The Annals of Probability
1978, Vol. 6, No. 2, 183–197

THE 1977 WALD MEMORIAL LECTURES

USES OF EXCHANGEABILITY

BY J. F. C. KINGMAN
University of Oxford

The Wald Memorial Lectures delivered in Seattle in August 1977, and summarised here, ranged over a variety of applications in both pure and applied probability of the idea of exchangeability, and particularly of de Finetti's theorem. Particular emphasis was placed on two contrasting themes, some recent work of Aldous on the subsequence principle, and consequences of de Finetti's theorem for certain problems in population genetics.

1. Combinatorial arguments in probability theory. The lectures on which this paper is based were not intended to be a systematic survey of the theory and applications of exchangeability; at most they were designed to illustrate the power and elegance of the concept in a variety of contexts. Accordingly the present account is a personal selection of topics which I have found interesting and illuminating.

The use of combinatorial arguments for probabilistic problems is of course as old as probability theory itself. Indeed, so long as the theory was concerned exclusively with the equiprobable case, all its results were necessarily combinatorial in character. In this century, however, more subtle use has been made by many authors, perhaps most effectively by Sparre Andersen and his colleagues (as for instance several papers in the collection [4] bear witness).

A typical argument might run as follows. Suppose that

$$Z = \Phi(X_1, X_2, \cdots, X_n) \tag{1.1}$$

is a function of independent, identically distributed random variables X_r, such that $\mathbf{E}(Z)$ exists. For any permutation π of $\{1, 2, \cdots, n\}$, Z has the same distribution as

$$Z_\pi = \Phi(X_{\pi 1}, X_{\pi 2}, \cdots, X_{\pi n}) , \tag{1.2}$$

and hence

$$\mathbf{E}(Z) = (n!)^{-1} \sum_\pi \mathbf{E}(Z_\pi) = \mathbf{E}\{\Psi(X_1, X_2, \cdots, X_n)\} , \tag{1.3}$$

Received May 31, 1977.
AMS 1970 *subject classification.* Primary 60G99.
Key words and phrases. Exchangeability, combinatorial argument, subsequences, population genetics.

where

$$\Psi(x_1, x_2, \cdots, x_n) = (n!)^{-1} \sum_\pi \Phi(x_1, x_2, \cdots, x_n) \tag{1.4}$$

is the symmetrised form of Φ, the sums extending over all $n!$ permutations π.

It may well happen that Ψ is a simpler function than Φ, so that (1.3) facilitates the calculation of $\mathrm{E}(Z)$. For example, if Ψ is identically zero, the identity

$$\mathrm{E}\{\Phi(X_1, X_2, \cdots, X_n)\} = 0 \tag{1.5}$$

is valid for all independent and identically distributed X_r. That such simple arguments can lead to very important results may be seen by considering Spitzer's identity, of which the original proof [44] was exactly of this type.

The argument can also be inverted, since [29] it is simple to show that, if (1.5) holds whatever the distribution of the X_r (or less stringently, for every distribution whose support contains at least n points), then necessarily $\Psi = 0$ identically. And then the argument in the original direction shows that (1.5) holds even if the X_r are not independent, *so long as the joint distribution of* $(X_{\pi 1}, X_{\pi 2}, \cdots, X_{\pi n})$ *is the same as that of* $(X_1, X_2, \cdots, X_n)$, *for every permutation* π.

The italicised clause is exactly what is meant by saying that $(X_1, X_2, \cdots, X_n)$ is *exchangeable* (or permutable or interchangeable or symmetrically dependent), so that the combinatorial argument leads directly to this concept. Every result which can be thrown into the form (1.5) and which is true under the assumptions of independence and common distribution is necessarily true under the much weaker assumption of exchangeability. In this sense, finite exchangeable sequences partake of some of the properties of independent, identically distributed sequences.

The technique is not confined to equalities. Suppose for example that (X_1, X_2) is exchangeable (which, of course, simply means that it has the same bivariate distribution as (X_2, X_1)), and that f and g are two increasing functions. Then

$$[f(x_1) - f(x_2)][g(x_1) - g(x_2)] \geqq 0$$

for all x_1, x_2 and so

$$\mathrm{E}\{[f(X_1) - f(X_2)][g(X_1) - g(X_2)]\} \geqq 0 \, .$$

Therefore, if the expectations exist, exchangeability implies that

$$\mathrm{E}\{f(X_1)g(X_1)\} \geqq \mathrm{E}\{f(X_1)g(X_2)\} \, . \tag{1.6}$$

This simple inequality is not quite trivial, since when X_1 and X_2 are independent it reduces to the "other" Tchebychev inequality [47]

$$\mathrm{E}\{f(X)g(X)\} \geqq \mathrm{E}\{f(X)\}\mathrm{E}\{g(X)\} \, . \tag{1.7}$$

Again, there are applications in which it is not necessary to consider all permutations π. If for any subset A of the group of permutations of $\{1, 2, \cdots, n\}$, Φ satisfies

$$\sum_{\pi \in A} \Phi(x_1, x_2, \cdots, x_n) = 0 \tag{1.8}$$

for all $(x_1, x_2, \cdots, x_n)$, then (1.5) holds for all exchangeable $(X_1, X_2, \cdots, X_n)$. Takács [45], [46] has shown, for example, that the "ballot theorems" of Bertrand and André follow from identities of the form (1.8) in which A consists of all cyclic permutations.

All these arguments apply to *finite* exchangeable sequences. For infinite sequences, much more can be said, but it is possible to feel that the glamour of the essentially infinitary results has been allowed to obscure the power of the purely finite.

2. De Finetti's theorem. The central result of exchangeability theory is the theorem proved in successively greater generality by de Finetti in [18] and [19], and by Hewitt and Savage [22]. It can be regarded from a number of different points of view; a functional analyst for example might see it as an integral representation theorem in the spirit of Choquet for symmetric measures on product spaces. For a probabilist it is perhaps most illuminating to follow a simple martingale argument (cf. Loève [38]).

An infinite sequence $\mathbf{X} = (X_1, X_2, \cdots)$ of random variables is said to be *exchangeable* if $(X_1, X_2, \cdots, X_n)$ is exchangeable for each $n \geqq 2$. Describe a random variable as n-symmetric if it is a function of $\mathbf{X}$ which is unchanged if the first n variables are permuted in any way: for example,

$$X_1X_2X_3 + X_4X_6$$

is 3-symmetric but not 4-symmetric. Let $\mathscr{F}_n$ be the smallest σ-algebra with respect to which all the n-symmetric random variables are measurable, and note that $\mathscr{F}_n \supseteqq \mathscr{F}_{n+1}$.

If f is a measurable function for which

$$\mathbf{E}|f(X_1)| < \infty ,$$

and if $Y = g(\mathbf{X})$ is a bounded n-symmetric random variable, the exchangeability of $\mathbf{X}$ implies that, for $1 \leqq j \leqq n$,

$$\begin{aligned}\mathbf{E}\{f(X_j)g(\mathbf{X})\} &= \mathbf{E}\{f(X_1)g(X_j, X_2, \cdots, X_{j-1}, X_1, X_{j+1}, \cdots)\} \\ &= \mathbf{E}\{f(X_1)g(\mathbf{X})\} ,\end{aligned}$$

so that

$$\mathbf{E}\{n^{-1} \textstyle\sum_{j=1}^n f(X_j)Y\} = \mathbf{E}\{f(X_1)Y\} .$$

For any $A \in \mathscr{F}_n$, Y may be taken as the indicator of A, so that

$$\textstyle\int_A n^{-1} \sum_{j=1}^n f(X_j)\, d\mathbf{P} = \int_A f(X_1)\, d\mathbf{P} \qquad (A \in \mathscr{F}_n) .$$

The integrand on the left is n-symmetric and therefore $\mathscr{F}_n$-measurable, and therefore

$$n^{-1} \textstyle\sum_{j=1}^n f(X_j) = \mathbf{E}\{f(X_1) \mid \mathscr{F}_n\} . \tag{2.1}$$

An elementary martingale convergence theorem ([13], Theorem 4.3) or direct computation ([38], Section 27.2) now shows that

$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{j=1}^n f(X_j) = \mathbf{E}\{f(X_1) \mid \mathscr{F}_\infty\} \tag{2.2}$$

almost surely, where

$$\mathscr{F}_\infty = \bigcap_{n=1}^\infty \mathscr{F}_n\,. \tag{2.3}$$

The existence of the limit (2.2) is an expression of the *strong law of large numbers* for exchangeable sequences. In particular, if f is the indicator of the interval $(-\infty, x]$, then with probability one,

$$\lim_{n\to\infty} n^{-1}\#\{j \leqq n;\ X_j \leqq x\} = F(x)\,, \tag{2.4}$$

where

$$F(x) = \mathbf{P}\{X_1 \leqq x \mid \mathscr{F}_\infty\} \tag{2.5}$$

is a random distribution function.

Now suppose that f is a bounded measurable function on R^k. The argument leading to (2.1) generalises at once to give, for $n \geqq k$,

$$\{n(n-1)\cdots(n-k+1)\}^{-1}\sum f(X_{j_1}, X_{j_2}, \cdots, X_{j_k}) = \mathbf{E}\{f(X_1, X_2, \cdots, X_k) \mid \mathscr{F}_n\}\,,$$

where the sum extends over distinct $j_1, j_2, \cdots, j_k \leqq n$. The martingale theorem then shows that

$$\begin{aligned}\mathbf{E}\{f(X_1, X_2, \cdots, X_k) \mid \mathscr{F}_\infty\} &= \lim_{n\to\infty}\{n(n-1)\cdots(n-k+1)\}^{-1}\sum f(X_{j_1}, X_{j_2}, \cdots, X_{j_k}) \\ &= \lim_{n\to\infty} n^{-k}\sum_{j_1=1}^n\sum_{j_2=1}^n\cdots\sum_{j_k=1}^n f(X_{j_1}, X_{j_2}, \cdots, X_{j_k})\,,\end{aligned}$$

since the contribution from terms with coincidences among the j_r is of order n^{k-1}. In particular, if

$$f(y_1, y_2, \cdots, y_k) = f_1(y_1)\,f_2(y_2)\cdots f_k(y_k)\,,$$

and f_r is the indicator of $(-\infty, x_r]$, (2.4) implies that

$$\mathbf{P}\{X_1 \leqq x_1, X_2 \leqq x_2, \cdots, X_k \leqq x_k \mid \mathscr{F}_\infty\} = F(x_1)\,F(x_2)\cdots F(x_k)\,. \tag{2.6}$$

This formula encapsulates de Finetti's theorem: *there is a σ-algebra conditional on which the X_j are independent and identically distributed.* Note that $F(x)$ is for each x an $\mathscr{F}_\infty$-measurable random variable. If $\mathscr{G}$ is any sub-σ-algebra of $\mathscr{F}_\infty$ with respect to which each $F(x)$ is measurable, then taking expectations of (2.6) conditional on $\mathscr{G}$ yields the same expression with $\mathscr{F}_\infty$ replaced by $\mathscr{G}$. In particular, if $\mathscr{G}$ is the σ-algebra generated by the variables $F(x)$, then in an obvious notation

$$\mathbf{P}\{X_1 \leqq x_1, X_2 \leqq x_2, \cdots, X_k \leqq x_k \mid F\} = F(x_1)\,F(x_2)\cdots F(x_k)\,. \tag{2.7}$$

Thus we have a recipe for constructing the most general exchangeable sequence. First construct a sequence of independent random variables having the same distribution function F. Then allow F itself to vary randomly. The randomisation destroys the independence while preserving exchangeability, and every infinite exchangeable sequence can be constructed in this way.

It is very important to stress that de Finetti's theorem is about infinite sequences, and cannot be applied to finite exchangeable sequences. There is a finite analogue, but it lies less deep and is much less useful (see Kendall [28]). It is worth making the distinction vivid by a simple argument. Let $(X_1, X_2, \cdots, X_n)$ be exchangeable, and suppose that X_1 has finite mean μ and (nonzero) variance σ^2; let ρ be the correlation coefficient of X_1 and X_2. Then the variance of $\sum_{j=1}^n X_j$ is

$$\sum_{j=1}^n \sigma^2 + \sum_{i \neq j} \rho\sigma^2 = n\sigma^2\{1 + (n-1)\rho\},$$

and since this is positive we have

$$\rho \geqq -(n-1)^{-1}. \tag{2.8}$$

In particular, the members of an infinite exchangeable sequence are positively correlated, as may also be proved directly from (2.7). However, in a finite exchangeable sequence negative correlations can arise: for example the lower bound (2.8) is attained if the X_r have a symmetric multinomial distribution. When $\rho < 0$ the X_r can never be conditionally independent and identically distributed.

3. Some complements to de Finetti's theorem. It would be surprising if such a beautiful result had not attracted a variety of generalizations, extensions and analogues. Again, I make no attempt at completeness, but will mention three different theoretical developments to illustrate the possibilities.

(a) *Conditioning on a single random variable.* It is perhaps less than satisfying that the conditioning in (2.6) and (2.7) should be on (respectively) an abstract σ-algebra and on a whole function, although of course both $\mathscr{F}_\infty$ and F have significant meanings in relation to the exchangeable sequence $\mathbf{X}$. It is therefore interesting that (2.7) can be replaced by an expression of the form

$$\mathbf{P}\{X_1 \leqq x_1, X_2 \leqq x_2, \cdots, X_k \leqq x_k \mid \zeta\} = F_\zeta(x_1) F_\zeta(x_2) \cdots F_\zeta(x_r), \tag{3.1}$$

where ζ is a single random variable, and the distribution function F_ζ depends now on ζ. In fact, (2.6), (2.7) and (3.1) are all really equivalent, since the σ-algebra generated by ζ, and that generated by F, are both identical with $\mathscr{F}_\infty$ up to events of zero probability (Olshen [41]).

(b) *Spherical symmetry.* A theorem usually ascribed to Maxwell states that, if independent random variables $X_1, X_2, \cdots, X_n$ are such that the joint distribution of $(X_1, X_2, \cdots, X_n)$ has spherical symmetry (in the sense that it is invariant under all n-dimensional rotations) then the X_r are normally distributed with zero mean and the same variance (regarding the degenerate distributions as normal with zero variance). In view of this and de Finetti's theorem, it is not very remarkable that something can be said about infinite sequences with spherical symmetry, even when independence is not assumed.

Thus suppose that $X_1, X_2, \cdots$ are random variables with the property that, for every $n \geqq 1$, the joint distribution of $(X_1, X_2, \cdots, X_n)$ is spherically symmetric.

Then it can be proved that

$$V = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} X_j^2 \tag{3.2}$$

exists almost surely, and that, conditional on V, the X_j are independent and normally distributed, with zero mean and variance V. (The result appears to be due to Freedman [21]; an incomplete proof may be found in [26] and a short one in [30].)

More interesting problems arise with matrices whose distributions are invariant under multiplication by orthogonal matrices. The picture here is still far from clear, but some valuable brushwork has been carried out by Wachter [48] and Dawid [12].

(c) *Selection.* Every exchangeable sequence $\mathbf{X}$ has the *selection property*, that for any integers $1 \leqq j_1 < j_2 < \cdots < j_k$ the joint distribution of $(X_{j_1}, X_{j_2}, \cdots, X_{j_k})$ is the same as that of $(X_1, X_2, \cdots, X_k)$. The converse is false for finite sequences, because for a sequence of N variables the selection property only gives information about the $(N-1)$-dimensional marginals. Rather surprisingly, the converse is valid for infinite sequences. It has been established by Dacunha-Castelle [9] in a more general setting; we sketch here a simple proof.

Suppose that the infinite sequence $\mathbf{X}$ has the selection property. Taking $j_r = r+1$ shows in particular that $\mathbf{X}$ is stationary, and the Birkhoff ergodic theorem then shows that (2.4) holds for some random distribution function F. We now use an easily proved fact of elementary real analysis: if for each $r \leqq k$ the bounded sequence $(a_r(j); j \geqq 1)$ has Cesàro limit

$$\alpha_r = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} a_r(j) ,$$

then

$$\binom{n}{k}^{-1} \sum_{j_1<j_2<\cdots<j_k\leqq n} a_1(j_1)\, a_2(j_2) \cdots a_k(j_k) \to \alpha_1 \alpha_2 \cdots \alpha_k \tag{3.3}$$

as $n \to \infty$. Applying this result with

$$a_r(j) = f_r(X_j) ,$$

and f_r the indicator of $(-\infty, x_j]$, we have that

$$\binom{n}{k}^{-1} \#\{j_1 < j_2 < \cdots < j_k \leqq n;\ X_{j_1} \leqq x_1, X_{j_2} \leqq x_2, \cdots, X_{j_k} \leqq x_k\} \to F(x_1)\, F(x_2) \cdots F(x_k) . \tag{3.4}$$

By the selection property, the expectation of the left-hand side is

$$\mathbf{P}\{X_1 \leqq x_1, X_2 \leqq x_2, \cdots, X_k \leqq x_k\}$$

and bounded convergence shows that this is equal to

$$\mathbf{E}\{F(x_1)\, F(x_2) \cdots F(x_k)\} .$$

The symmetry of this last expression shows that $\mathbf{X}$ is exchangeable.

Thus an apparently weaker property than exchangeability, the selection property, suffices for the conclusion of de Finetti's theorem.

4. The subsequence principle. In 1967 Komlós established a conjecture of Steinhaus by proving the remarkable theorem [36] that, if $\mathbf{X} = (X_1, X_2, \cdots)$ is any sequence of random variables with $\mathbf{E}|X_n|$ bounded, then there exists a nonrandom sequence $1 \leqq n_1 < n_2 < n_3 < \cdots$ of integers such that the subsequence

$$X_r^* = X_{n_r} \tag{4.1}$$

has the property that the limit

$$\lim_{m\to\infty} m^{-1} \sum_{r=1}^{m} X_r^* \tag{4.2}$$

exists almost surely. In other words, the strong law of large numbers applies to some subsequence of every L_1-bounded sequence.

Inspired by this result Chatterji, in a long series of papers exemplified by and listed in [7] and [8], established corresponding analogues of other classical properties of independent sequences, and propounded his subsequence principle. This asserts that every limit property enjoyed by all independent, identically distributed sequences satisfying some moment condition is shared by some (nonrandomly chosen) subsequence of every sequence which satisfies the moment condition uniformly. Chatterji devised a separate (and usually very complex) proof of each instance of the principle, and it became an obvious challenge to establish a general result to embrace all the known special cases.

Any such general approach inevitably involves exchangeability. Suppose that a property $\mathscr{P}$ determined by joint distributions is enjoyed by some subsequence of every random sequence. Then it is enjoyed by some subsequence of every exchangeable sequence, and by the selection property this subsequence has the same stochastic structure as the original sequence. Thus $\mathscr{P}$ is enjoyed by every exchangeable sequence. On the other hand, de Finetti's theorem shows that any property shared by all independent and identically distributed sequences has an analogue for exchangeable sequencs; for instance the classical strong law implies the existence of the limit (4.2) whenever (X_r^*) is exchangeable and $\mathbf{E}|X_1^*| < \infty$.

The tension between these two facts makes it clear that the first step towards a precise version of the Chatterji principle is to reformulate it as a conjecture: *every limit property enjoyed by all exchangeable sequences is shared by some subsequence of every tight sequence.* (The adjective "tight," which means that

$$\lim_{\lambda\to\infty} \sup_n \mathbf{P}\{|X_n| \geqq \lambda\} = 0 , \tag{4.3}$$

is necessary to stop probability disappearing to infinity.)

The conjecture would for instance be true if every tight sequence contained an exchangeable subsequence. This is rather obviously false, but it is approximately true. More precisely, it can be shown that every tight sequence (X_n) contains a subsequence (4.1) which is *asymptotically exchangeable* in the sense that the joint distribution of the sequence

$$(X_r^*, X_{r+1}^*, X_{r+2}^*, \cdots) \tag{4.4}$$

converges as $r \to \infty$ to the distribution of some exchangeable sequence.

This result seems to have arisen independently from several sources; it is essentially contained in [9], in [17] and in unpublished work of my own. The simplest proofs derive it from the celebrated combinatorial theorem of Ramsey [42]. It is natural to hope that it might be powerful enough to establish the above conjecture, and I was optimistic enough to speculate (having in mind Theorem 3.1.1 of Skorokhod [43]) that, for any asymptotically exchangeable sequence (X_n), there is an exchangeable sequence (Y_n) on the same (or perhaps a larger) probability space, such that

$$X_n - Y_n \to 0 \tag{4.5}$$

in probability as $n \to \infty$.

This line of speculation was brought to an abrupt halt by a decisive counter-example of Aldous [1]. Let ε_n $(n \geqq 1)$ be independent with

$$\mathbf{P}(\varepsilon_n = 0) = \mathbf{P}(\varepsilon_n = 1) = \tfrac{1}{2} ,$$

and let

$$\xi = \textstyle\sum_{n=1}^{\infty} \varepsilon_n 2^{-n} , \tag{4.6}$$

so that ξ is uniformly distributed on (0, 1). It is not difficult to see that the sequence

$$X_n' = \xi + \varepsilon_n \tag{4.7}$$

is asymptotically exchangeable, the limiting distribution being that of the exchangeable sequence

$$Y_n' = \eta + \varepsilon_n , \tag{4.8}$$

where η is uniformly distributed on (0. 1) but is independent of the ε_n.

Now suppose, if possible, that, on some probability space, there are sequences (X_n), (Y_n) such that the joint distribution of (X_n) is the same as that of (X_n'), that of (Y_n) the same as that of (Y_n'), and (4.5) holds. By (4.7)

$$\mathbf{P}(X_m' - X_n' \in \{-1, 0, 1\}) = 1 ,$$

so that

$$\mathbf{P}(X_m - X_n \in \{-1, 0, 1\}) = 1 ,$$

and thus the fractional part of X_n does not change with n. By (4.8) the same is true of Y_n, and by (4.5) these two fractional parts are the same. Hence there is a random variable $\zeta \in (0, 1)$ and random variables ε_n', ε_n'' in $\{0, 1\}$ such that

$$X_n = \zeta + \varepsilon_n' , \qquad Y_n = \zeta + \varepsilon_n'' .$$

From the known distributions of the sequences (X_n) and (Y_n) it follows that ε_n' is a function of ζ (the nth digit in its binary expansion), while ε_n'' is independent of ζ. Therefore

$$\mathbf{P}\{|X_n - Y_n| = 1 \mid \zeta\} = \mathbf{P}\{\varepsilon_n'' \neq \varepsilon_n' \mid \zeta\} = \tfrac{1}{2}$$

and thus

$$\mathbf{P}\{|X_n - Y_n| = 1\} = \tfrac{1}{2} ,$$

contradicting (4.5).

The remarkable fact is that, despite this example, the subsequence conjecture can be made precise, and it is then true, as has been shown in a brilliant display of sustained argument by Aldous [1], [2], [3]. For the details the reader must turn to his own account, to the concept of a limit statute and a very general theorem subsuming all the known cases of the subsequence principle. It is impossible here to do justice to a major achievement in probability theory, and I must confine myself to paying tribute to it.

5. Exchangeability in population genetics. As a sharp contrast to the austere elegance of the subsequence problem, I now turn to more earthy applications of exchangeability theory. From time to time (as for example in [5] and [28]) the relevance of the concept to the description of variation in biological populations has been noted, a relevance which depends on the fact that in many models the members of the population are supposed to be indistinguishable.

More recently the theory has found rather deeper application to problems in population genetics. One such is expounded in detail in [32], but rather than recounting this work, the possibilities will be illustrated with reference to a rather different model which may prove to be appropriate for some genetical situations [34].

Consider a population of a fixed size N evolving in nonoverlapping generations. In genetical terms, it will be taken as haploid and interest will centre on the gene at a particular locus. From this point of view, each individual can be described by an element of a set S of possible alleles at that locus.

A generation G_t ($t = 0, 1, 2, \cdots$) therefore consists of N elements (not necessarily distinct) of a fixed set S, which in practice is finite but large and for modelling purposes might be allowed to be countably infinite. Each member of G_t then produces daughters which are also identified with elements of S, and which make up the next generation G_{t+1}. It is assumed that the probability that the numbers of daughters born to the members of G_t are respectively $d_1, d_2, \cdots, d_N$ is of the multinomial form

$$N!/N^N d_1!\, d_2! \cdots d_N! \quad (d_1 + d_2 + \cdots + d_N = N)\,, \tag{5.1}$$

which means that the reproduction is selectively neutral. However, to allow the possibility of mutation, we assume that the daughter of a member of G_t at i in S is at j in S with a given probability $p(i, j)$, the $p(i, j)$ satisfying

$$p(i, j) \geqq 0\,, \qquad \textstyle\sum_{j \in S} p(i, j) = 1\,, \tag{5.2}$$

and therefore constituting a stochastic matrix

$$P = (p(i, j);\, i, j \in S)\,.$$

We therefore have a discrete-generation model embodying genetic drift (the random fluctuations implied by (5.1)) and mutation, but not selection (or the manifold complications of diploid, multi-locus behaviour). The sequence (G_t; $t = 0, 1, 2, \cdots$) is a Markov chain. To examine its properties it is convenient to label the members of G_t in random order as

$$X_r(t) \quad (r = 1, 2, \cdots, N) . \tag{5.3}$$

Thus $X_r(t) \in S$ is the rth element to be drawn from G_t in random sampling without replacement; G_t is described by the finite sequence (5.3), which by definition is exchangeable.

It is now easy to see that, conditional on G_t, the random variables $X_r(t+1)$ are independent, with

$$\mathbf{P}\{X_r(t+1) = j \mid G_t\} = N^{-1} \textstyle\sum_{\alpha=1}^{N} p(X_\alpha(t), j) . \tag{5.4}$$

(This utilises the special nature of the multinomial distribution (5.1).) In particular, (5.4) implies that

$$\pi_t(j) = \mathbf{P}\{X_1(t) = j\} \tag{5.5}$$

satisfies

$$\pi_{t+1}(j) = N^{-1} \textstyle\sum_{\alpha=1}^{N} \mathbf{E}\{p(X_\alpha(t), j)\} = \mathbf{E}\{p(X_1(t), j)\} = \sum_{i \in S} \pi_t(i) p(i, j) .$$

Now suppose that (and it is at this point that we part company from [32]) P is irreducible, aperiodic and positive recurrent, with stationary probability measure $(\pi(j); j \in S)$. Then, whatever G_0,

$$\pi_t(j) \rightarrow \pi(j) \tag{5.6}$$

as $t \rightarrow \infty$, for all j.

From (5.4) again, the joint probability

$$\pi_t(j_1, j_2) = \mathbf{P}\{X_1(t) = j_1, X_2(t) = j_2\} \tag{5.7}$$

satisfies

$$\pi_{t+1}(j_1, j_2) = N^{-2} \textstyle\sum_{\alpha,\beta=1}^{N} \mathbf{E}\{p(X_\alpha(t), j_1) p(X_\beta(t), j_2)\} .$$

By exchangeability, the summand equals

$$\mathbf{E}\{p(X_1(t), j_1) p(X_2(t), j_2)\} = \textstyle\sum_{i_1, i_2 \in S} \pi_t(i_1, i_2) p(i_1, j_1) p(i_2, j_2)$$

if $\alpha \neq \beta$, and

$$\mathbf{E}\{p(X_1(t), j_1) p(X_1(t), j_2)\} = \textstyle\sum_{i \in S} \pi_t(i) p(i, j_1) p(i, j_2)$$

if $\alpha = \beta$, so that

$$\begin{aligned} \pi_{t+1}(j_1, j_2) = {} & (1 - N^{-1}) \textstyle\sum_{i_1, i_2 \in S} \pi_t(i_1, i_2) p(i_1, j_1) p(i_2, j_2) \\ & + N^{-1} \textstyle\sum_{i \in S} \pi_t(i) p(i, j_1) p(i, j_2) . \end{aligned} \tag{5.8}$$

Regarding the $\pi_t(i)$ as known, this is a recurrence relation on t for the quantities (5.7), from which it is easy to show that

$$\pi(j_1, j_2) = \lim_{t \rightarrow \infty} \pi_t(j_1, j_2) \tag{5.9}$$

exists and is independent of G_0, being determined by the equations

$$\begin{aligned} \pi(j_1, j_2) = {} & (1 - N^{-1}) \textstyle\sum_{i_1, i_2 \in S} \pi(i_1, i_2) p(i_1, j_1) p(i_2, j_2) \\ & + N^{-1} \textstyle\sum_{i \in S} \pi(i) p(i, j_1) p(i, j_2) . \end{aligned} \tag{5.10}$$

A similar analysis may now be carried out, recursively on n, for the probabilities

$$\pi_t(j_1, j_2, \cdots, j_n) = \mathbf{P}\{X_1(t) = j, X_2(t) = j_2, \cdots, X_n(t) = j_n\}, \tag{5.11}$$

which converge to limits $\pi(j_1, j_2, \cdots, j_n)$ which are independent of G_0 and are determined by equations analogous to (5.10). In this way one could in principle climb up to $n = N$, and one would then have calculated the equilibrium distribution of G_t.

In practice the algebra becomes far too complex, particularly since the cases of genetical interest have very large values of N. This however is balanced by the fact that mutation is usually a rare phenomenon, so that we can write

$$p(i, j) = (1 - u)\delta_{ij} + uq(i, j), \tag{5.12}$$

where $Q = (q(i, j); i, j \in S)$ is stochastic and u is very small.

It turns out that, for many purposes, the geneticist is concerned with large values of N and small values of u such that

$$\theta = 2Nu \tag{5.13}$$

takes moderate values. When this is so, the mathematician can help by pointing out the easily verified fact that, as $N \to \infty$ and $u \to 0$ with θ fixed, each equilibrium probability converges to a limit

$$\Pi_n(j_1, j_2, \cdots, j_n) = \lim \pi(j_1, j_2, \cdots, j_n), \tag{5.14}$$

which satisfies the equations

$$\begin{aligned} n(n - 1 + \theta)&\Pi_n(j_1, j_2, \cdots, j_n) \\ &= \theta \textstyle\sum_{\alpha=1}^n \sum_{i\in S} \Pi_n(j_1, \cdots, j_{\alpha-1}, i, j_{\alpha+1}, \cdots, j_n)q(i, j_\alpha) \\ &\quad + \textstyle\sum_{\alpha=1}^n \Pi_{n-1}(j_1, \cdots, j_{\alpha-1}, j_{\alpha+1}, \cdots, j_n)\nu_\alpha, \end{aligned} \tag{5.15}$$

where

$$\nu_\alpha = \#\{\beta \neq \alpha; j_\beta = j_\alpha\}. \tag{5.16}$$

If Π_{n-1} is known, (5.15) determines Π_n by the obvious iteration. Thus Π_n can be calculated recursively on n, starting from the fact that $(\Pi_1(j))$ is the equilibrium distribution for P (or equivalently for Q). The algebra is still complicated, but much less so than before, and (5.15) is probably well adapted to automatic computation in particular cases.

Note that Π_n is a symmetric function of its n arguments, and that

$$\textstyle\sum_{j\in S} \Pi_n(j_1, \cdots, j_{n-1}, j) = \Pi_{n-1}(j_1, \cdots, j_{n-1}) \qquad (n \geqq 2), \tag{5.17}$$

and

$$\textstyle\sum_{j\in S} \Pi_1(j) = 1.$$

Hence there is an exchangeable sequence $\mathbf{X}$ with values in S for which

$$\mathbf{P}\{X_1 = j_1, X_2 = j_2, \cdots, X_n = j_n\} = \Pi_n(j_1, j_2, \cdots, j_n) \tag{5.18}$$

for all n. De Finetti's theorem therefore shows that there is a random probability measure

$$\mathbf{p} = (p(j); j \in S) \tag{5.19}$$

on S such that

$$\Pi_n(j_1, j_2, \cdots, j_n) = \mathbf{E}\{\mathbf{p}(j_1)\, \mathbf{p}(j_2) \cdots \mathbf{p}(j_n)\}\,. \tag{5.20}$$

The distribution of $\mathbf{p}$ has a natural interpretation ([32], [23]) as the limiting distribution as $N \to \infty$ of the empirical distribution of the N points of G_t. It is easy to check that it is not degenerate, and this presistence of ramdomness has important genetical consequences [39], [40].

6. Random partitions. The argument of the preceding section does not by any means exhaust the usefulness of de Finetti's theorem in population genetics. Another example is that of random partitions arising in models of "infinite alleles" type. Suppose that a sample of n gametes is taken from a large population, and that the gene at a particular locus can be determined for each gamete. Thus the sample of n is partitioned in a random way according to the different alleles represented. A model for the allelic partition (ignoring now, as is often biologically realistic, any labelling of the alleles) is then a probability distribution P_n over the set of partitions of the integer n.

Since n may be chosen at the experimenter's convenience, P_n must be specified for every $n \geqq 1$ (or more strictly for every n less than the population size, but this constraint will be ignored). Moreover, the different distributions P_n must be related by a consistency property. This arises because, for $m < n$, one way of taking a random sample of size m from the population is first to take one of size n and then to take a subsample of m from this sample. Thus P_m must be determined by a relation of the form

$$P_m = \sigma_{mn} P_n\,, \tag{6.1}$$

where σ_{mn} is a certain linear mapping between the relevant spaces of probability measures.

One is therefore led to the study of families (P_n) satisfying (6.1), where P_n is for each $n \geqq 1$ the distribution of a random partition of n. It would be nice to have a representation theorem for such families, but no completely satisfactory result of this type is yet known. However, for genetical purposes it is enough to know that every consistent family satisfying a further condition (discussed in detail in [35]) can be written in the form

$$P_n(\pi) = \int_\nabla \phi_\pi \, d\mu\,, \tag{6.2}$$

where μ is a probability measure on the space ∇ of sequences $\mathbf{x} = (x_1, x_2, \cdots)$ satisfying

$$x_n \geqq x_{n+1}\,, \qquad \sum_{n=1}^\infty x_n = 1\,. \tag{6.3}$$

The function ϕ_π on ∇ is defined for each partition π of n by requiring that $\phi_\pi(\mathbf{x})$

be the probability that a sample of size n from an infinite population, in which alleles are presented with respective frequencies $x_1, x_2, \cdots$, has the allelic partition π.

In genetical applications in which selective pressures are important, the measure μ is usually degenerate. But in selectively neutral models, μ is usually diffuse, and in many of these a particular measure depending on a single parameter θ results. This is the Poisson–Dirichlet distribution [31], which yields in (6.2) the celebrated Ewens sampling formula [15], whose consistency in the sense (6.1) was first noted by Kelly [27]. Thus in (6.2) there is a sharp distinction between selective and neutral models, which can perhaps be exploited in the analysis of allele frequency data.

The representation (6.2) can be looked upon as a version of de Finetti's theorem. If we sample the gametes sequentially, and if we could recognise the genes as corresponding to previously named alleles, we would have an exchangeable sequence of random variables with values in the set of possible alleles. De Finetti's theorem would then give (6.2), where μ is the distribution of the allele frequencies in the population when arranged in descending order. The analysis of [35] shows that this conclusion may still apply if the alleles have not previously been recognised, so that the only information lies in the unlabelled partitions. Although it does not seem possible to derive this result directly from de Finetti's, the methods of proof follow those described in Section 2 above.

7. Some other areas of application. De Finetti's original purpose was to free the theory of inference using repeated experiments from the unsatisfactory features of the frequentist theory (for a more recent account of his ideas see [20]), although it could be argued that, because of its infinitary nature, his theorem comes as near as mathematics can to justifying the frequentist position. A series of observations whose stochastic structure is unaltered by permutations, and which is capable in principle of indefinite extension, can always be regarded as a sequence of independent variables with a common distribution function F, and F (and F alone) can be estimated with arbitrary accuracy from sufficiently many observations.

There seems no further contribution which exchangeability can make to the theory of inference when the observations form a simple series. But the concept finds more subtle application when the data have a more complex structure. This is the case in some sampling problems (e.g., [14]) and in others involving the analysis of variance [37]. There are here some intriguing theoretical questions concerned with invariance under smaller groups of permutations [10], [11].

A related statistical problem is that of multivariate data in which the sample size, though large, is comparable with the dimensionality of the observations. Here the methods of exchangeability, applied not to the observations but to the singular values of the data matrix, have proved valuable in establishing deep limiting properties, notably in the work of Wachter [48].

Finally I would mention the work of Kallenberg ([23], [24], [25]; see also [6]) on processes with exchangeable increments. It would be interesting to relate this to the statistical problem by way of the relationship [32] between the gamma process (regarded as a particular Kallenberg process) and the Dirichlet random measure whose statistical relevance has been stressed by Ferguson [16]. Such a circle of ideas might also have considerable relevance to the structure of genetical populations (cf. [33]).

REFERENCES

[1] ALDOUS, D. J. (1976). *Subsequences of sequences of random variables.* Smith's prize essay, Univ. of Cambridge.

[2] ALDOUS, D. J. (1977a). Subsequences of sequences of random variables. *Bull. Amer. Math. Soc.* **83** 121–123.

[3] ALDOUS, D. J. (1977b). Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **40** 59–82.

[4] ANDERSEN, SPARRE E., (ed.) (1962). *Combinatorial Methods in Probability Theory.* Aarhus.

[5] CANE, V. R. (1966). A note on the size of epidemics and the number of people hearing a rumour. *J. Roy. Statist. Soc. B* **28** 487–490.

[6] CANE, V. R. (1977). A class of non-identifiable stochastic models. *J. Appl. Probability* **14** 475–482.

[7] CHATTERJI, S. D. (1974a). A subsequence principle in probability theory. *Bull. Amer. Math. Soc.* **80** 495–497.

[8] CHATTERJI, S. D. (1974b). A principle of subsequences in probability theory. *Advances in Math.* **13** 31–54.

[9] DACUNHA-CASTELLE, D. (1974). Indiscernability and exchangeability in L^p spaces. *Proc. Seminar on Random Series, Convex Sets and Geometry of Banach Spaces.* Aarhus.

[10] DAWID, A. P. (1972). Contribution to the discussion of [37], 29–30.

[11] DAWID, A. P. (1977a). Invariant distributions and analysis of variance models. *Biometrika* **64** 291–297.

[12] DAWID, A. P. (1977b). Spherical matrix distributions and a multivariate model. *J. Roy. Statist. Soc. B* **39** 254–261.

[13] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[14] ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations. *J. Roy. Statist. Soc. B* **31** 195–233.

[15] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3** 87–112.

[16] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[17] FIGIEL, T. and SUCHESTON, L. (1976). An application of Ramsey sets in analysis. *Advances in Math.* **20** 103–105.

[18] FINETTI, B. DE (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Ser. 6.* Memorie, Classe di Scienze Fisiche, Mathematiche e Naturali **4** 251–299.

[19] FINETTI, B. DE (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1–68.

[20] FINETTI, B. DE (1974). *Theory of Probability.* Wiley, London.

[21] FREEDMAN, D. A. (1963). Invariants under mixing which generalise de Finetti's theorem: continuous time parameter. *Ann. Math. Statist.* **34** 1194–1216.

[22] HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

[23] KALLENBERG, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 309–321.

[24] KALLENBERG, O. (1974). Path properties of processes with independent and interchangeable increments. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 257–271.

[25] KALLENBERG, O. (1975). Infinitely divisible processes with interchangeable increments and random measures under convolution. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 309–321.

[26] KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalisation. *Sankhyā A* **32** 419–430.

[27] KELLY, F. P. (1976). The equilibrium behaviour of stochastic models of interaction and flow. Ph. D. thesis, Univ. of Cambridge.

[28] KENDALL, D. G. (1967). On finite and infinite sequences of exchangeable events. *Studia Sci. Math. Hungar.* **2** 319–327.

[29] KINGMAN, J. F. C. (1965). Contribution to the discussion of [46], 395.

[30] KINGMAN, J. F. C. (1972). On random sequences with spherical symmetry. *Biometrika* **59** 492–493.

[31] KINGMAN, J. F. C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. B* **37** 1–22.

[32] KINGMAN, J. F. C. (1976). Coherent random walks arising from some genetical problems. *Proc. Roy. Soc. A* **351** 19–31.

[33] KINGMAN, J. F. C. (1977a). The population structure associated with the Ewens sampling formula. *Theor. Pop. Biol.* **11** 274–283.

[34] KINGMAN, J. F. C. (1977b). A note on multidimensional models of neutral mutation. *Theor. Pop. Biol.* **11** 285–290.

[35] KINGMAN, J. F. C. (1978). Random partitions in population genetics. *Proc. Roy. Soc. A.* To appear.

[36] KOMLÓS, J. (1967). A generalisation of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.* **18** 217–229.

[37] LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. B* **34** 1–41.

[38] LOÈVE, M. (1960). *Probability Theory.* Van Nostrand, Princeton.

[39] MORAN, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theor. Pop. Biol.* **8** 318–330.

[40] MORAN, P. A. P. (1976). A selective model for electrophoretic profiles in protein polymorphisms. *Genet. Res.* **28** 47–54.

[41] OLSHEN, R. (1973–1974). A note on exchangeable sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 317–321.

[42] RAMSEY, F. P. (1930). On a problem in formal logic. *Proc. London Math. Soc.* (2) **30** 264–286.

[43] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

[44] SPITZER, F. (1956). A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.

[45] TAKÁCS, L. (1962). Ballot problems. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 154–158.

[46] TAKÁCS, L. (1965). Applications of ballot theorems in the theory of queues. *Proc. Symp. Congestion Theor.* Univ. North Carolina, 337–398.

[47] TCHEBYCHEV, P. L. (1948). *Complete Collected Works* **3** 128–131.

[48] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probability* **6** 1–18.

MATHEMATICAL INSTITUTE
24-29 ST. GILES
OXFORD OX1 3LB
ENGLAND