Local Spatiotemporal Point Process Regression

Yao Xie

January 2020

1 Formulation

Consider the setting where we have N spatial observations. Assume the spatial regions are discretized. We observe counts y_i for the ith region, $i = 1, \ldots, N$, and in the period from $[0, T]$. Suppose we observe the covariates $x_i \in \mathbb{R}^p$ in the ith region during the same period. There are p covariates. Assume the location (or centriod) of the ith location is given by z_i.

Now we would like to build a model to predict the future intensity. We will model the intensity of the point processes are Poisson process. Assume the intensity at the ith location is given by λ_i. Also assume the intensity may depend on the covariates:

$$\lambda_i = \mu_i + \beta_i^T x_i, \tag{1}$$

where μ_i is the base intensity for the ith region, and $\beta_i \in \mathbb{R}^p$ denotes the p-dimensional model coefficient for the ith region.

Assume the observations y_i are independent conditioning on (μ_i, β_i). Then we can write the likelihood function as

$$\prod_{i=1}^{N} e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!}$$

After taking log and perform simplification, we obtain the negative log-likelihood function as a function of the parameters given by

$$\ell(\lambda_1, \ldots, \lambda_N) = \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} y_i \log \lambda_i.$$

where λ_i are specified in (1) and we have drop the constant term from the log-likelihood function since they do not matter for the maximization.

Since the model above clearly has many parameters (Np parameters). To fit model while using data more efficiently, we will make some additional assumption that there are similarities between the models in the neighborhood. We assume the difference between the coefficients in neighborhoods are proportional to their distance:

$$\|\beta_i - \beta_j\|_2^2 \propto d(z_i, z_j)/\alpha$$

for some $\alpha > 0$; where $d(\cdot, \cdot)$ is a distance measure between two locations (e.g., Euclidean distance).

Based on this, we can formulate the regularized maximum likelihood estimate for the model as

$$\begin{aligned} \min_{\{\lambda_i, \mu_i, \beta_i\}, i=1,\ldots,N} \quad & \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} y_i \log \lambda_i + \alpha \sum_{i \neq j} \frac{1}{d(z_i, z_j)} \|\beta_i - \beta_j\|_2^2 \\ \text{subject to} \quad & \lambda_i = \mu_i + \beta_i^T x_i, \quad i = 1, \ldots, N \end{aligned} \tag{2}$$

Note that this is a convex optimization problem.

1.1 Possible Bayesian formulation

Assume the $\{\mu_i\}$, $i = 1, \ldots, N$ are Beta(a, b) distribution, which is a conjugate prior for Poisson distribution. From this we may obtain further interpretation of the baseline intensity. This is related to the so-called "FAB inference" by Hoff, Yu, Burris.