Statistics in Molecular Biology
IMS Lecture Notes – Monograph Series (1999) Volume 33

LIKELIHOODS ON COALESCENTS: A MONTE CARLO SAMPLING APPROACH TO INFERRING PARAMETERS FROM POPULATION SAMPLES OF MOLECULAR DATA

BY JOSEPH FELSENSTEIN, MARY K. KUHNER, JON YAMATO AND PETER BEERLI

Department of Genetics, University of Washington, Box, 7360, Seattle WA 98195-7360

When population samples of molecular data, such as sequences, are taken, the members of the sample are related by a gene tree whose shape is affected by the population processes, such as genetic drift, change of population size, and migration. Genetic parameters such as recombination also affect that genealogy. Likelihood inference of these parameters involves summing over all possible genealogies. There is a vast number of these, so that exact computation is not possible. Griffiths and Tavaré have proposed computing these likelihoods by Monte Carlo integration. Our group is doing this by the Metropolis-Hastings method of Markov Chain Monte Carlo integration. We now have, in our **LAMARC** package, programs to do this for constant-sized and growing populations, and for geographically structured populations. The bias of the estimator of population growth rate is discussed. One can also allow for samples stratified in time, as with fossil DNA or sequential samples from the population of a virus in a patient. A program for recombining sequences is in progress, and we hope to put together an object-oriented environment which can cope with a variety of evolutionary forces.

1. Introduction. Samples of genes from natural populations of organisms are related by a genealogy, which is usually unknown. At the level of the copies of the genes, such a genealogy would specify where each copy of the gene came from. Thus, a particular copy that we sample may have come from the mother of that individual, from her father, from his father, from his mother, and so on, back in time. Other copies are doing the same. As we go back, occasionally two of these lineages will coalesce, as it happens that two copies of a gene are descended from the same parental copy. Thus, my great-great-great-grandmother might happen to be the sibling of your great-great-great-grandfather, and the genes we possess might then turn out to be copied from the same copy in one of their parents. Such coalescences are inevitable in natural populations.

Figure 1 shows such a pattern of ancestry. Each circle is an individual who has two copies of the gene; we are concerned not just with the genealogy of the individuals, but with the genealogy at the gene level. In the figure, time

---

Supported by NIH grant R01 GM-51929 and NSF grant BIR-9527687
AMS 1991 *subject classifications.* Primary 62F03, 92D15; secondary 92D10, 92D20.
Key words and phrases. Coalescent, molecular evolution, genetics.