Bayesian Marked Spatial Point Process with Applications to Professional Basketball

Project Proposal

Jieying Jiao, Guanyu Hu

October 5, 2018

Basketball data records the shoting data for each basketball player. This includes the shoting location and time information, and indicator showing made or miss. Previous studies focused on modeling the location using spatial point process. Here we want to consider the accuracy of shot, i.e, whether this shot is made or miss. So we decide to use marked spatial point process.

Possion process is a widely used parametric distribution for a finite set of count variable. We use it to model the location part of shot. A poisson process $\boldsymbol{Y} = (y(s_1), ..., y(s_n))$ is a random countable subset on $\mathcal{B} \subseteq \mathcal{R}^2$, where $s_i \in \mathcal{B}, i = 1, \cdots, n$. Then for any areal $A \subseteq \mathcal{B}$, we define the counting process $N_{\boldsymbol{Y}}(A) = \sum_{i=1}^{n} 1(s_i \in A)$. If the poisson process $\boldsymbol{Y}$ has intensity function $\lambda(\boldsymbol{s})$, then $N_{\boldsymbol{Y}}(A) \sim Poi(\lambda(A))$, where $\lambda(A) = \int_A \lambda(s)ds$. Then the log-likelihood function on A is $l = \sum_{i=1}^{n} \log\lambda(s_i) - \int_A \lambda(s)ds$. Usually, we often has following regression form for the spatially varying intensity: $\lambda(s_i) = \lambda_0 \exp(\boldsymbol{\beta X_1}(s_i))$, where λ_0 is the baseline of intensity function, $\boldsymbol{\beta}$ is regression coefficient, and $\boldsymbol{X_1}(s_i)$ is spatially varying covariates.

For the marks part, we have binary variable, i.e, made or miss. So we decide to use logistic regression to model marks, i.e, $m(s) \sim Ber(p(s))$, $logit(p(s)) = \alpha_0\lambda(s) + \boldsymbol{\alpha X_2}(s)$, where $\alpha_0, \boldsymbol{\alpha}$ are regression coefficients, $\boldsymbol{X_2}(s_i)$ is spatially varying covariates. This is called intensity dependent model.

Then we have the joint model likelihood for marked point process:

$$L = \prod_{i=1}^{n} p(m(s_i)|\lambda(s_i)) \times \prod_{i=1}^{n} \lambda(s_i) \times \exp\left(-\int_A \lambda(s)ds\right)$$

The prior distribution can be as follow: $\lambda_0 \sim \mathrm{G}(a_1, b_1), \boldsymbol{\beta} \sim \mathrm{MVN}(0, \sigma^2 I), \sigma^2 \sim \mathrm{IG}(a_2, b_2), \boldsymbol{\alpha} \sim \mathrm{MVN}(0, \delta^2 I), \tau^2 \sim \mathrm{IG}(a_3, b_3), \delta^2 \sim \mathrm{IG}(a_4, b_4)$.

We also plan to conduct Bayesian model selection for real data and simulation study.