Sampling Ecological Occurrence Matrices

Koko K. Kayibi, Muhammad Ali Khan, S. Pirzada

Ecological occurrence matrices, such as Darwin finches tables, are $0, 1$-matrices whose rows are species of animals and colums are islands, and the (i, j) entry is 1 if animal i lives in island j, and is 0 otherwise. Moreover the row sums and columns sums are fixed by field observation of these islands. These occurence matrices are thus just bipartite graphs G with a fixed degree sequence and where $V_1(G)$ is the set of animals and $V_2(G)$ is the set of islands. The problem is, given an occurrence matrix, how to tell whether the distribution of animals is due to competition or to chance. Thus, researchers in Ecology are highly interested in sampling easily and uniformly ecological occurrence tables so that, by using Monte Carlo methods, they can approximate test statistics allowing them to prove or disprove some null hypothesis about competitions amongs animals.

Several algorithms are known to construct realizations on n vertices and m edges of a given degree sequence, and each one of them has its strengths and limitations. Most of these algorithms can be fitted in two categories: MonteCarlo Markov chains methods that are based on *edge-swappings*, and *sequential sampling* methods that are based on starting from an empty graph on n vertices and adding edges sequentially according to some probability scheme. We present a new algorithm that samples uniformly all simple bipartite realizations of a degree sequence and whose basic ideas may be seen as implementing a *dual sequential method*, as it inserts sequentially vertices instead of edges.

The running time of our algorithms is $\mathcal{O}(m)$ where m is the number od edges in any realization. The best algorithms that we know of are the one implicit in [1] that has a running time of $\mathcal{O}(ma_{max})$ where a_{max} is the maximum of the degrees, but is not uniform. Similarly, the algorithm presented by Chen and all [3] does not sample uniformly, but nearly uniformly. Moreover the edge-swapping Markov Chains pionneered by Brualdi [2] and Kannan and all in [5], and much used by reseachers in Ecology, have just been proven in [4] to be fast mixing for semi-regular degree sequences only.

References

[1] M. Bayati, J. H. Kim, and A. Saberi, A sequential algorithm for generating random graphs, Algorithmica 58(2010), 860–910.

[2] R. A. Brualdi, Matrices of zeroes and ones with fixed row and column sum vectors. Lin. Algebra Appl. 33(1980), 159–231

[3] Y. Chen, P. Diaconis, S. Holmes, J. S. Liu, Sequential Monte Carlo methods for statistical analysis of tables, J. Am. Stat. Assoc. 100 (2005), 109–120.

[4] P. L. Erdös, I. Miklós and Z. Toroczkai, A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs, Electron. J. Combin. 17(1) (2010) #R66.

[5] R. Kannan, P. Tetali and S. Vempala, Simple Markov-chain algorithms for generating bipartite graphs and tournaments, Random Struct. Algorithms 14(4) (1999), 293–308.