Non-parametric Bayesian inference for Y-linked bisexual branching processes through MCMC methods

Miguel González, Cristina Gutiérrez, Rodrigo Martínez, Manuel Mota

1 Department of Mathematics, University of Extremadura. 06071 Badajoz, SPAIN.
E-mails: mvelasco@unex.es; cgutierrez@unex.es; rmartinez@unex.es; mota@unex.es

Keywords: sex-linked genes, two-sex stochastic models, non-parametric Bayesian inference, Gibbs sampler

It is well-known that in some animal populations the sex of the individuals is determined by a pair of chromosomes X and Y. A female has XX chromosomes, while a male has XY chromosomes. Certain characteristics are due to genes carried on the X chromosome (X-linked). Others due to genes carried on the Y chromosome (Y-linked) and still others to genes carried on both chromosome (XY-linked). Taking into account this fact, females and males with different genotypes and/or phenotypes appear in the population. Females and males in a generation form mating units in order to produce offspring. Each mating unit has a genotype which is formed from the genotypes of the two individuals in the unit. Finally, an individual offspring will receive its genetic profile as specified by the inheritance rules associated with that species.

From a practical viewpoint, it is of interest to model and analyze the evolution of sex-linked genes from generation to generation. Recently, González et al. (2006) have introduced a two-dimensional bisexual branching process to analyze the evolution of Y-linked genes, where preference of females for males with a specific genetic characteristic determined by an allele of the gene is assumed.

The behaviour of these kinds of Y-linked genes is strongly related to the reproduction law of each genotype. In practice these offspring distributions are usually unknown and their estimation is necessary. Until now, the whole family tree up to a given generation must be available to obtain accurate estimators (see González et al. (2008)). However, in most of the populations it is not possible to observe it and only the total number of individuals in each generation is available. In this work, we deal with the estimation problem for the offspring distribution of each genotype of a Y-linked gene when the only data we suppose available is the total number of males of each genotype and the total number of females in each generation. We set out the problem from a Bayesian outlook in a non-parametric framework. We provide an algorithm, based on the Gibbs sampler and gaussian kernel density estimators, to approach the posterior distribution of the reproduction law of each genotype. From the posterior density estimates, we also obtain HPD (high posterior density) credible regions for the reproduction mean of each genotype and we approximate the posterior predictive distribution of the future population size. Finally, by way of several simulated examples we illustrate the accuracy of the algorithm introduced. These examples have been developed with the statistical software R.

Acknowledgement

The research was supported by the Ministerio de Educación y Ciencia and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2006-08891.

References
