

Seminario Aleatorio

Sesión 416

Structure learning of graphical models for count data, with applications to single-cell RNA sequencing

Thi Kim Hue Nguyen

Department of Statistical Sciences, University of Padova

Abstract

The problem of estimating the structure of a graph from observed data is of growing interest in the context of high-throughput genomic data and single-cell RNA sequencing in particular. These, however, are challenging applications since the data consists of high-dimensional counts with high variance and over-abundance of zeros. Here, we present general frameworks based on the zero-inflated negative binomial distribution for learning the structure of a graph from single-cell RNA-seq data. In detail, we employ a variant of the PC algorithm with a sequence of deviance test statistics. We demonstrate with simulations that our approaches are able to retrieve the structure of a graph in a variety of settings, and we show the utility of the approach on the olfactory epithelium data.

Paper published in The Annals of Applied Statistics

DOI: 10.1214/23-AOAS1732

**Viernes 20 de octubre de 2023,
13:00 horas de CDMX**

Transmisión en línea:

<https://itam.zoom.us/j/94446870531?pwd=cnBuaklwWDFmR2lvZkxLSU5hcFQyUT09>

Meeting ID: 944 4687 0531

Passcode: 774808

El Seminario Aleatorio del Departamento de Estadística del ITAM está destinado tanto a profesores como a estudiantes, por lo que se agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.