Statistical Science
1994, Vol. 9, No. 3, 305–306

In This Issue

GENETICS AND MOLECULAR BIOLOGY

The revolution in biology that followed the elucidation of the structure of DNA by Watson and Crick in 1953, and the development of recombinant DNA techniques in the 1970s, has led to a wide range of new questions being presented to statisticians. A special feature of this issue is a group of five papers describing some statistical aspects of modern genetics and molecular biology.

The paper, by W. Navidi and N. Arnheim, concerning the polymerase chain reaction (PCR), is a fitting introduction. This reaction—whose inventor, Kary B. Mullis, was awarded the 1993 Nobel prize in chemistry—is now used routinely to produce data of the kind discussed by the other authors, and is creating its own revolution within biology. An understanding of PCR is an essential prerequisite for any statistician wishing to get involved in this area, and the explanation of it by Navidi and Arnheim also includes a discussion of errors in the reaction. They go on to review some of the innovative applications of PCR to genetic mapping and genetic disease diagnosis in whose development they played a key role.

The U.S.-initiated but now essentially worldwide Human Genome Project has as its aim the sequencing of the entire human genome. Two intermediate goals of the project are production of high-resolution genetic and physical maps of all the human chromosomes. Genetic maps are discussed by Navidi and Arnheim, and also by E. A. Thompson, and it is in the review by David Nelson and Terence Speed that the notion of physical map is explained. In order to grasp just what physical maps are, why they are needed, and the possible use of statistics in their construction, it is necessary to have met the concepts and terminology of recombinant DNA procedures. Our authors provide this, as well as a review of the statistical and probabilistic techniques that have proved useful in one of the major U.S. physical mapping projects.

The remaining three papers describe novel statistical methods for analysing genetic map and DNA sequence data. Thompson reviews likelihood methods for performing linkage analysis using human pedigree data, and then focuses on what has come to be known as homozygosity mapping: linkage mapping of recessive diseases using inbred affected individuals. She describes a variety of Markov chain Monte Carlo approaches to linkage mapping with pedigree data, and goes on to give a detailed analysis using the approach which is particularly suited to homozygosity mapping.

Michael Waterman and Martin Vingron are concerned with DNA and protein sequence comparison, the querying of a database to locate one or more sequences which match a given query sequence more than might be expected by chance. After going over the basic algorithms used to compare DNA or protein sequences, they review the theory available and then go on to extend Poisson approximation techniques to obtain a practical method for estimating the significance of sequence alignment scores.

The final paper in our group, by R. C. Griffiths and Simon Tavaré, gives readers a glimpse of the wide range of questions concerning our species that can be addressed using DNA sequences from a sample of individuals. Many will have heard about "mitochondrial Eve," the most recent common female ancestor of us all, at least as far as the maternally inherited mitochondrial genome is concerned. "Her" place in space and time was estimated from a sample of mitochondrial sequence data not unlike that discussed by Griffiths and Tavaré. However, unlike some earlier and now discredited analyses, the approach of these authors is based on probabilistic models that attempt to embody the common ancestry of the individuals sampled and aspects of the substitution process. Their novel Markov chain simulation method should also have wide appeal, as it is applicable beyond the present context.

THE BOOTSTRAP

The paper by G. Alastair Young is a critical essay on the evolution of bootstrap research, using as a springboard his reactions to several books that have now appeared on the subject. Emphasizing the need for investigations of direct relevance to statistical practice, Young succinctly summarizes the bootstrap paradigm and its accompanying strong body of theoretical work, then indicates gaps in our knowledge, voices worries about misuse of this powerful technology, and points toward directions for future research effort. Commentary is provided by Rudolf Beran; Bradley Efron; Patricia Grambsch, Mary Kathryn Cowles and Thomas Louis; D. V. Hinkley; Michael Meredith and Jorge Morel; William Navidi; and Mark Schervish. Many of these discussants share Young's concerns and elaborate with additional examples; Beran indicates disagreements; and Efron