a more general one, namely

$$\binom{s}{r} r^{\alpha}\theta^{r}(1-\theta)^{s-r} \Big/ \Big[\sum_{j=1}^{s}\binom{s}{j} j^{\alpha}\theta^{j}(1-\theta)^{s-j}\Big]; \qquad r=1,\cdots,s$$

(where the parameter $\alpha \geqq 0$), could be based on two different sets of assumptions, one saying that the probability that a family was ascertained, was proportional to r^{α} (for $r \geqq 1$) and another set saying that each family has been ascertained through a *single* affected child (the *proband*) and that the probability that the proband was the first affected child, was π and that it was the ith affected child ($2 \leqq i \leqq r$), was $\pi(i^{\alpha}-(i-1)^{\alpha})$. (The parameter π is referring to family specific properties.) By choosing $\alpha=0$ or $\alpha=1$ the first two models were obtained.

By means of this observation a multiple test method for choice between single-proband models was constructed. The method was based on records of the birth number of the proband among the affected children. Further details have been given in Stene (1977), (1978).

Extensions to models for family data ascertained through simultaneous information about m affected children in each family were also considered. It was shown that the conditional probability that a family in the material had r affected children, given that it had s children altogether and had been ascertained through m probands by a procedure specified analogously to that before, was

$$\binom{s}{r}\binom{r}{m}^{\alpha}\theta^{r}\cdot(1-\theta)^{s-r} \Big/ \sum_{j=m}^{s}\binom{s}{j}\binom{j}{m}^{\alpha}\theta^{j}(1-\theta)^{s-j}$$

where $m \leqq r \leqq s$. Here m is a parameter and not a random variable.

References

STENE, J. (1977) Assumptions for different ascertainment models. *Biometrics* **33,** 523–527.
STENE, J. (1978) Choice of ascertainment model I. *Ann. Hum. Genet.* **42,** 219–229.

Time reversion and age distributions

S. TAVARÉ, *University of Sheffield*

Adv. Appl. Prob. 11, 10–12, 1979

We consider Markov chains X_n with state space S, absorbing states C, and transient states T. The return process, $\bar{X}_n$, is generated from the X process by specifying distributions over S which determine how the X process is restarted whenever C is hit. The resulting process is taken to be irreducible, aperiodic,

and positive recurrent. A special case of such a process was considered by Levikson (1977) in relation to determining the distribution of G_j, the age of $\{X_n\}$ given current position $j \in T$. The distribution is given by

$$P(G_j = n) = \lim_{m \to \infty} P(\bar{X}_{m-n} \in C, \bar{X}_{m-k} \notin C \quad 0 < k < n \mid \bar{X}_m = j).$$

The initial state is assumed fixed. However, under the stated conditions, the limit above is independent of the initial distribution of $\bar{X}_0$. It follows that if we assume that the chain is stationary, we can reverse time to obtain a new Markov chain $\{\bar{Y}_n\}$. The restarting method of Levikson is modified to ensure that $\bar{Y}_n$ and $\bar{X}_n$ have the same state spaces. (See Tavaré (1978), Pakes (1978).) This chain may or may not be probabilistically identical to $\{X_n\}$ (cf. Watterson (1978)), but has a transition matrix which is readily computable in terms of that of $\{\bar{X}_n\}$, and therefore that of $\{X_n\}$. To determine properties of the age distribution specified above, we form the absorbing process $\{Y_n\}$, which has transition matrix identical to that of $\{\bar{Y}_n\}$, but with states in C made absorbing. It should now be clear that the distribution of G_j is the absorption time distribution of $\{Y_n\}$, given $Y_0 = j$. The last visit probabilities for $\{X_n\}$, which were interpreted by Levikson, in a genetic context, as the probabilities of different alleles being the oldest, are simply the absorption probabilities for $\{Y_n\}$. It follows immediately from the representation of the age in terms of properties of a readily specified Markov chain that the age can be represented as the sum of a sequence of geometric random variables, whose parameters can be related to the original chain.

Time-reversed processes have appeared in a genetic context before. Seneta (1965) discusses the interpretation of quasi-stationary distributions of absorbing Markov chains in terms of certain reversed processes. The use of time reversion is also a useful method for deriving properties of age distributions for continuous-time processes with arbitrary restarting distributions.

We can also use a similar method for those cases in which $\{\bar{X}_n\}$ is null-recurrent, if we suppose that $\bar{X}_0$ is distributed as the stationary measure. The results so obtained agree with those derived by Pakes (1978) for cases where the strong ratio limit property holds.

Also discussed was a way of interpreting formal applications of Wright's formula for the stationary distribution of diffusion processes (Wright (1937)) in cases where no stationary distribution can exist. The result was described in terms of a simple genetic mechanism, and discussed in the light of other solutions of this problem.

References

LEVIKSON, B. (1977) The age distribution of Markov processes. *J. Appl. Prob.* **14**, 492–506.
PAKES, A. G. (1978) On the age distribution of a Markov chain. *J. Appl. Prob.* **15**, 65–77.

SENETA, E. (1965) Quasi-stationary distributions and time reversion in genetics. *J. R. Statist. Soc.* **B28,** 253–277.

TAVARÉ, S. (1978) Age distributions for Markov processes in genetics (abstract). *Adv. Appl. Prob* **10,** 17–19.

WATTERSON, G. A. (1978) Reversibility and the age of an allele II. Two allele models with selection and mutation. *Theoret. Popn Biol.* **12,** 179–196.

WRIGHT, S. (1937) The distribution of gene frequencies in populations. *Proc. Natn. Acad. Sci. U.S.A.* **23,** 307–320.

Extinction probabilities and pedigree structure

E. A. THOMPSON, *University of Cambridge*

The evolution of a population is the joint survial of its genes. The mechanism of Mendelian segregation and the limited number of paths of descent for genes in a small population result in negative correlations in the extinction and survival of disjoint sets of genes. The characterisation of genealogical structure is a long-standing problem in human genetics, and these correlations appear to be good candidates for useful structural parameters. As described by Thompson (1978) the recursive method of Cannings et al. (1978) may be used to compute extinction probabilities for arbitrary sets of founder genes in pedigrees of, in theory, arbitrary size and complexity.

However, if extinction correlations are to be used as parameters of structure, a base-point for comparison must first be provided by an analysis of their behaviour under specific mating structures. The situations of outbreeding, random mating and various regular mating systems have been considered and show that extinction correlations can provide consistent characterisations of a mating structure, and can be used as a basis for comparison of relationships between different sets of individuals within a structure. (When we speak of 'relationships' between individuals in this context, we mean relationships with regard to the descent of their genes and hence the extent of co-descendance in the current population rather than the extent of co-ancestry amongst original members.)

For the population of the isolated island of Tristan da Cunha, there are twelve early founders who contributed a large majority of the genes of the 1961 sampled population of 243 individuals. The extinction correlations between these founders have been analysed and show interesting results. Correlations are of course normally largest between the two genes within an individual, next between spouses, next between original founder couples, next between these couples and their founder sons-in-law, and so on. However, some high correlations between 'distant' founders illuminate patterns of co-descendants that are not immediately obvious, and have not emerged from previous analyses of the pedigree by traditional methods.