potentiel, Ch. I-IV; Ch. V-VIII. Hermann, Paris.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. Synthese 36, 271-281.

Diaconis, P. and Freedman, D. (1980a). Finite exchangeable sequences. Ann. Probability 8, 745-764.

Diaconis, P. and Freedman, D. (1980b). de Finetti's theorem for Markov chains. Ann. Probability 8, 115-130.

Diaconis, P. and Freedman, D. (1980c). de Finetti's generalizations of exchangeability. In Studies in Inductive Logic and Probability II (R. C. Jeffrey, ed.). University of California Press, Berkeley.

Diaconis, P. and Freedman, D. (1981). On the statistics of vision: the Julesz conjecture. J. Math. Psychol. 24, 112-138.

Diaconis, P. and Freedman, D. (1982). Partial exchangeability and sufficiency. Preprint, Department of Statistics, Stanford University (to appear in Sankhyā).

Diaconis, P. and Ylvisaker, Y. (1979). Conjugate priors for exponential families. Ann. Statist. 7, 269-281.

Dohler, R. (1980). On the conditional independence of random events. Theory Probability Appl. 25, 628-634.

Doob, J. L. (1953). Stochastic Processes. Wiley, New York.

Dubins, L. E. (1982). Towards characterizing the set of ergodic probabilities. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 61-74.

Dubins, L. E. (1983). Some exchangeable probabilities which are singular with respect to all presentable probabilities. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 64, 1-6.

Dubins, L. E. and Freedman, D. (1979). Exchangeable processes need not be mixtures of independent identically distributed random variables. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete **48**, 115-132.

Dynkin, E. B. (1978). Sufficient statistics and extreme points. Ann. Probability **6**, 705-730.

Eagleson, G. K. (1979). A Poisson limit theorem for weakly exchangeable events, J. Appl. Prob. **16**, 794-802.

Eagleson, G. K. (1982). Weak limit theorems for exchangeable random variables. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 251-268.

Eagleson, G. K. and Weber, N. C. (1978). Limit theorems for weakly exchangeable arrays. Math. Proc. Cambridge Phil. Soc. **84**, 123-130.

Eaton, M. (1981). On the projections of isotropic distributions. Ann. Statist. **9**, 391-400.

Feller, W. (1971). An Introduction to Probability Theory and its Applications, vol. 2. Wiley, New York.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. **1**, 209-230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. Ann. Statist. **2**, 615-629.

Figiel, T. and Sucheston, L. (1976). An application of Ramsey sets in analysis. Advances in Math. **20**, 103-105.

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

de Finetti, B. (1938). Sur la condition d'equivalence partielle. Trans. in Studies in Inductive Logic and Probability II (R. C. Jeffrey, ed.). University of California Press, Berkeley.

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

de Finetti, B. (1974). *Theory of Probability*. Wiley, New York.

Freedman, D. (1962a). Mixtures of Markov processes. *Ann. Math. Statist.* **33**, 114-118.

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

Freedman, D. (1963). Invariants under mixing which generalize de Finetti's theorem: continuous time parameter. *Ann. Math. Statist.* **34**, 1194-1216.

Freedman, D. (1980). A mixture of independent identically distributed random variables need not admit a regular conditional probability given the exchangeable σ-field. *Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete* **51**, 239-248.

Freedman, D. and Diaconis, P. (1982). de Finetti's theorem for symmetric location families. *Ann. Statist.* **10**, 184-189.

Galambos, J. (1982). The role of exchangeability in the theory of order statistics. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 75-86.

Grigorenko, L. A. (1979). On the σ-algebra of symmetric events for a countable Markov chain. *Theor. Probability Appl.* **24**, 199-204.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.

Heath, D. and Sudderth, W. (1976). de Finetti's theorem for exchangeable random variables. *Amer. Statistician* **30**, 188-189.

Helland, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79-94.

Heller, A. (1965). On stochastic processes derived from Markov chains. *Ann. Math. Statist.* **36**, 1286-1291.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. Trans. Amer. Math. Soc. 80, 470-501.

Hida, T. (1980). Brownian Motion. Springer-Verlag, New York.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. Preprint.

Hoover, D. N. (1982). Row-column exchangeability and a generalized model for exchangeability. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 281-291.

Horowitz, J. (1972). Semilinear Markov processes, subordinators and renewal theory. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 24, 167-193.

Hu, Y.-S. (1979). A note on exchangeable events. J. Appl. Prob. 16, 662-664.

Ignatov, Ts. (1982). On a constant arising in the theory of symmetric groups and on Poisson-Dirichlet measures. Theory Prob. Appl. 27, 136-147.

Isaac, R. (1977). Generalized Hewitt-Savage theorems for strictly stationary processes. Proc. Amer. Math. Soc. 63, 313-316.

Jaynes, E. (1983). Some applications and extensions of the de Finetti representation theorem. In Bayesian Inference and Decision Techniques (P. K. Goel and A. Zellnor, eds.). North-Holland, Amsterdam.

Johnson, N. L. and Kotz, S. (1977). Urn Models and their Applications. Wiley, New York.

Kallenberg, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 27, 23-36.

Kallenberg, O. (1974). Path properties of processes with independent and interchangeable increments. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 28, 257-271.

Kallenberg, O. (1975). Infinitely divisible processes with interchangeable increments and random measures under convolution. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 32, 309-321.

Kallenberg, O. (1976). Random Measures. Akademie-Verlag, Berlin.

Kallenberg, O. (1976-81). On the structure of stationary flat processes, I-III. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 37, 157-174; 52, 127-147; 56, 239-253.

Kallenberg, O. (1982a). Characterizations and embedding properties in exchangeability. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete 60, 249-281.

Kallenberg, O. (1982b). A dynamical approach to exchangeability. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 87-96.

Kallenberg, O. (1982c). The stationary-invariance problem. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 293-295.

Kallenberg, O. (1982d). Conditioning in point processes. Preprint, University of Göteborg.

Kallenberg, O. (1983). The local time intensity of an exchangeable interval partition. Preprint, University of Göteborg.

Kelly, F. P. (1979). Reversibility and Stochastic Networks. Wiley, New York.

Kemperman, J. H. B. (1973). Moment problems for sampling without replacement I. Indagationes Math. 35, 149-164.

Kendall, D. G. (1967). On finite and infinite sequences of exchangeable events. Studia Scient. Math. Hung. 2, 319-327.

Kerov, S. V. and Vershik, A. M. (1982). Characters of infinite symmetric groups and probability properties of Robinson-Shenstead-Knuth's algorithm. Preprint.

Kimberling, C. H. (1973). Exchangeable events and completely monotonic sequences. Rocky Mtn. J. 3, 565-574.

Kindermann, R. and Snell, J. L. (1980). Markov Random Fields and their Applications. American Math. Soc., Providence.

Kingman, J. F. C. (1978a). Uses of exchangeability. Ann. Probability 6, 183-197.

Kingman, J. F. C. (1978b). The representation of partition structures. J. Lond. Math. Soc. 18, 374-380.

Kingman, J. F. C. (1978c). Random partitions in population genetics. Proc. Roy. Soc. 361, 1-20.

Kingman, J. F. C. (1980). Mathematics of Genetic Diversity. SIAM, Philadelphia.

Kingman, J. F. C. (1982a). Exchangeability and the evolution of large populations. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 97-112.

Kingman, J. F. C. (1982b). The coalescent. Stochastic Processes Appl. 13, 235-248.

Knuth, D. E. (1981). The Art of Computer Programming, Vol. 2, 2nd Ed. Addison-Wesley, Reading, Mass.

Kolchin, V. F. and Sevast'yanov, B. A. (1978). Random Allocations. Winston, New York.

Komlós, J. (1967). A generalisation of a problem of Steinhaus. Acta Math. Acad. Sci. Hungar. 18, 217-229.

Lauritzen, S. L. (1982). Statistical Models as Extremal Families. Aalborg University Press.

Letac, G. (1981a). Isotropy and sphericity: some characterizations of the normal distribution. Ann. Statist. 9, 408-417.

Letac, G. (1981b). Problèmes classiques de Probabilité sur un couple de Gelfand. In Analytic Methods in Probability Theory (D. Dugué et al., eds.). Springer Lecture Notes in Mathematics 861, New York.

Loève, M. (1960). Probability Theory. Van Nostrand, Princeton.

Lynch, J. (1982a). On a representation for row-column-exchangeable arrays. Technical Report 82, Department of Mathematics, University of South Carolina.

Lynch, J. (1982b). Canonical row-column-exchangeable arrays. Technical Report 84, Department of Mathematics, University of South Carolina.

Maitra, A. (1977). Integral representations of invariant measures. Trans. Amer. Math. Soc. **229**, 209-225.

Mandelbaum, A. and Taqqu, M. S. (1983). Invariance principle for symmetric statistics. Preprint, Operations Research, Cornell University.

Mansour, B. (1981). Le cube comme couple de Gelfand. Thesis, University of Toulouse.

Marshall, A. W. and Olkin, I. (1979). Inequalities: Theory of Majorization and its Application. Academic Press, New York.

Milgrom, P. R. and Weber, R. J. (1983). Exchangeable affiliated random variables. Preprint.

Moran, P. A. P. (1973). A central limit theorem for exchangeable variates with geometrical applications. J. Appl. Prob. **10**, 837-846.

O'Brien, G. L. and Vervaat, W. (1983). Self-similar processes with stationary increments generated by point processes. Preprint.

Olshen, R. A. (1971). The coincidence of measure algebras under an exchangeable probability. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete **18**, 153-158.

Olshen, R. A. (1973). A note on exchangeable sequences. Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete **28**, 317-321.

Ornstein, D. S. (1973). An application of ergodic theory to probability theory. Ann. Probability **1**, 43-65.

Palacios, J. (1982). The exchangeable σ-field of Markov chains. Ph.D. thesis, University of California, Berkeley.

Parthasarathy, K. R. (1967). Probability Measures on Metric Spaces. Academic Press, New York.

Pavlov, Y. L. (1981). Limit theorem for a characteristic of a random mapping. Theory Prob. Appl. **26**, 829-834.

Pfeiffer, P. E. (1979). Conditional Independence in Applied Probability. UMAP, Educational Development Center, Newton, Mass.

Phelps, R. R. (1966). Lectures on Choquet's Theorem. Van Nostrand, Princeton.

Pitman, J. (1978). An extension of de Finetti's theorem. Adv. Appl. Prob. **10**, 268-270.

Pittel, B. (1983). On the distributions related to transitive classes of random finite mappings. Ann. Probability **11**, 428-441.

Quine, M. P. (1978). A functional central limit theorem for a class of exchangeable sequences. Preprint, University of Sydney.

Quine, M. P. (1979). A functional central limit theorem for a generalized occupancy problem. Stochastic Proc. Appl. **9**, 109-115.

Rényi, A. (1963). On stable sequences of events. Sankhyā Ser. A **25**, 293-302.

Rényi, A. and Révész, P. (1963). A study of sequences of equivalent events as special stable sequences. Publ. Math. Debrecen. **10**, 319-325.

Ressel, P. (1983). de Finetti type theorems: an analytical approach. Preprint, Department of Statistics, Stanford University.

Ridler-Rowe, C. J. (1967). On two problems on exchangeable events. Studia Sci. Math. Hungar. 2, 415-418.

Ryll-Nardzewski, C. (1957). On stationary sequences of random variables and the de Finetti's equivalence. Colloquium Math. 4, 149-156.

Saunders, R. (1976). On joint exchangeability and conservative processes with stochastic rates. J. Appl. Prob. 13, 584-590.

Schoenberg, I. J. (1938). Metric spaces and positive definite functions. Trans. Amer. Math. Soc. 44, 522-536.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

Shaked, M. and Tong, Y. L. (1983). Some partial orderings of exchangeable random variables by positive dependence. Preprint, Mathematics Department, University of Arizona.

Shields, P. C. (1979). Stationary coding of processes. IEEE Trans. Information Theory 25, 283-291.

Sigalotti, L. (1982). On particular renewal processes interesting reliability theory. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 303-311.

Slud, E. V. (1978). A note on exchangeable sequences of events. Rocky Mtn. J. 8, 439-442.

Smith, A. M. F. (1981). On random sequences with centered spherical symmetry. J. Roy. Statist. Soc. Ser. B 43, 208-209.

Speed, T. P. (1982). Cumulants, k-statistics and their generalisations. Preprint, University of Western Australia.

Spitzer, F. (1975). Markov random fields on an infinite tree. Ann. Probability 3, 387-398.

Spizzichino, F. (1982). Extendibility of symmetric probability distributions and related bounds. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 313-320.

Stein, C. (1983). Approximate computation of expectations. Unpublished lecture notes.

Strassen, V. (1965). The existence of probability measures with given marginals. Ann. Math. Statist. 36, 423-439.

Surabian, C. G. (1979). A survey of exchangeability. M.Sc. thesis, Tufts University.

Taqqu, M. S. (1982). Self-similar processes and related ultraviolet and infrared catastrophes. In Random Fields (J. Fritz, ed.). North-Holland, Amsterdam.

Teicher, H. (1982). Renewal theory for interchangeable random variables. In Exchangeability in Probability and Statistics (G. Koch and F. Spizzichino, eds.). North-Holland, Amsterdam, 113-121.

Troutman, B. M. (1983). Weak convergence of the adjusted range of cumulative sums of exchangeable random variables. J. Appl. Prob. 20, 297-304.

Vershik, A. M. and Schmidt, A. A. (1977). Limit measures arising in the theory of symmetric groups I. Theory Prob. Appl. 22, 70-85.

Watterson, G. A. (1976). The stationary distribution of the infinite-many neutral alleles diffusion model. J. Appl. Prob. 13, 639-651.

Weber, N. C. (1980). A martingale approach to central limit theorems for exchangeable random variables. J. Appl. Prob. 17, 662-673.

Zabell, S. L. (1982). W. E. Johnson's "sufficientness" postulate. Ann. Statist. 10, 1091-1099.

Zachary, S. (1983). Countable state space Markov random fields and Markov chains on trees. Ann. Probability. to appear.