

The same relationship also makes it possible to deduce a mapping formula for any proposed distribution of the chiasmata. The probability of no chiasmata occurring in a segment is expressed in terms of a parameter which is related to the length of the segment. Both the recombination fraction and the average number of chiasmata are expressed in terms of this parameter, which is then eliminated from the expressions. If we assume that there is at least one chiasma in every chromosome arm, with the remaining chiasmata distributed at random, we obtain a simple mapping formula which gives positive co-incidence within the arms and no co-incidence across the centromere. It is also possible to derive mapping formulae allowing for chromatid interference. The recombination fraction depends simply on the relative excess of pairs of adjacent chiasmata that involve all four chromatids over those that involve only two.

For any given level of chromatid interference the relationship between the distribution of the chiasmata along the chromosome and the mapping formula is many-one so that the second of my practical questions, that of the goodness of fit of the visible 'chiasmata' with the expected cross-overs is not so easy to answer, although it is the more useful question to pose.

Age distributions for Markov processes in genetics

Adv. Appl. Prob.
10, 17-19, 1978.

S. TAVARÉ, *University of Sheffield*

There are several situations in which we might want to estimate the age of a process. In a genetic context, the usual problem is to find the age of an allele, given its current frequency.

There are essentially two different approaches to this problem. We could adopt a statistical approach by forming a likelihood based on our observations, and then estimate the age by, for instance, maximum likelihood. For examples of this method, see Stigler (1970), Thompson (1976). Alternatively, we can define our age in terms of some random variable, and find its distribution. Estimation procedures are then based on this distribution. In what follows, we discuss only the second method.

When the underlying process is a diffusion, Watterson (1976), (1977) exploits the ideas of reversibility to unify some results of Maruyama (1974), Kimura and Ohta (1973), and Levikson (1977). We consider age distributions for Markov chains. Consider a population of fixed size M , in which each individual is of type A or type B at a locus. Let X_n be the number of A individuals at time n . Suppose that $\{X_n, n \geq 0\}$ is a Markov chain, with transition matrix P , state space $\{0, 1, \dots, M\}$, and absorbing barriers at 0 and M . Following

Levikson (1977), construct from $\{X_n\}$ the return chain $\{\bar{X}_n\}$ by returning the X -process to state 1 whenever it hits 0, or to state $M-1$ whenever it hits M . In genetic terms, such a construction might correspond to a (rare) mutation reintroducing a particular allele into the population.

Although Levikson considered such returns to be instantaneous, the return chain is most easily described in terms of the original chain if we consider that it takes one step for such a return to occur. We can now define the age of the X -process by the time since a barrier was last hit. Thus the age at time m , A_m , is defined by

$$\{A_m = n\} \equiv \{\bar{X}_{m-n} = 0 \text{ or } M, \bar{X}_{m-k} \in \{1, \dots, M-1\}, 0 \leq k < n\}.$$

Since $\{\bar{X}_n\}$ is ergodic, we can define a limiting age, A , given 'now' at j , by

$$P(A = n | j) = \lim_{m \rightarrow \infty} P(\bar{X}_{m-n} = 0 \text{ or } M, \bar{X}_{m-k} \in \{1, \dots, M-1\} \text{ for } 0 \leq k < n | \bar{X}_m = j).$$

For the unconditioned chain, let $\pi_i(j)$ be the probability of absorption at state i , given $X_0 = j$, ($i = 0, M$), and let $N = (n_{ij})$ be the mean sojourn time matrix, so that

$$n_{ij} = E(\text{number of visits to } j \text{ before absorption} | X_0 = i).$$

Proofs, based on elementary Markov chain theory, show that

$$P(A = n | j) = \frac{\pi_0(M-1)P_{1j}^{(n-1)} + \pi_M(1)P_{M-1,j}^{(n-1)}}{\pi_0(M-1)n_{1j} + \pi_M(1)n_{M-1,j}}, \quad n \geq 1$$

and that

$$P(A \text{ allele is the oldest} | \text{now at } j) = \frac{n_{M-1,j}\pi_M(1)}{\pi_0(M-1)n_{1j} + \pi_M(1)n_{M-1,j}}.$$

When the return chain $\{\bar{X}_n\}$ is reversible we can connect the mean age with the mean (future) absorption time, and the probability of allele A being the oldest with its (future) fixation probability. Conditions for such reversibility were given, and the results were illustrated by Moran's model (Moran (1958)). Results and problems that arise in the case of one absorbing barrier were also presented.

References

- KIMURA, M. AND OHTA, T. (1973) The age of a neutral mutant persisting in a finite population. *Genetics* **75**, 199-212.

- LEVIKSON, B. (1977) The age distribution of Markov processes. *J. Appl. Prob.* **14**
- MARUYAMA, T. (1974) The age of an allele in a finite population. *Genet. Res. Camb.* **23**, 137-143.
- MORAN, P. A. P. (1958) Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60-71.
- STIGLER, S. M. (1970) Estimating the age of a Galton-Watson branching process. *Biometrika* **57**, 505-512.
- THOMPSON, E. A. (1976) Estimation of age and rate of increase of rare variants. *Amer. J. Hum. Genet.* **28**, 442-452.
- WATTERSON, G.A. (1976) Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. *Theoret. Popn Biol.* **10**, 239-253.
- WATTERSON, G. A. (1977) Reversibility and the age of an allele. II. Two allele models, with selection and mutation. To appear.

Impossible gene identity states

E. A. THOMPSON, *University of Cambridge*

A genealogical relationship between two individuals may be summarized by the nine identity coefficients of Gillois (1965). These coefficients are the probabilities, D_i , of the nine possible states of identity-by-descent between the four genes of the two individuals, at an autosomal locus. Thus the space of relationships may be written as

$$(1) \quad \Delta = \left\{ (D_i; i = 1, \dots, 9); D_i \geq 0, \sum_1^9 D_i = 1 \right\},$$

where we use the ordering of the nine states given (for example) by Jacquard (1974).

In the case of non-inbred individuals only the three last states have non-zero probability, and relationships may then be represented in the space

$$(2) \quad \Delta^* = \{(D_7, D_8, D_9); D_i \geq 0, D_7 + D_8 + D_9 = 1\}.$$

Thompson (1976) has shown that in fact only a fraction, 1/3, of this space can actually be attained by the identity state probabilities since

$$(3) \quad D_8^2 \geq 4D_7D_9.$$

In view of this it is of interest to determine the general space of attainable identity coefficients—a subspace of (1). There are several alternative approaches. One is to consider what classes of states *must* be attainable, the second is to consider what classes of states *cannot* be attainable, and the third is to parametrize certain classes of relationship, and obtain corresponding classes of attainable identity coefficients.