Project Proposal: Clustered Poisson Process

Xiaokang Liu

Method

Clustered spatial point processes are popular tools to model clustered events in spatial point pattern data. In general, the clustered Poisson process consists of a parent process and a child process. For example, the Neyman-Scott process consists of a Poisson process generating events at location c with a 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 $x(s) = (x_1(s), \ldots, x_p(s))$ representing a p-dimensional location related vector of covariates, and $\beta = (\beta_0, \ldots, \beta_p)$ regression the vector of coefficients. The superposition of the child processes $Y = \cup_{c \in C} Y_c$ defined a Neyman-Scott process with intensity function $\lambda(s; \beta) = \exp\{x(s)'\beta^*\}$, where $\beta^* = (\log(\kappa) + \beta_0, \ldots, \beta_p)'$. In some studies, p maybe very large, thus it is of interest to identify important factors underlying such spatial point patterns. Thurman 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, formulated 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 the goodness of fit and leads to a weighted estimating equation. For the regularization, an adaptive Lasso penalty is used.

In this project, we will realize the algorithm used to solve this estimation problem using R. Also, we try to achieve the generalization from modeling one single clustered Poisson process to simultaneously model several process in order 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 between-species relationship to improve variable selection accuracy. Toward 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 (BCI). In this dataset, the 50-hectare permanent tree plot was established in 1980 in the tropical moist forest of 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. The considered location related factors including elevation, slope and 13 soil characteristics.

Thurman, Andrew L, Rao Fu, Yongtao Guan, and Jun Zhu. 2015. "Regularized Estimating Equations for Model Selection of Clustered Spatial Point Processes." *Statistica Sinica*. JSTOR, 173–88.