AN URN MODEL AND THE COALESCENT IN NEUTRAL INFINITE-ALLELES GENETIC PROCESSES

David Branson
University of Essex

Abstract

An urn model is proposed as a useful computational device for obtaining results within the context of infinite-alleles genetic processes incorporating selective neutrality. The model is based on Kingman's coalescent, as developed by Watterson and by Donnelly and Tavaré, and mimics the procedure of tracing a sample's ancestry backwards in time, noting the appearance of common ancestors or new mutants. The structure of the model is possessed by a time-inhomogeneous linear birth-and-death model with immigration, by Moran's model and by a class of other models including the Wright-Fisher model in the limit of large population size. Elementary combinatorial arguments connected with the partial or complete emptying of the urn give rise to a number of results which can be interpreted in terms of the allelic composition of genetic samples and populations. The relationship to other urn models is discussed.

1. Introduction. Some years ago Kingman (1982 a,b,c) introduced the fruitful concept of the coalescent in which a sample of genes is taken from a population and its ancestry traced back in time, noting where there are common ancestors, until one reaches a single common ancestor. This idea has been developed by others, in particular by Watterson (1984) and by Donnelly and Tavaré (1986) who consider in detail the consequences of mutation. In tracing the ancestry backwards we may come to a mutant which introduced a new allele into the population and which was therefore the originating ancestor for a family represented in the sample, all members of the family carrying the same allele. We assume mutation is non-recurrent, that is each mutation produces a previously unknown allele and we assume that selection is absent. We further suppose that the organism we are considering is a monoecious haploid reproducing either asexually or by self-fertilization, so that each individual has just one parent and, in the absence of mutation, the genetic composition of an individual is identical to that of its parent.

In Section 2 we describe an urn model which mimics the coalescent and which keeps track of the time-ordering of the various mutant and non-mutant births, should this be of interest. The model gives a method of constructing the jump chain of Donnelly and Tavaré's coalescent with ages and is put forward as