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

ESTIMATION OF CONDITIONAL MULTILOCUS GENE IDENTITY AMONG RELATIVES

BY ELIZABETH A. THOMPSON AND SIMON C. HEATH

University of Washington

Genetic Analysis Workshop 10 identified five key factors contributing to the resolution of the genetic factors affecting complex traits. These include analysis with multipoint methods, use of extended pedigrees, and selective sampling of pedigrees. By sampling the affected individuals in an extended pedigree, we obtain individuals who have an increased probability of sharing genes identical by descent (IBD) at marker loci that are linked to the trait locus or loci. Given marker data on specified members of a pedigree, the conditional IBD status among relatives can be assessed, but exact computation is often impractical for multiple linked markers on complex pedigrees. The use of Markov chain Monte Carlo (MCMC) methods greatly extends the range of models and data sets for which analysis is computationally feasible. Many forms of MCMC have now been implemented in the context of genetic analysis. Here we propose a new sampler, which takes as latent variables the segregation indicators at marker loci, and jointly updates all indicators corresponding to a given meiosis. The sampler has good mixing properties. Questions of irreducibility are also addressed.

1. Introduction. Relatives share common ancestors. A single gene in such an ancestor may therefore descend via repeated segregations to each of the relatives. Such genes, which are copies of a single ancestral gene within a defined pedigree, are said to be *identical by descent* (IBD). Disregarding mutation, IBD genes must be of like type. It is the sharing of IBD genes that underlies phenotypic similarities among relatives. The probabilities of patterns of gene identity by descent are determined by the pedigree structure, and in turn determine the probability distribution of observed data on individuals of the pedigree.

Genetic linkage is the dependent cosegregation of genes at different loci on the same chromosome. Linkage detection and linkage analysis on the basis of data observed on related individuals require the computation of multilocus probabilities of observed phenotypic data on pedigree structures. Genetic Analysis Workshop 10 identified five key factors contributing to the resolution of the genetic factors affecting complex traits (Wijsman and Amos 1997). These include analysis with multipoint methods, use of extended pedigrees, and selective sampling of pedigrees. Here we consider an approach to linkage detection which uses only data on affected individuals. However, calculation of multilocus probabili-

Work supported in part by NIH grant GM-46255 and NSF grant BIR-9305835.
AMS 1991 *subject classifications.* Primary 62F03 ; secondary 92D10.
Key words and phrases. Markov chain Monte Carlo, meiosis samplers, linkage detection, irreducibility, segregation indicators.