Genetic Analysis Workshop 7: Recent progress in statistical methods

D.T.

Bishop

Genetic Epidemiology Laboratory, Imperial Cancer Research Fund, Leeds (UK)

Introduction

There are a number of important new developments in the statistical analysis of pedigrees that potentially offer the opportunity to expand the repertoire of models that are routinely fitted in segregation and linkage analyses. While the feasibility of these methods is not proven, the excitement comes as much from the fact that epidemiologists, biostatisticians, and computer scientists have recognized that recent advances can be applied in genetic epidemiology as from these specific methods.

The most interesting of these methods is the Gibbs sampler. Gibbs sampling can be thought of as being analogous to the E-M algorithm except that instead of replacing a parameter by its maximum likelihood estimator as would be performed by the E-M algorithm, the parameter is instead replaced by a sample from the appropriate conditional distribution in Gibbs sampling. The result of the application of Gibbs sampling is not merely the value of the estimator, but instead, the appropriate posterior distribution of that parameter. In the approach adopted by Thomas at this workshop, one other interesting twist to the problem is the acknowledgement that, for instance, for a single major locus model, the genotype of each individual is itself a parameter to be estimated. In one cycle of sampling, each parameter is replaced by a sample from its conditional distribution given the value of all of the other parameters (genotypes and model parameters) and all of the observations on the pedigrees. At this workshop, this method was applied to an analysis of the Utah melanoma and nevus counts, but similar interest in these methods has been shown by Thompson and Wijsman (1990) and Sheehan and Thomas (in press). Developments in computer science were emphasized by Szolovits. While computer scientists have discussed probabilistic networks and geneticists have discussed likelihoods on pedigrees, the two problems have been examined separately (apparently to the detriment of both areas). The approach discussed by Szolovits is to use methods for factoring algebraic formulae. To illustrate the formulation, a prototype program GENINFER-II has been developed for the Macintosh. This program is aimed at genetic counselors and neatly combines the 'point and click' friendliness of a Macintosh interface with computational sophistication to allow the counselor to construct pedigrees on the screen and to calculate risks for identified individuals within the pedigree. Pedigree drawing has been a natural focus of attention for computer scientists, but the difficulty of providing a two dimensional representation of a complex pedigree has limited progress. The success of the approach defined by Landre et al. (1972) did not come until Drs. Bennett Dyke, Jean MacCluer, and colleagues adapted the algorithm for the Macintosh (Mamelka et al., 1989). Another approach to pedigree plotting was reported by Round, who has developed an algorithm which is incorporated in his analysis program SCHESIS. This pedigree plotting is based on a two

stage process of developing a conceptual plot based on marriage nodes and then working out optimal representations of the conceptual plot to form the final pedigree. With this approach, which is similar to that presented by Thomas (1988), the number of line crossings can be minimized to provide more readability to the pedigrees.

A number of current linkage analysis problems were also addressed at this meeting. Terwilliger and Ott described a bootstrap approach to the examination of lod scores when the mode of inheritance model is examined jointly with the evidence for linkage. The general approach of joint estimation is appealing since we cannot really expect to know the mode of inheritance precisely in many situations of interest for complex diseases, but the statistical problem of inflating the evidence for linkage by so doing has not been solved. Weeks (1990) described a simulation approach to the determination of critical values for a predefined set of genetic models. Terwilliger and Ott have applied resampling methods to the same problem to approximate the results with much less intensive computing.

Vieland et al. examined single locus linkage analysis of diseases which are in fact determined by two loci. The results are reassuring since, at least in the situations examined, while the estimates of recombination were biased, the overall evidence for linkage was not seriously damaged. The generalization of these results to more complex models is an important topic for further research since, as elucidated by Risch (1990), there seems to be evidence of interacting loci in various disease etiologies. Finally, Lewis calculated expected lod scores (ELODs) by mating types for three loci. These calculations are important for locus ordering and risk calculations for disease loci where the polymorphism and availability of markers do not always allow definitive information. With the formulae contained in Lewis, analytical examinations of ELODs can be performed without recourse to numerical methods.

References

Landre M-F, Valat M-T, Jutier P: Reconnaissance automatique des liens de parente. Trace automatique d'arbre genealogique. Mathematique, Informatique and Statistique Bio-Medicales:10 (Centre de Calcul et de Statistique des Facultes de Medecine de Paris, Paris, 1972). Mamelka PM, Dyke B, MacCluer JW: Pedigree/Draw for the Apple Macintosh. PGL Technical Report No. 1 (Population Genetics Laboratory, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, 1989).

132

Bishop

Risch N: Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222-228 (1990).

Sheehan NA, Thomas A: On the irreducibility of a Markov chain defined on the genotypes of a pedigree by a sampling scheme. Biometrics (in press).

Thomas A: Drawing pedigrees. IMA Journal of Mathematics Applied in Medicine and Biology 5:201-213 (1988).

Thompson EA, Wijsman EM: The Gibbs Sampler on extended pedigrees: Monte Carlo methods for the genetic analysis of complex traits. Technical Report