



Estimating High-Dimensional Autoregressive Point Processes

Rebecca Willett

Abstract

Vector autoregressive models characterize a variety of time series in which linear combinations of current and past observations can be used to accurately predict future observations. For instance, each element of an observation vector could correspond to a different node in a network, and the parameters of an autoregressive model would correspond to the impact of the network structure on the time series of observations at each network node. Of particular interest are autoregressive point processes, in which observations consist of the times at which each node participates in some event or activity. Such data is common in spike train observations of biological neural networks, interactions within a social network, and pricing changes within financial networks. However, very little is known about how many events must be recorded before we may accurately infer the underlying autoregressive models. I will describe sparsity-regularized methods and associated performance bounds which provide new insight into the sample complexity of these problems in high dimensions. While sparsity-regularization is well-studied in the statistics and machine learning communities, common assumptions from that literature (such as the restricted eigenvalue condition) are difficult to verify in this setting because of the correlations and heteroscedasticity of the observations. A novel analysis method leveraging a combination of Martingale concentration inequalities and high-dimensional linear regression characterizes how much data must be collected to ensure reliable inference depending on the size and sparsity of the autoregressive parameters, and these bounds are supported by several empirical studies. *This is joint work with Eric Hall and Garvesh Raskutti.*

E. Hall, G. Raskutti, and R. Willett, “Inference of High-dimensional Autoregressive Generalized Linear Models.” arXiv:1605.02693, 2016.