

Smoothing discrete distributions

Paulo Eduardo Oliveira

University of Coimbra, Portugal

Abstract

We will discuss estimation of probability distributions on discrete, finite or infinite, space using nonparametric methods. This model includes of course, categorical distributions. Although the smoothing implied in nonparametric methods may seem, at first glance, unnatural, smoothing does improve upon the naïve frequency estimator. Discretizations of the kernel estimator and the correspondent characterizations of asymptotic properties are discussed.

When dealing with categorical distributions one is often faced with relatively few observations, when compared to the support size. This leads to considering error criteria better adapted to this sparse estimation problem. Asymptotics with respect to these sparse criteria is discussed. These results do not really fall into the general approach to nonparametric estimation, as they imply that the base space should be updated as the sample size grows.

Other error criteria, such as relative errors, are commonly considered in parametric problems. We will adapt relative error criteria to our nonparametric estimation problem. The estimator found can be explicitly written but their asymptotics is harder to describe, in some cases only doable indirectly. However, their finite sample performance is, depending on the properties of the true probability distribution, good. We will also discuss the integration into the estimator of partial known information about the true probability.

Keywords

Discrete distributions, Local polynomial estimator, Relative errors, Asymptotics, Sparse observations.