Project Proposal: Clustered Poisson Process

Xiaokang Liu

Method

Clustered spatial point processes are popular models for spatial point pattern data that contain clusters of events. For example, this model can be used to characterize the distribution of one species of trees in a forest. In general, the clustered Poisson process is consists of a parent process and a child process. For example, Neyman-Scott process is consists of a Poisson process generating events at location as c with constant intensity function $\kappa > 0$ (parent process) and a poisson process Y_c (child process) with inhomogeneous intensity function $\lambda_c(s; \beta, w) = h(s - c; w) \exp\{x(s)'\beta\}$, where $h(s - c; w)$ is a density function parameterized by w, with $x(s) = (x_1(s), \dots, x_p(s))$ representing a p-dimensional location related covariates vector and regression coefficient vector $\beta = (\beta_0, \dots, \beta_p)$. The superposition of the child processes $Y = \cup_{c \in C} Y_c$ defined a Neyman-Scott process with the intensity function $\lambda(s; \beta) = \exp\{x(s)'\beta^*\}$, where $\beta^* = (\log(\kappa) + \beta_0, \dots, \beta_p)'$. In some of such studies, p maybe very large, thus it is of interest to identify important factors underlying such spatial point patterns. Andrew et al. (2015) proposed a regularized method to solve the variable selection problem in clustered spatial point processes, which minimizes a combination of a goodness of fit term and a model complexity penalization term. In their paper, instead of using Monte Carlo methods to approximate the maximum likelihood estimates, estimates were obtained by solving the estimating equation, which is obtained by getting the first order derivative of the log-likelihood function for the Poisson process. This estimation procedure can greatly improve computation efficiency. Further, for incorporating the information about the interaction of events, the weighted quasi-log-likelihood function is used to measure goodness of fit and leads to weighted estimating equation. For the regularization, an adaptive Lasso penalty is used, which can compel some small coefficients to be exact zero thus conduct variable selection.

In this project, we will realize the algorithm used to solve this estimation problem by R. Also, we try to achieve the generalization from modeling one single clustered Poisson process to simultaneously model several process together to take into consideration the potential correlation between different species. For example, there maybe coexistence or compulsion effect among different tree species. We want to leverage strength from such kind of relationship between species to improve variable selection. For achieving this goal, some other regularization methods are considered, like group Lasso penalty to achieve row-sparsity of the coefficient matrix.

Application

We are going to apply the multivariate point process to model the distribution of several species of trees in a forest plot on the barro Colorado Island. In this dataset, the 50-hectare permanent tree plot was established in 1980 in the tropical moist forest of Barro Colorado Island (BCI) in Gatun Lake in central Panama. All free-standing woody stems at least 10 mm diameter at breast height were identified, tagged, and mapped. Over 350,000 individual trees have been censused over 35 years. We will fit the model to some selected tree species, including B. pendula, O. whitei and P. panamense these three species. The considered location related factors including elevation, slope and 13 soil characteristics.