On modeling and sampling complex systems

Matteo Marsili

The study of complex systems is limited by the fact that only few relevant variables are accessible for modeling and sampling. In addition, empirical data most often undersample the space of possible states. We discuss the consequences of this in a simple framework inspired by maximum entropy considerations. Our arguments suggest that models can be predictable only when the number of relevant variables is less than a critical threshold. Within our framework, the undersampling regime can be distinguished from the regime where the sample becomes informative of system's behavior. In the undersampling regime, the most informative frequency size distributions have power law behavior and Zipf's law emerges at the crossover between the undersampled regime and the regime where the sample contains enough statistics to make inference on the behavior of the system. These ideas are illustrated in some applications, showing that they can be used to identify relevant variables or to select most informative representations of data, e.g. in data clustering.

Proceedings 27th European Conference on Modelling and Simulation ©ECMS Webjørn Rekdalsbakken, Robin T. Bye, Houxiang Zhang (Editors)
ISBN: 978-0-9564944-6-7 / ISBN: 978-0-9564944-7-4 (CD)