Ruriko Yoshida (Kentucky)

Markov Bases for the Toric Homogeneous Markov Chain Models

Abstract for the Colloquium 2014 December 2

Discrete time Markov chains are often used in statistical models to fit the observed data from a random physical process. Sometimes, in order to simplify the model, it is convenient to consider time-homogeneous Markov chains, where the transition probabilities do not depend on the time T. While under the time-homogeneous Markov chain model it is assumed that the row sums of the transition probabilities are equal to one, under the toric homogeneous Markov chain (THMC) model the parameters are free and the row sums of the transition probabilities are not restricted.

In order for a statistical model to reflect the observed data, a goodness-of-fit test is applied. For instance, for the time-homogeneous Markov chain model, it is necessary to test if the assumption of time homogeneity fits the observed data. In 1998, Diaconis–Sturmfels developed a Markov Chain Monte Carlo method (MCMC) for goodness-of-fit test by using Markov bases. A Markov basis is a set of moves between elements in the conditional sample space with the same sufficient statistics so that the transition graph for the MCMC is guaranteed to be connected for any observed value of the sufficient statistics. In algebraic terms, a Markov basis is a generating set of a toric ideal defined as the kernel of a monomial map between two polynomial rings. In algebraic statistics, the monomial map comes from the design matrix (configuration) associated with a statistical model.

In this talk we will consider a Markov basis and a Groebner basis for the toric ideal associate with the design matrix defined by the THMC model with $S \geq 2$ states without initial parameters for any time $T \geq 3$. First we will show the upper bound of the Markov degree, the degree of a minimal Markov base, of the THMC model with $S=3$ for $T \geq 3$. In order to compute the upper bound, we use the model polytope—the convex hull of the columns of the design matrix. Here we will show the model polytope has only 24 facets for $T \geq 5$ and a complete description of the facets for $T \geq 3$. Finally, we will show a condition when the THMC with any $S \geq 2$ states for $T \geq 3$ has a square-free quadratic Groebner basis and Markov basis. One such example is the embedded discrete Markov chain (jump chain) of the Kimura three-parameter model.

This is joint work with Davis Haws (IBM), Abraham Martin del Campo (IST Austria), and Akimichi Takemura (University of Tokyo).
