Math 12230: Spatio-Temporal Statistics for Big Data Midterm Exam 1
*Due*: November 18, 2015 in class

1. (10 points) Show that the variogram $2\gamma(\cdot)$ defined by

$$2\gamma(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Var}\big(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\big), \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d,$$

satisfies the following properties:

(a) $2\gamma(\cdot)$ is conditionally negative definite, i.e.,

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \le 0,$$

for all $\mathbf{s}_i \in \mathbb{R}^d$ and $a_i \in \mathbb{R}$ with $\sum_{i=1}^{m} a_i = 0$.

(b) For any $c > 0$, $e^{-c\gamma(\cdot)}$ is positive definite, i.e.,

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j e^{-c\gamma(\mathbf{s}_i - \mathbf{s}_j)} \ge 0,$$

for all $\mathbf{s}_i \in \mathbb{R}^d$ and $a_i \in \mathbb{R}$.

2. (10 points) Let $\mathbf{Z} = \big(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\big)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, $\mathbf{B} = (b_{ij})$ with $b_{ii} = 0$, $\mathbf{C} = (c_{ij})$ with $c_{ii} = 0$, $\boldsymbol{\Lambda} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\mathbf{M} = \mathrm{diag}(\tau_1^2, \ldots, \tau_n^2)$.

(a) Show that the simultaneously specified Gaussian model

$$Z(\mathbf{s}_i) = \mu_i + \sum_{j=1}^{n} b_{ij}\big(Z(\mathbf{s}_j) - \mu_j\big) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$, implies $\mathbf{Z} \sim N_n\big(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B}^T)^{-1}\big)$, provided $\mathbf{I} - \mathbf{B}$ is invertible.

(b) Show that the conditionally specified Gaussian model

$$Z(\mathbf{s}_i) \,|\, \{z(\mathbf{s}_j)\colon j \neq i\} \sim N\left(\mu_i + \sum_{j=1}^{n} c_{ij}\big(z(\mathbf{s}_j) - \mu_j\big), \tau_i^2\right), \quad i = 1, \ldots, n,$$

implies $\mathbf{Z} \sim N_n\big(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}\big)$, provided $\mathbf{I} - \mathbf{C}$ is invertible and $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$ is symmetric and positive definite.

3. (10 points) Let $\mathbf{s}_1, \ldots, \mathbf{s}_n \in A \subset \mathbb{R}^d$ be a realization of an inhomogeneous Poisson process with intensity $\lambda_{\boldsymbol{\theta}}(\cdot)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Show that the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \lambda_{\boldsymbol{\theta}}(\mathbf{s}_i) \exp\left\{-\int_A \lambda_{\boldsymbol{\theta}}(\mathbf{s})\nu(d\mathbf{s})\right\},$$

where $\nu$ is the Lebesgue measure.

4. (15 points) Consider the *change-of-support* problem in a regression setting, where the response process $Y(\cdot)$ and the predictor process $X(\cdot)$ have different supports. Here the *support* refers to the region over

which the process is averaged. Suppose data on $Y(\cdot)$ are observed at monitoring sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and data on $X(\cdot)$ are output from a physical model as block averages

$$X(B_i) = \frac{1}{|B_i|} \int_{B_i} X(\mathbf{s})\, d\mathbf{s}, \quad i = 1, \ldots, m.$$

How would you model such data and estimate the parameters? You may summarize a state-of-the-art method from the literature (and cite the reference) or propose your own solutions.

5. (15 points) Consider a type of change-of-support problem in spatial point process modeling. Let $N$ be a Cox process with intensity $\Lambda(\cdot)$ that depends on a Gaussian predictor process $X(\cdot)$. Suppose we observe $N(B_i)$, the number of events in block $B_i$, $i = 1, \ldots, m$, and data on $X(\cdot)$ at monitoring sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Describe an existing approach (and cite the reference) or propose your own method for modeling and parameter estimation.