

- 59-82.
- [2] Aldous, D., Weak convergence and the general theory of processes, Projected monograph.
 - [3] Aldous, D., Representations for partially exchangeable arrays of random variables. *J. Multivariate Analysis*, (to appear).
 - [4] Billingsley, P., *Probability and Measure* (Wiley, Chichester, 1979).
 - [5] Brown, I.C., *Compensators and Cox convergence*. *Math. Proc. Cambridge Phil. Soc.* 90 (1981) 305-319.
 - [6] Bühmann, H., *Austauschbare stochastische Variablen und ihre Grenzwertsätze*, *Univ. Calif. Publ. Statist.* 3 (1960) 1-35.
 - [7] Dacunha-Castelle, D., Indiscernability and exchangeability in L^p spaces. In: *Proc. Seminar on Random Series, Convex Sets and Geometry of Banach Spaces*, pp. 50-56 (Aarhus, 1974).
 - [8] Diaconis, P. and Freedman, D., de Finetti's theorem for Markov chains, *Ann. Probab.* 8 (1980) 115-130.
 - [9] Horowitz, J., *Semilinear Markov processes, subordinators and renewal theory*. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 24 (1972) 167-193.
 - [10] Itô, K., *Poisson point processes attached to Markov processes*. In: *Proc. 6th Berkeley Sym. Math. Statist.* Probab. 3 (1972) 225-239.
 - [11] Kallenberg, O., Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 27 (1973) 23-36.
 - [12] Kallenberg, O., *Characterizations and embedding properties in exchangeability*, to appear in *Zeit. Wahr. verw. Gebiete*.
 - [13] Kingman, J.F.C., *Regenerative Phenomena* (Wiley, London, 1972).
 - [14] Kingman, J.F.C., An intrinsic description of local time, *J. London Math. Soc.* (2), 6 (1973) 725-731.
 - [15] Kingman, J.F.C., *Uses of exchangeability*, *Ann. Probab.* 6 (1978) 183-197.
 - [16] Métivier, M., *Reelle und Vektorwertige Quasimartingale und die Theorie der Stochastischen Integration*, *Lecture Notes in Math.* No. 607, (Springer, Berlin, 1977).
 - [17] Williams, D., *Diffusions, Markov Processes and Martingales*, Vol. 1, (Wiley, Chichester, 1979).

O. Kallenberg
 Department of Mathematics
 Chalmers University of Technology
 and the University of Göteborg
 S-41296 Göteborg - Sweden

EXCHANGEABILITY AND THE EVOLUTION OF
 LARGE POPULATIONS

J.F.C. Kingman

Mathematical Institute
 University of Oxford, U.K.

Although de Finetti's theorem relates to infinite exchangeable sequences, there are problems in which it, and other ideas of exchangeability, can apply to large finite collections. This is illustrated by some recent results in population genetics, where an invariance principle involving a limiting Markov chain, the n -coalescent, has proved valuable.

1. A GENERAL STRATEGY

It is perhaps ironic that the theory of exchangeable random variables, developed by de Finetti to give mathematical support to a controversial theory of inference, should have found its most significant applications in philosophically uncontentious areas of pure and applied probability. The representation theorem for infinite exchangeable sequences, which justly bears his name, delimits the models which can be proposed for any phenomenon which has a natural symmetry or invariance property.

A few of the uses of de Finetti's theorem were described in [7], which included an account of a particular model in population genetics. The detail perhaps obscured an interesting feature, that although de Finetti's theorem is about infinite collections, it is there applied where the finiteness of the population is an essential ingredient.

In a biological population of large, but finite, size N , the randomness of the reproductive mechanism ('random genetic drift') typically leads to variances per generation of order N^{-1} . These can often be ignored, where for instance selective differences are much larger than N^{-1} , and then the evolution of the population can be studied deterministically. But there are many biological situations where selective differences, mutation rates and similar quantities are at most of order N^{-1} , and then a deterministic treatment can be seriously misleading.

In such problems one approach which is often fruitful is to consider samples

of size n much smaller than N . This may in any case be appropriate if that is the way in which data are collected (as in [15], for instance). Again, some important questions can be expressed in terms of samples: the mean homozygosity is the probability that a sample of size $n=2$ is monomorphic, and certain covariances studied by Moran [13] are similar probabilities when $n=4$.

Suppose then that a sample of size n is drawn at random from the population at a particular time, and that some (possibly multivariate) measurement yields values X_1, X_2, \dots, X_n for the n members of the sample. Any model for the evolution of the population should lead to a joint distribution P_{Nn} for these random variables, which will depend on N and on other parameters like mutation rates (if the problem is a genetical one) or migration rates (if geographical distribution is relevant). Because of the sampling procedure, P_{Nn} must be exchangeable, but de Finetti's theorem does not yet apply because there are only finitely many X_j .

Suppose however that we wish to model a situation in which N is large and the other parameters are of order N^{-1} . Mathematically this involves investigating P_{Nn} in the limit as $N \rightarrow \infty$, assuming that the parameters when multiplied by N converge to finite limits. In particular models like that of [5] one can see explicitly that, for each n , P_{Nn} converges to a limit P_n . Each P_n is an exchangeable joint distribution for n random variables X_1, X_2, \dots, X_n .

Moreover, the P_n for different values of n are related by the fact that one way of generating a sample of size n is to draw one of size $(n+1)$ and then reject the last member drawn. Thus P_n is obtained from P_{n+1} by integrating out one variable. It follows from the Daniell-Kolmogorov theorem that there is an infinite sequence of random variables X_1, X_2, \dots such that, for each n , P_n is the joint distribution of X_1, X_2, \dots, X_n . Since each P_n is exchangeable, de Finetti's theorem applies to the infinite sequence (X_n) . Hence there is a (possibly random) probability distribution P such that P_n is the joint distribution of n random variables, conditionally independent given P , and each having conditional distribution P .

Attention therefore concentrates on P , which one can regard as the distribution of the X_j in a hypothetical infinite population, even though the finiteness of N is part of the structure of the model. To elucidate the properties of P may be a problem of considerable difficulty, but there is an important class of models in which progress is possible.

2. THE N-COALESCENT

To proceed further we shall make two key assumptions, expressed in the adjectives haploid and neutral.

A haploid population is one in which each individual has just one parent. This is true of some organisms, but most interesting populations are diploid. However, a good deal of genetics is done in terms of the haploid gametes which carry genetical structure from one generation to the next. In ecological problems, one can sometimes work just in terms of a single sex. Nevertheless, it would be very useful to extend the ideas of the present analysis to fully diploid models.

A neutral model in genetics is one in which there are no selective forces acting at the loci in question. The significance of such models has been disputed (see for example [3] or [9]), but they at least serve as null hypotheses in the sense that, if the experimental data are consistent with neutrality, they provide no evidence about the way in which selection might be acting.

The form in which we use neutrality is in an assumption that the individuals present at a given time are exchangeable in their propensity to reproduce, and that information about their X -values, or about their parentage, is irrelevant. This has the effect that one can study the genealogy in terms of some reproductive model, and then ask how this is reflected in the X -values of a sample.

This programme is carried out in detail in [10] for a particular model (of Wright-Fisher type), but the same results arise from large classes of different models. We here describe the limiting process of [10], and then go on in the next three sections to accumulate evidence in favour of the assertion of robustness.

The idea is to consider n individuals drawn at random from the population at time T , and to trace back their ancestry, noting when there are common ancestors of some or all of the sample. Time is here measured in terms of a natural time unit, usually of the order of N generations (but see §5). Since N is large, we consider time as varying continuously.

The genealogy is described in terms of equivalence relations R_t ($t \geq 0$) on the set $\{1, 2, \dots, n\}$, where R_t contains the pair (i, j) if and only if the i th and j th members of the sample have a common ancestor alive at time $T-t$. Clearly R_0 is the identity relation Δ which contains only the pairs (i, i) , and

$$(2.1) \quad R_s \subseteq R_t \quad (s < t).$$

Thus ($R_t; t \geq 0$) is, under any particular model, a random process taking values in

the finite set E_n of equivalence relations on $\{1, 2, \dots, n\}$, satisfying $R_0 = \Delta$ and the monotonicity condition (2.1).

The assertion of [10] is that, in the limit as $N \rightarrow \infty$, the process (R_t) is a particular continuous-time Markov process on E_n called the n-coalescent. This can be defined by its transition rates

$$(2.2) \quad q_{\xi\eta} = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}\{R_{t+h} = \eta | R_t = \xi\}$$

($\xi, \eta \in E_n, \xi \neq \eta$), which take the form

$$(2.3) \quad q_{\xi\eta} = 1 \quad \text{if } \xi \prec \eta \\ = 0 \quad \text{otherwise,}$$

where the notation $\xi \prec \eta$ means that η is obtained from ξ by combining two of the equivalence classes of ξ into a single equivalence class of η .

With the initial condition $R_0 = \Delta$, (2.3) determines the stochastic structure of the n-coalescent, and its properties have been studied in [10] and [11]. Perhaps the most vivid picture is given by the paintbox construction. Let random points $U_1, U_2, \dots, U_n, V_1, V_2, \dots$ be independent and uniformly distributed on the interval $(0, 1)$. Let $(D_t; t \geq 0)$ be a pure death process on the positive integers, with death rate from k to $k-1$ equal to

$$(2.4) \quad d_k = \frac{1}{2} k(k-1),$$

which is independent of the U_i and the V_r . Since $\sum d_k^{-1}$ converges, we can impose the initial condition

$$(2.5) \quad \lim_{t \rightarrow 0} D_t = \infty.$$

Define R_t for $t > 0$ to be the relation on $\{1, 2, \dots, n\}$ which contains the pair (i, j) if and only if there is no point V_r with $1 \leq r \leq D_t - 1$ lying between U_i and U_j . Then (R_t) is an n-coalescent.

This result can be verified by direct calculation, but it really arises from a development of de Finetti's theorem to deal, first with random partitions ([6], [8]) and then with random equivalence relations [11].

3. A GENERAL DISCRETE-GENERATION MODEL

In this section we set the Wright-Fisher model in a much more general context, and show that under weak conditions the n-coalescent arises in the limit as the

population size tends to infinity. Assume therefore that the population evolves in discrete, nonoverlapping generations, and that each generation is of fixed size N . Label the members of the r th generation in random order, and let v_j be the number of children born to the j th member. Then the random variables v_1, v_2, \dots, v_N are exchangeable and satisfy the condition

$$(3.1) \quad \sum_{j=1}^N v_j = N.$$

We assume that the v_j are independent of the family sizes in other generations (neutrality) and that the joint distribution of the v_j does not depend on r .

The Wright-Fisher model is the very special case in which this joint distribution is of symmetric multinomial form

$$(3.2) \quad \mathbb{P}\{v_j = n_j \text{ (} j=1, 2, \dots, N)\} = \frac{N! N^{-N}}{n_1! n_2! \dots n_N!}$$

Fix $n \leq N$ and $r \geq 1$, and choose n members at random from the r th generation. For $0 \leq s \leq r$, let R_s be the equivalence relation which contains the pair (i, j) if and only if the i th and j th members of this sample have a common ancestor in the $(r-s)$ th generation. Then $R_0 = \Delta$ and $R_s \subseteq R_{s+1}$. Moreover, the conditional distribution of R_{s+1} given R_1, R_2, \dots, R_s clearly depends only on R_s and not on R_1, R_2, \dots, R_{s-1} or on r . In other words, $(R_s; s=0, 1, 2, \dots, r)$ is a Markov chain with values in E_n , whose transition probabilities depend only on N, n and the joint distribution of the v_j . Although R_s is only defined for $s \leq r$, its distributions do not depend on r , and are therefore those of a Markov chain $(R_s; s=0, 1, 2, \dots)$. Conditions under which this latter chain, in a suitable time scale, converges in distribution to the n-coalescent, are given by the following theorem.

Theorem 1 Suppose that, as $N \rightarrow \infty$, the variance of v_j converges to a non-zero limit σ^2 , and that

$$(3.3) \quad \sup_N \mathbb{E}(v_1^k) < \infty$$

for all $k \geq 1$. Then the finite-dimensional distributions of the process

$$(3.4) \quad R_t = R_{\lfloor N\sigma^{-2}t \rfloor}$$

converges to those of the n-coalescent.

Proof We compute first the probability

$$P_{\xi\eta} = \mathbb{P}(R_{S+1} = \eta | R_S = \xi)$$

for $\xi, \eta \in E_n$. This is zero unless $\xi \subseteq \eta$, in which case the equivalence classes of η may be labelled as C_α ($\alpha=1, 2, \dots, a$) and those of ξ as $C_{\alpha\beta}$ ($\alpha=1, 2, \dots, a; \beta=1, 2, \dots, b_\alpha$), where

$$C_{\alpha\beta} = \bigcup_{\beta=1}^{b_\alpha} C_\alpha$$

$$k = \sum_{\alpha=1}^a b_\alpha$$

The $C_{\alpha\beta}$ correspond to

particular individuals in the daughter generation, and $P_{\xi\eta}$ is the probability that, if k individuals are selected at random from this generation and labelled as $C_{\alpha\beta}$, then all the $C_{\alpha\beta}$ for each fixed α have the same parent, and that these parents for different values of α are distinct. Conditional on the family sizes v_j in the parent generation, this probability is

$$\sum (v_{j_1} b_1) (v_{j_2} b_2) \dots (v_{j_a} b_a) / (N)_k,$$

where the summation extends over all distinct j_1, j_2, \dots, j_a in $1 \leq j \leq N$ and we use the notation

$$(N)_k = N(N-1) \dots (N-k+1).$$

Hence

$$(3.5) \quad P_{\xi\eta} = \mathbb{E} \left\{ (N)_k^{-1} \sum (v_{j_1} b_1) (v_{j_2} b_2) \dots (v_{j_a} b_a) \right\}$$

when $\xi \subseteq \eta$, and zero otherwise.

Consider first the case $\xi \subset \eta$, in which $a = k-1$ and we can, without loss of generality, take $b_1 = 2, b_2 = \dots = b_a = 1$.

Then the sum in (3.5) is

$$S = \sum v_{j_1} (v_{j_1} - 1) v_{j_2} \dots v_{j_{k-1}} \ll \sum_{j=1}^N v_j (v_j - 1) N^{k-2}$$

so that

$$P_{\xi\eta} \ll \mathbb{E} \left\{ (N)_k^{-1} N^{k-2} \sum_{j=1}^N v_j (v_j - 1) \right\}$$

$$= (N)_k^{-1} N^{k-1} \mathbb{E}(v_1 (v_1 - 1))$$

$$= N^{-1} \sigma^2 + o(N^{-1}),$$

using the exchangeability of the v_j and the fact that, from (3.1), $\mathbb{E}(v_1) = 1$. On the other side,

$$S \geq \sum_{j=1}^N v_j (v_j - 1) \left\{ (N - v_j)^{k-2} - \binom{k-2}{2} \sum_{i \neq j} v_i^{k-4} \right\}$$

$$\geq \sum_{j=1}^N v_j (v_j - 1) N^{k-2} - (k-2) \sum_{j=1}^N v_j^{k-3} - \binom{k-2}{2} \sum_{i \neq j} v_i^2 v_j^{k-4}$$

so that

$$P_{\xi\eta} \geq N^{-1} \mathbb{E}(v_1 (v_1 - 1)) - (k-2) N^{-2} \mathbb{E}(v_1^3) - \binom{k-2}{2} N^{-2} \mathbb{E}(v_1^2 v_2^2)$$

$$= N^{-1} \sigma^2 + o(N^{-1}),$$

using (3.3) and the Schwarz inequality. Thus, when $\xi \subset \eta$,

$$(3.6) \quad P_{\xi\eta} = N^{-1} \sigma^2 + o(N^{-1}).$$

Now consider the case when $\xi \subset \eta$ but $\xi \subset \eta$ is false. Then $a \leq k-2$ and (3.5) gives

$$P_{\xi\eta} \leq (N)_k^{-1} \sum \mathbb{E} \left\{ v_{j_1}^{b_1} v_{j_2}^{b_2} \dots v_{j_a}^{b_a} \right\}$$

$$\leq (N)_k^{-1} \sum \mathbb{E} \left\{ \binom{k}{j_1} \binom{k}{j_2} \dots \mathbb{E} \left\{ v_{j_a}^{b_a/k} \right\} \right\}$$

$$= (N)_k^{-1} (N)_a \mathbb{E}(v_1^k) = o(N^{-2}),$$

using Hölder's inequality and (3.3). Thus, for any $\xi, \eta \in E_n$,

$$(3.7) \quad P_{\xi\eta} = N^{-1} \sigma^2 q_{\xi\eta} + o(N^{-1}),$$

and as in [10] this means that the stochastic matrix $P_N = (P_{\xi\eta})$ satisfies

$$(3.8) \quad \lim_{N \rightarrow \infty} P_N [N \sigma^{-2} t] = e^{t\Omega}$$

for all $t \geq 0$, where Ω is the infinitesimal generator of the n -coalescent. This suffices to prove the theorem.

Condition (3.3) does not seem to exclude any reproductive processes of interest, but the condition that σ^2 be non-zero does fail in one case, the discrete-

neration form of Moran's model [12]. As explained in [10], this is because a different natural time unit, of $\frac{1}{2} N^2$ generations, is appropriate. We return to this problem in §5.

4. THE EFFECT OF VARYING POPULATION SIZE

The model of §3 has the serious drawback that the total population size N is required to be fixed. It is often true in practice that N fluctuates in some manner determined by external factors. The analysis of Theorem 1 still holds in such cases, to give (3.7) where N is the value at that time, but the step from (3.7) to (3.8) requires modification. What happens is that the natural time scale is non-linear, and is defined by assigning the step from one generation to the next a 'time' interval equal to the local value of $N^{-1} \sigma^2$. If this is done, the resulting process (R_t) still converges in distribution to the n -coalescent.

A more delicate problem arises when the variation in N is endogeneous rather than exogeneous, in that it comes from the reproductive mechanism itself rather than from external factors. Then N cannot be treated as a constant in the expectation (3.5). The natural time scale now assigns the step from one generation to the next an interval

$$(4.1) \quad \delta = \mathbb{E} (N^{-1} v_1 (v_1 - 1)),$$

and convergence to the n -coalescent of the scaled process

$$(4.2) \quad R_t = R_{\lfloor t/\delta \rfloor}$$

takes place if

$$P_{\xi n} = \delta q_{\xi n} + o(\delta)$$

as $\delta \rightarrow 0$.

The proof of Theorem 1 can be carried through in this more general situation, so long as estimates of the form

$$(4.4) \quad \mathbb{E} \left\{ N^{-2} \begin{matrix} b_1 & b_2 & b_a \\ v_1 & v_2 & \dots & v_a \end{matrix} \right\} = o(\delta)$$

are available for $a, b_1, b_2, \dots, b_a \geq 1$, remembering that N is now a random variable, but still satisfies (3.1). By Hölder's inequality, the left hand side of (4.4) is not greater than

$$\prod_{\alpha=1}^a \left\{ \mathbb{E} (N^{-2} \frac{k_\alpha}{v_\alpha}) \right\}^{b_\alpha/k_\alpha} = \mathbb{E} (N^{-2} \frac{k}{v}),$$

and we have the following theorem.

Theorem 2 If δ is defined by (4.1) and (R_t) by (4.2), and if the joint distribution of the v_j varies with δ in such a way that

$$(4.5) \quad \mathbb{E} (N^{-2} \frac{k}{v_j}) = o(\delta)$$

as $\delta \rightarrow 0$, for all $k \geq 1$, then (R_t) converges in distribution to the n -coalescent.

5. MORAN'S MODEL

An early alternative to the Wright-Fisher model was proposed by Moran in [12]. This is most naturally expressed in continuous time, when it assumes that, in a time interval $(T, T+h)$, there is a probability $\lambda h + o(h)$ of a change of state, independently of the past. A change of state consists of a birth to one of the N individuals (chosen at random), the population size being held fixed by the simultaneous death of a randomly chosen individual.

As before, fix $T > 0$ and $n \leq N$, and select a sample of size n from the population at time T . Define R_t ($0 \leq t \leq T$) as the equivalence relation in E_n which contains (i, j) if and only if the i th and j th members of the sample have a common ancestor at time $T-t$. The next theorem shows that, in a suitable time scale, (R_t) is an n -coalescent. This is an exact result, not an asymptotic property as $N \rightarrow \infty$. It has affinities with reversibility properties of the Moran process, such as have been exploited by Watterson [18] and Kelly [4].

Theorem 3 If time is measured so that

$$(5.1) \quad \lambda = \frac{1}{2} N^2,$$

then (R_t) is the restriction to $[0, T]$ of an n -coalescent.

Proof Consider first the case $n = N$, and note the possible changes of state of the population in $[0, T]$, as one moves back in time from T . A typical change selects two members i and j (with replacement) from the population. It reverses the birth of i out of j , and replaces it by the member who died at that instant. If $i = j$ there is no such replacement, j being restored from death in childbirth. In terms of common ancestors, this amounts if $i \neq j$ to a coalescence of i and j , and if $i = j$ to no effective change.

In terms of R_t , a change of state when R_t has k equivalence classes amounts to a coalescence of two of these with probability $k(k-1)/N^2$, and to no change

otherwise. Thus (R_t) is a Markov process, with probability

$$\lambda k(k-1)N^{k-2}h + o(h) = \frac{1}{2} k(k-1)h + o(h)$$

of making a transition in $(t, t+h)$. If it makes a transition from $R_t = \xi$, this takes it to η , where η is equally likely to be any of the $\frac{1}{2} k(k-1)$ relations with $\xi \prec \eta$. This shows that (R_t) is an N -coalescent.

To deal with the case $n < N$, we have only to use the observation in [10] that the restriction to $\{1, 2, \dots, n\}$ of an N -coalescent is an n -coalescent. This completes the proof.

6. THE EWENS SAMPLING FORMULA

The results of the previous sections suggest quite strongly that the n -coalescent can be taken as a robust description of the genealogy of a sample of size $n \ll N$ from a neutral haploid population of large size N . The details of the reproductive mechanism affect only the time scale in which the genealogy is to be observed.

It follows that any aspect of the sample, which depends on the history of the population only through the family tree of the sample, has a robust limit as $N \rightarrow \infty$ which can be derived from the n -coalescent. This is an invariance principle just like the classic one of Donsker [1], and the n -coalescent plays the same role as does the Wiener process (brownian motion) for Donsker.

A good example is given by models of neutral mutation in population genetics. Whether two gametes have the same allele at one particular chromosome locus depends on the mutations which have occurred in their lines of descent from their last common ancestor. Thus it is only necessary to build onto the genealogy a model for mutation (independent of (R_t) because of the assumption of neutrality). One can therefore conclude without further calculation that two models with the same mutation structure (expressed in the natural time scale), which both fall in the domain of attraction of the n -coalescent, will both have the same limiting distributions for the genetical composition of a sample.

This applies to the complicated models where mutation is represented by a random walk or Markov chain ([14], [13], [9]), but the most useful case is that of non-recurrent mutation. This assumes that mutation always produces a completely novel allele, so that two gametes have the same allele if and only if there is no mutation (at that locus) in either line of descent from their common ancestor.

Such a process can be formulated directly in terms of the n -coalescent. Suppose that at time t there are k equivalence classes in R_t . Then it is assumed that the probability that a mutation affects any one of these classes in $(t, t+h)$ is $\frac{1}{2} \theta h + o(h)$, independently of (R_t) and of mutations in other time intervals or to other classes. The constant θ is a measure of the mutation rate (the factor $1/2$ being conventional and convenient). Thus mutations affect equivalence classes in a Poisson manner between coalescences.

Now define a random equivalence relation R on $\{1, 2, \dots, n\}$, as being that which contains the pair (i, j) if and only if no mutation affects the equivalence classes containing i or j in the interval $(0, \tau_{ij})$, where

$$(6.1) \quad \tau_{ij} = \inf \{t \geq 0; (i, j) \in R_t\}.$$

The next theorem derives the distribution of R .

Theorem 4 For any $\xi \in E_n$ whose equivalence classes are of sizes $\lambda_1, \lambda_2, \dots, \lambda_k$,

$$(6.2) \quad \mathbb{P}\{R = \xi\} = \frac{\theta^{k-1}}{(\theta+1)(\theta+2)\dots(\theta+n-1)} \prod_{\alpha=1}^k (\lambda_\alpha - 1)!$$

Proof In the interval $(0, h)$ the possible events (ignoring those with probability $o(h)$) are a transition in (R_t) creating an equivalence class of two members, and a mutation to one of the n singletons of R_0 . Hence

$$P_n(\xi) = \mathbb{P}\{R = \xi\}$$

satisfies

$$P_n(\xi) = (1 - \frac{1}{2} n(n-1)h - \frac{1}{2} n\theta h) P_n(\xi) + h \sum_1 P_{n-1}(\eta) + \frac{1}{2} \theta h \sum_2 P_n(\zeta) + o(h),$$

where (i) \sum_1 is the sum over all pairs $(i, j) \in \xi$ with $i < j$, and η is the restriction of ξ to $\{1, 2, \dots, j-1, j+1, \dots, n\}$, relabelled to bring it into E_{n-1} , and (ii) \sum_2 is the sum over all singletons i of ξ and all relations $\zeta \in E_n$ which yield ξ when i is removed from its equivalence class.

Hence

$$(6.3) \quad n(\theta+n-1) P_n(\xi) = 2 \sum_1 P_{n-1}(\eta) + \theta \sum_2 P_n(\zeta),$$

and this equation determines $P_n(\xi)$ uniquely for all $n \geq 1$, $\xi \in E_n$, by induction on n and on the number of equivalence classes of ξ .

Thus the theorem is proved if (6.2) can be shown to satisfy (6.3). If (6.2) is substituted into (6.3) and common factors removed, we have to show that

$$n^{(\theta+n-1)} = 2^{(\theta+n-1)} \sum_{\{\alpha; \lambda \geq 1\}} \frac{1}{2} \lambda_{\alpha} + \sum_{\{\alpha; \lambda = 1\}} \sum_{\beta \neq \alpha} [\theta + \sum_{\beta} \lambda_{\beta}]$$

and this is true since the right hand side is

$$(\theta+n-1) \sum_{\{\alpha; \lambda \geq 1\}} \lambda_{\alpha} + \sum_{\{\alpha; \lambda = 1\}} (\theta+n-1) = (\theta+n-1) \sum_{\alpha=1}^k \lambda_{\alpha}$$

Any $\xi \in E_n$ defines an induced partition

$$(6.4) \quad \pi(\xi) = 1 \quad 2 \quad \dots \quad n \quad \begin{matrix} a_1 & a_2 & \dots & a_n \\ \dots & \dots & \dots & \dots \end{matrix}$$

where a_r is the number of equivalence classes of size r . The probability (6.2) depends on ξ only through $\pi(\xi)$, since (6.2) can be written as

$$(6.5) \quad \mathbb{P}(R = \xi) = \{\theta(\theta+1) \dots (\theta+n-1)\}^{-1} \prod_{r=1}^n \frac{a_r}{\{\theta(r-1)!\}^{a_r}}$$

The number of equivalence relations with a given partition is

$$n! / \prod_{r=1}^n \{ (r!)^{a_r} a_r! \},$$

and therefore

$$(6.6) \quad \mathbb{P} \left\{ \pi(R) = 1 \quad 2 \quad \dots \quad n \quad \begin{matrix} a_1 & a_2 & \dots & a_n \\ \dots & \dots & \dots & \dots \end{matrix} \right\} = \frac{n!}{\theta^{a_1} \prod_{r=1}^n \frac{a_r^{a_r}}{r^{a_r} a_r!}}$$

This is the sampling formula first propounded by Ewens [2] and since established by direct analysis of several different models (see for example [17] and papers cited there). We see that the proof of Theorem 4 is really redundant, since once (6.6) has been proved for any model in the domain of attraction of the n -coalescent, it is automatically true for any other such model. An obvious exchangeability argument leads from (6.6) back to (6.2).

Similar arguments hold for more complex models of neutral mutation. The results for these are much more involved, and less explicit than (6.2) and (6.6), but they are equally robust to details of the reproductive mechanism.

7. THE COMMON ANCESTOR PROBLEM

Not every interesting question can be posed in terms of a sample of fixed size n , and therefore the robust limiting results derived from the n -coalescent may not be directly applicable. It may still be possible to use these results to derive some useful information. We illustrate the difficulties and the possibilities by considering the problem: if the genealogy of a particular generation is traced back, how far is it necessary to go before there is an ancestor common to the whole gen-

eration?

Let A_s be the number of individuals, s generations back, who are ancestors of some members of the given generation. In the general model of §3, $(A_s; s=0,1,2,\dots)$ is a Markov chain because the probability (3.5) depends only on a, b_1, b_2, \dots, b_a and not on the sizes of the $C_{\alpha\beta}$. Of course,

$$N = A_0 \geq A_1 \geq A_2 \geq \dots \geq 1,$$

and the problem relates to the passage time

$$(7.1) \quad T_N = \inf \{s; A_s = 1\}$$

from N to 1.

The theory of the n -coalescent can be applied to T_{Nn} , the number of generations one must go back before a particular sample of n individuals has a common ancestor. In fact [10], Theorem 1 implies that

$$(7.2) \quad \mathbb{P} \{N_{\sigma}^{-1} 2 T_{Nn} \leq t\} \rightarrow \psi_n(t)$$

as $N \rightarrow \infty$ for fixed n and t , where ψ_n is the distribution function of

$$(7.3) \quad \sum_{m=2}^n \tau_m,$$

the τ_m being independent and exponentially distributed with

$$(7.4) \quad \mathbb{E}(\tau_m) = 2/m(m-1).$$

Since $T_N \geq T_{Nn}$ ($n \leq N$), (7.2) implies that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \{N_{\sigma}^{-1} 2 T_N \leq t\} \leq \psi_n(t),$$

and since this holds for all n ,

$$(7.5) \quad \limsup_{N \rightarrow \infty} \mathbb{P} \{N_{\sigma}^{-1} 2 T_N \leq t\} \leq \psi(t),$$

where $\psi = \lim_{n \rightarrow \infty} \psi_n$ is the distribution function of the variable

$$(7.6) \quad T = \sum_{m=2}^{\infty} \tau_m,$$

which has

$$(7.7) \quad \mathbb{E}(T) = \sum_{m=2}^{\infty} \frac{2}{m(m-1)} = 2$$

and is therefore almost certainly finite.

It would be useful to strengthen (7.5) to

$$(7.8) \quad \lim_{N \rightarrow \infty} \mathbb{P} \{N_{\sigma}^{-1} 2 T_N \leq t\} = \psi(t)$$

by bounding T_N from above. Notice that (7.5) implies that

$$\limsup_{N \rightarrow \infty} N^{-1} \mathbb{E}(T_N) \geq 2\sigma^{-2},$$

and (7.8) would follow if this could be strengthened to

$$(7.9) \quad \lim_{N \rightarrow \infty} N^{-1} \mathbb{E}(T_N) = 2\sigma^{-2}.$$

This seems in general a delicate problem, but it is possible to prove (7.9) for the Wright-Fisher model (see [16] for a related discussion).

Theorem 5 In the Wright-Fisher model (as described in §3 with (3.2))

$$(7.10) \quad \mathbb{E}(T_N) \leq 2(N-1),$$

and therefore (7.8) and (7.9) hold with $\sigma^2 = 1$.

Proof We begin with a variant of an inequality proved in [5]. First show by calculus that, for $n \geq 2$ and $0 \leq x \leq 1$,

$$x^{n-1} \geq \frac{nx - (n-2)}{n - (n-2)x}.$$

Set $x = (k-1)/k$ and rearrange to give

$$2nk^{n-1} \geq (2k+n)k^{n-1} - (2(k-1)+n)(k-1)^{n-1}.$$

The left hand side is

$$2nk^{-1} \sum_{r=1}^n S(n,r)(k)_r = 2n \sum_{r=1}^n S(n,r)(k-1)_{r-1},$$

where $S(n,r)$ is the usual Stirling number, so that

$$(7.11) \quad 2n \sum_{r=1}^n S(n,r)(k-1)_{r-1} \geq (2k+n)k^{n-1} - (2(k-1)+n)(k-1)^{n-1}.$$

Sum from $k=1$ to $k=N$ to give

$$(7.11) \quad 2n \sum_{r=1}^n S(n,r)(N)_r^{-1} \geq (2N+n)N^{n-1},$$

for $n \geq 2$, $N \geq 1$.

For the Wright-Fisher model the transition probabilities of (A_s) are

$$P_{nr} = \mathbb{P}(A_{s+1} = r | A_s = n) = S(n,r)(N)_r N^{-n}$$

so that (7.11) implies that

$$(7.12) \quad \sum_{r=1}^n P_{nr} \frac{1}{r} \geq \frac{1}{2N} + \frac{1}{n} \quad (n \geq 2).$$

We now prove by induction on n that the expected passage time of (A_s) from n to 1 satisfies

$$(7.13) \quad \mu_n \leq 2N(1-n^{-1}).$$

This is trivially true for μ_1 . Suppose that it is true for $\mu_1, \mu_2, \dots, \mu_{n-1}$ and use the identity

$$\mu_n = 1 + \sum_{r=1}^n P_{nr} \mu_r \quad (n \geq 1)$$

to deduce that

$$\begin{aligned} \mu_n(1 - P_{nn}) &= 1 + \sum_{r=1}^{n-1} P_{nr} \mu_r \\ &\leq 1 + \sum_{r=1}^{n-1} P_{nr} 2N(1-r^{-1}) \\ &\leq 1 + 2N - 2N\left(\frac{1}{2N} + \frac{1}{n}\right) - P_{nn} 2N\left(1 - \frac{1}{n}\right) \\ &= 2N(1-n^{-1})(1 - P_{nn}). \end{aligned}$$

Hence the induction succeeds and (7.13) is proved. In particular, (7.10) follows on setting $n=N$.

It follows that

$$\int_0^\infty \mathbb{P}(N^{-1}T_N > t) dt \leq 2(1-N^{-1}) \leq 2,$$

while (7.7) shows that

$$\int_0^\infty (1 - \Psi(t)) dt = 2.$$

This together with (7.5) and Fatou's lemma implies (7.8) and (7.9) with $\sigma^2 = 1$.

Hence the theorem is proved.

It is an open question whether a similar analysis is possible for the general model of §3. However, it is in a sense rather unsatisfactory to arrive at conclusions like (7.8) and (7.9) by direct estimation. It would be more in the spirit of Donsker to establish a 'functional limit theorem' from which results like Theorem 5 could be deduced at once from convergence of 'finite-dimensional' distributions (which here means distributions of samples of fixed, but arbitrary, size).

REFERENCES

[1] Donsker, M.D., An invariance principle for certain probability limit theorems, *Mem. Amer. Math. Soc.* 6 (1951).
 [2] Ewens, W.J., The sampling theory of selectively neutral alleles, *Theor. Pop. Biol.* 3 (1972) 87-112.
 [3] Ewens, W.J., *Mathematical Population Genetics*, Springer, Berlin, 1979.

- [4] Kelly, F.P., Exact results for the Moran neutral allele model, Adv. Appl. Prob. 9 (1977) 197-201.
- [5] Kingman, J.F.C., Coherent random walks arising in some genetical models, Proc. Roy. Soc. A 351 (1976) 19-31.
- [6] Kingman, J.F.C., Random partitions in population genetics, Proc. Roy. Soc. A 361 (1978) 1-20.
- [7] Kingman, J.F.C., Uses of exchangeability, Ann. Prob. 6 (1978) 183-197
- [8] Kingman, J.F.C., The representation of partition structures, J. London Math. Soc. 18 (1978) 374-380.
- [9] Kingman, J.F.C., Mathematics of Genetic Diversity, (Soc. Industr. Appl. Math., Philadelphia, 1980).
- [10] Kingman, J.F.C., On the genealogy of large populations (to appear).
- [11] Kingman, J.F.C., The coalescent (to appear).
- [12] Moran, P.A.P., Random processes in genetics, Proc. Camb. Phil. Soc. 54 (1958) 60-71.
- [13] Moran, P.A.P., Wandering distributions and the electrophoretic profile, Theor. Pop. Biol. 8 (1975) 318-330.
- [14] Ohta, T. and Kimura, M., A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population, Genet. Res. 22 (1973) 201-204.
- [15] Singh, R.S., Lewontin, R.C. and Felton, A.A., Genetic heterogeneity within electrophoretic 'alleles' of xanthine dehydrogenase in *Drosophila pseudoobscura*, Genetics 84 (1976) 609-629.
- [16] Watterson, G.A., On the number of segregating sites in genetical models without recombination, Theor. Pop. Biol. 7 (1975) 256-276.
- [17] Watterson, G.A., The stationary distribution of the infinitely many neutral alleles diffusion model, J. Appl. Prob. 13 (1976) 639-651.
- [18] Watterson, G.A., Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model, Theor. Pop. Biol. 10 (1976) 239-253.

J.F.C. Kingman
 Mathematical Institute
 University of Oxford
 24-29 St. Giles
 Oxford OX1 3LB - U.K.

RENEWAL THEORY FOR INTERCHANGEABLE
 RANDOM VARIABLES

H. Teicher
 Rutgers University
 New Brunswick, New Jersey

If $T_c(\alpha) = \inf \{n \geq 1: \sum_{i=1}^n X_i > cn^\alpha\}$, $c > 0$, $0 \leq \alpha < 1$ where $\{X_n, n \geq 1\}$ are interchangeable random variables with positive mean and zero covariance, it is shown that $ET_c(\alpha) \sim (c/\mu)^{1/\alpha}$ as $c \rightarrow \infty$. Under the additional hypotheses of finite variance and $\text{Cov}(X_1^2, X_2^2) = 0$, a Central Limit Theorem is proved for both $T_c(\alpha)$ and $N_c(\alpha) = \sup \{n \geq 1: \sum_{i=1}^n X_i \leq cn^\alpha\}$.

1. INTRODUCTION

Classical renewal theory, beginning with positive i.i.d. random variables $\{X_n, n \geq 1\}$, has been concerned with $N_c = \sup \{n \geq 1: \sum_{i=1}^n X_i \leq c\}$, $c > 0$ and in particular with the behavior of its first and second moments as c tends to infinity. Extensions have been made to a non-constant boundary cn^α , $0 \leq \alpha < 1$ and to i.i.d. random variables with positive mean. Moreover, a central limit theorem has been established for N_c when the underlying random variables $\{X_n, n \geq 1\}$ have finite variance.

Here, extensions of these results to interchangeable random variables with positive mean will be obtained. The only prior discussion known to the author of renewal theory in the context of interchangeable random variables is in [2].

2. MAINSTREAM

Let $\{X_n, n \geq 1\}$ be interchangeable random variables (r.v.'s) on some probability space. Then there exists a σ -algebra G such that $\{X_n, n \geq 1\}$ are conditionally independent and identically distributed (i.i.d.) given G and so if $EX_1 = \mu$ and EX_1^2 exists,