Computational Algebraic Statistics

organized by
Jesus A. De Loera, Steven Fienberg, Serkan Hosten, Bernd Sturmfels, and Alan Karr

Workshop Summary

This workshop helped define an exciting new subject at the interface of mathematics and statistics. Algebraic statistics is concerned with applying tools from commutative algebra and algebraic geometry to statistical problems. The key idea is that many statistical models for discrete data are actually algebraic varieties, which makes them amenable to the tools of computational algebra, such as Gröbner bases.

The collection of participants at this workshop was a mix of mathematicians and statisticians, many of them quite young, and a substantial amount of time was spent simply establishing a common language and defining key problems. The program involved both traditional talks and intensive working sessions. Persi Diaconis and Steve Fienberg gave introductory lectures on the first morning. Further one-hour lectures were given by Akimichi Takemura, Russell Steele, Frantisek Matus, Elizabeth Allman, Seth Sullivant, Henry Wynn, Donald Richards, and Jesus De Loera. A group of four graduate students presented their research in shorter talks. A few additional research talks were given by others.

Particularly successful were the working sessions: Serkan Hosten and Chris Meek led a session on Graphical Models; Inga Hreinsdottier and Lior Pachter led a session on Computational Biology; Ian Dinwoodie and Mike Stillman led a session on Software; and Reinhard Laubenbacher and Eva Riccomagno led a session on Linear Polynomial Models. Several of the open questions discussed at these sessions were actually answered during the conference. For instance, Berkeley graduate student Nick Eriksson solved two problems, one posed by Persi Diaconis on permutation data and one posed by Steve Fienberg on three dimensional tables. This will lead to two research papers to be submitted to the special volume mentioned below. Joseph Landsberg, a differential geometer, started a very interesting collaboration with Elizabeth Allman and John Rhodes on the defining equations of statistical models for phylogenetic trees, and Bernd Sturmfels started a collaboration with Inga Hreinsdottir on two-locus genetic linkage models. During a final working session on maximum likelihood estimation (MLE) many of us realized that understanding the algebraic structure of MLE and the EM-algorithm is a next big challenge.

A concrete outgrowth of this workshop will be a special issue of the Journal of Symbolic Computation (JSC) on "Algebraic Statistics". The Editor-in-Chief of JSC, Hoon Hong, was enthusiastic about this plan and he appointed Serkan Hosten and Chris Meek to act as guest editors. The submission deadline is May 1, 2004, and we are expecting many good papers, mostly from conference participants but also a few others.

In summary, many wonderful things are happening in Algebraic Statistics, and many people in applied areas (such as machine learning, bioinformatics etc.) are beginning to learn about the new algebraic tools. Conversely, statistics provides many good open problems and truly new directions for inquiry for algebraic geometers. The AIM conference provided a

unique opportunity to make these connections and it will have a lasting impact for many years to come.