STAT 534 - Lecture 24: Key

Modeling Point Patterns using NHPPs

- Statistical modeling depends on a sampling model for the data and the associated likelihood function.

- *Conditional on the number of points, $Num(D) = n$, the location density can be specified as:*

$$f(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n | Num(D) = n) = \prod_i \frac{\lambda(\boldsymbol{s}_i)}{(\lambda(D))^n}$$

- Then condidering the location and number of points simultaneously, *we have*

$$f(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, Num(D) = n) = \prod_i \frac{\lambda(\boldsymbol{s}_i)}{(\lambda(D))^n} \times (\lambda(D))^n \frac{\exp(-\lambda(D))}{n!}$$

- *Finally, the likelihood is*

$$\mathcal{L}(\lambda(\boldsymbol{s}), \boldsymbol{s} \in D; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n,) = \prod_i \lambda(\boldsymbol{s}_i) \exp(-\lambda(D))$$

- Alternatively, consider a fine partition for D, then using the Poisson assumption, the likelihood is the product of the counts across the partitions.

$$\prod_l \exp(-\lambda(A_l))(\lambda(A_l))^{Num(A_l)}$$

  As the grid becomes finer, $Num(A_l) \in \{0, 1\}$.

- *The likelihood is a function of the entire intensity surface* $\lambda(\boldsymbol{s})$*. Hence, a functional description of* $\lambda(\boldsymbol{s})$ *is necessary.*

- **Discuss**: how could $\lambda(\boldsymbol{s})$ be obtained?

- *There are two general approaches for modeling* $\lambda(\boldsymbol{s})$*: parametric and non-parametric.*

- A simple, but "silly" example of a parametric form would be $\lambda(\boldsymbol{s}) = \sigma\lambda_0(\boldsymbol{s})$. *Realistically,* $\lambda_0(\boldsymbol{s})$ *would not be available.*

- Suppose remoted-sensed covariate information is available on a grid. How might this be used for constructing $\lambda(\boldsymbol{s})$? *Markov random field properties (Areal data) could be used.*

- In general a parametric function $\lambda(\boldsymbol{s};\theta)$ could be specified, but it would required a richly specified class that would need to be non-negative. A common approach for this type of problem is to use a set of basis functions such that $\lambda(\boldsymbol{s};\theta) = \sum_k a_k g_k(\boldsymbol{s})$.

- Sketch a set of basis functions to create a multimodal curve in 1D.

- *Sometimes a trend surface (as a function of latitude and longitude) can be used to specify* $\lambda(\boldsymbol{s})$

- The most common approach is to specify $\log(\lambda(\boldsymbol{s})) = X^t(\boldsymbol{s})\gamma$. *In otherwords, spatial covariates are used in a regression model for the intensity surface.*

- Specifying covariates still requires integrating $\int_D \exp(X^t(\boldsymbol{s})\gamma)d\boldsymbol{s}$, which is challenging as $X^t(\boldsymbol{s}$ is not often specified in a functional form.

- *For a fine resolution, and assuming constant intensity,* $\int_B \exp(X^t(\boldsymbol{s})\gamma)d\boldsymbol{s} = |B|\exp(X^t(\boldsymbol{s})\gamma)$

- Now, considering $\lambda(\boldsymbol{s})$ through the non-parametric lens, *let*

  $$\lambda(\boldsymbol{s}) = g(X^t(\boldsymbol{s})\gamma)\lambda_0(\boldsymbol{s})$$

  where $g() \geq 0$ *and* $\lambda_0(\boldsymbol{s})$ *as the error process (with mean 1). This is referred to as a Cox process.*

- Suppose $\lambda(\boldsymbol{s}) = \exp(Z(\boldsymbol{s}))$ where $Z(\boldsymbol{s})$ is a realization from a spatial Gaussian process with mean $X^t(\boldsymbol{s})\gamma$ and covariance function $\sigma^s\rho()$. *Then this is referred to as a Log Gaussian Cox Process (LGCP).*

- The LGCP provides a prior for $\lambda(\boldsymbol{s})$. This can be written as $X^t(\boldsymbol{s})\gamma + w(\boldsymbol{s})$ where $w(\boldsymbol{s}) = \log\lambda_0(\boldsymbol{s})$.

- **Discuss**: How to fit a model for a LGCP?

- In the parametric setting, *inference proceeds using* $\mathcal{L}(\theta|\boldsymbol{s}_\infty, \ldots, \boldsymbol{s}_\backslash)$ *in the normal manner.*

- In the non-parametric setting, $\lambda(\boldsymbol{s}) = \exp(Z(\boldsymbol{s}))$, *where* $Z(\boldsymbol{s}) = X^t(\boldsymbol{s})\gamma + w(\boldsymbol{s})$ *is a realization of a GP (GP prior on* $\log \lambda_D$*)*

- *Thus priors are needed for the covariance function of* $Z(\boldsymbol{s}$*, likely* σ^2 *and* ϕ*.*

-In general algorithms for this type of data are computationally intensive.

- Due to the challenges in modeling LGCP, the data is sometimes considered on a grid. Then the intensity for grid i could be estimated as:
$$\log \lambda_i = X_i^t \beta + \phi_i.$$
However, the results can be sensitive to the grid specification.