Strathmore
UNIVERSITY

INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN STATISTICAL SCIENCES
END OF SEMESTER EXAMINATION
STA 8408: SPATIAL STATISTICS

DATE: Monday, August 13, 2016 Time: 2 Hours

**Instructions**

1. This examination consists of **FOUR** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

**Question 1 (20 Marks)**

a) Enumerate the key components of the Cressie and Cassie (1993) **classification of spatial statistics**. Describe each component of this classification, explaining (where necessary) any difference between various components of the classification.

(**6 Marks**)

b) Explain the significance of the following concepts in spatial statistics:
   i) "***Spatial autocorellation***"
   ii) "***Tobler's first law of geography***"

(**6 Marks**)

a) Describe a stationary Gaussian process and hence explain the meaning of the terms isotropy and anisotropy.

(**7** marks)

**Question 2**

a) The variogram is an important tool in geostatics.
   i) How is the variogram used in the assessment of spatial autocorrelation?
   ii) The nugget, sill, and range are important components of a variogram. Describe each component.
   iii) Explain, mathematically, the relationship between the variogram and the covariance function in a stationary Gaussian process.

(**8** marks)

b) Kriging is an import part of geostatistics. Explain what Kriging is and hence distinguish between simple and ordinary kriging.

(**4** marks)

c) Consider a spatial process $\{Y(\boldsymbol{s}): s \in D\}$ with values $Y(\boldsymbol{s_1}), \dots, Y(\boldsymbol{s_n})$ at sites $\boldsymbol{s_1}, \dots, \boldsymbol{s_n}$. If we also assume that $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}), \varepsilon(\boldsymbol{s}) \sim MVN(\boldsymbol{0}, \Sigma(\boldsymbol{\theta}))$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \phi, \tau^2)$,derive an expression for the simple kriging estimator of $Y(\boldsymbol{s_0})$, the outcome variable at site $\boldsymbol{s_0}$.

(**8** marks)

**Question 3 (20 Marks)**

a) Distinguish between the terms "**clusters**" and "**clustering**." By means of appropriate examples, suggest spatial statistical methods that can be used to detect each of the two phenomena in Areal data sets.

(**4 Marks**)

b) **Spatial error**, **Spatial Lag** and **Spatial Durbin** models are common approaches used to model Areal data. Distinguish between these models, presenting an expression for each model in matrix form and providing a justification for its use.

(**6 Marks**)

c) For the spatial error model,

i) Derive an expression for the generalized least squares estimator of the vector parameters in the model;

(**6 Marks**)

ii) Determine the mean and variance of the vector of parameter estimates obtained in part (i)**.**

(**4 Marks**)

**Question 4 (20 Marks)**

a) Let $Y_i$ denote the observed counts of a disease in region $i = 1, \dots, n$. Consider the poisson-gamma model, where $Y_i|\theta_i \sim Poisson(\theta_i E_i)$ and where $\theta_i \sim Gamma(\alpha, \beta)$. Prove that the posterior mean of the for this model is a weighted average of the data-based SMR for region i and the prior mean $\mu$.

(**7 Marks**)

b) The ***conditional autoregressive model*** and the ***intrinsic autoregressive model*** are widely used as prior distribution for random spatial effects in Bayesian models.
Provide a mathematical description of each of these models, explaining how the two models relate to each other.

(**7 Marks**)

c) The results presented in the Appendix arise from fitting Bayesian Hierarchical spatial models to routinely collected incident Tuberculosis cases in Kenya. This data aggregated at District level is used to assess the relationship between TB incidence and the following predictors: Proportion poor in the district, ***proppoor***; District level HIV prevalence, ***hivprev***; Mean household size, ***meanhhsize***; Altitude, **altitude**; Proportion of urban households, ***urban;*** and illiteracy
The results of fitting a Bayesian Poisson regression model (***Model 1***), a Bayesian Poisson regression model with a spatially unstructured random effect (***Model 2***), a Bayesian Poisson regression model with a spatially structured random effect (***Model 3***), and a convolution model (***Model 4***) are presented in Appendix.

i) Determine the best fitting model.

(**2 Marks**)

ii) For the convolution model, which of the two random effects dominates the explained variability in the data.

(**2 Marks**)

iii) Discuss the direction of effect for each parameter in the best fitting model.

(**2 Marks**)

# APPENDIX

**Model 1. The standard poisson regression model**

$$log(\mu_i) = log(E_i) + \alpha_1 + \alpha_2 Proppoor + \alpha_3 HIVprev + \alpha_4 meanHHsize + \alpha_5 Altitude + \alpha_6 Urban + \alpha_7 Illiteracy$$
$$\alpha_1 \sim dflat(\ ), and\ \alpha_j \sim N(0.0, 10^4), j = 1, \ldots, 7$$

**DIC**

**Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes**

| | Dbar | Dhat | pD | DIC |
|---|---|---|---|---|
| O2008 | 11259.600 | 11252.700 | 6.916 | 11266.500 |
| total | 11259.600 | 11252.700 | 6.916 | 11266.500 |

Node statistics

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| alpha[1] | -0.263 | 0.005125 | 1.222E-4 | -0.273 | -0.2631 | -0.2529 | 1 | 10000 |
| alpha[2] | -0.1471 | 0.004109 | 7.097E-5 | -0.1553 | -0.147 | -0.1392 | 1 | 10000 |
| alpha[3] | 2.867 | 0.04697 | 0.001137 | 2.773 | 2.868 | 2.959 | 1 | 10000 |
| alpha[4] | 0.2757 | 0.03148 | 0.002763 | 0.2156 | 0.2753 | 0.3363 | 1 | 10000 |
| alpha[5] | -0.2957 | 0.03171 | 0.00278 | -0.3566 | -0.2959 | -0.2356 | 1 | 10000 |
| alpha[6] | 0.6241 | 0.007888 | 1.242E-4 | 0.6086 | 0.6242 | 0.6393 | 1 | 10000 |
| alpha[7] | 0.1512 | 0.005391 | 1.19E-4 | 0.1407 | 0.1511 | 0.1619 | 1 | 10000 |

---

**Model 2. The Poisson regression model with a spatially unstructured random effect**

$$log(\mu_i) = log(E_i) + \alpha_1 + \alpha_2 Proppoor + \alpha_3 HIVprev + \alpha_4 meanHHsize + \alpha_5 Altitude + \alpha_6 Urban + \alpha_7 Illiteracy + v_i$$
$$\alpha_1 \sim dflat(\ ), \quad \alpha_j \sim N(0.0, 10^4), j = 1, \ldots, 7, v_i \sim N(0.0, {\tau_v}^{-1}), and\ \tau_v \sim Gamma(0.5, 0.005)$$

**DIC**

**Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes**

| | Dbar | Dhat | pD | DIC |
|---|---|---|---|---|
| O2008 | 677.118 | 606.948 | 70.169 | 747.287 |
| total | 677.118 | 606.948 | 70.169 | 747.287 |

Node statistics

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| alpha[1] | -0.3667 | 0.06846 | 0.006637 | -0.4812 | -0.3758 | -0.208 | 1 | 10000 |
| alpha[2] | -0.08904 | 0.06304 | 0.006139 | -0.2123 | -0.08558 | 0.02145 | 1 | 10000 |
| alpha[3] | 3.814 | 0.5401 | 0.05196 | 2.66 | 3.821 | 4.741 | 1 | 10000 |
| alpha[4] | 0.1562 | 0.06689 | 0.006513 | 0.0238 | 0.1635 | 0.2727 | 1 | 10000 |
| alpha[5] | -0.1974 | 0.08377 | 0.008251 | -0.4057 | -0.1974 | -0.04384 | 1 | 10000 |
| alpha[6] | 0.5937 | 0.2279 | 0.02264 | 0.1318 | 0.6051 | 1.061 | 1 | 10000 |
| alpha[7] | 0.1603 | 0.06912 | 0.006695 | 0.004021 | 0.1688 | 0.2772 | 1 | 10000 |
| sigma.v | 0.3976 | 0.03609 | 8.945E-4 | 0.3352 | 0.3954 | 0.4732 | 1 | 10000 |
| tau.v | 13.87 | 671.9 | 7.399 | 4.466 | 6.398 | 8.9 | 1 | 10000 |

---

**Model 3. The Poisson regression model with a spatially structured random effect**

$$log(\mu_i) = log(E_i) + \alpha_1 + \alpha_2 Proppoor + \alpha_3 HIVprev + \alpha_4 meanHHsize + \alpha_5 Altitude + \alpha_6 Urban + \alpha_7 Illiteracy + v_i$$
$$\alpha_1 \sim dflat(\ ), \quad \alpha_j \sim N(0.0, 10^4), j = 1, \ldots, 7, u_i \sim N(0.0, {\tau_v}^{-1}), and\ \tau_v \sim Gamma(0.5, 0.005)$$

**DIC**

**Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes**

| | Dbar | Dhat | pD | DIC |
|---|---|---|---|---|
| O2008 | 680.861 | 607.104 | 73.757 | 754.618 |
| total | 680.861 | 607.104 | 73.757 | 754.618 |

Node statistics

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| alpha[1] | -0.3277 | 0.04145 | 0.004029 | -0.4054 | -0.3233 | -0.243 | 1 | 10000 |
| alpha[2] | -0.02648 | 0.07251 | 0.007125 | -0.1336 | -0.04623 | 0.1283 | 1 | 10000 |
| alpha[3] | 3.175 | 0.6713 | 0.06572 | 1.7 | 3.143 | 4.468 | 1 | 10000 |
| alpha[4] | -0.1392 | 0.1213 | 0.01207 | -0.3538 | -0.1593 | 0.1189 | 1 | 10000 |
| alpha[5] | 0.1485 | 0.1298 | 0.01293 | -0.1224 | 0.1784 | 0.3281 | 1 | 10000 |
| alpha[6] | 0.5423 | 0.1923 | 0.01905 | 0.1409 | 0.5368 | 0.8602 | 1 | 10000 |
| alpha[7] | 0.08137 | 0.07836 | 0.00767 | -0.08195 | 0.08627 | 0.2298 | 1 | 10000 |
| sigma.u | 0.9775 | 0.09005 | 0.001965 | 0.8222 | 0.972 | 1.166 | 1 | 10000 |
| tau.u | 9.841 | 683.4 | 8.771 | 0.736 | 1.058 | 1.48 | 1 | 10000 |

---

**Model 4. The Poisson regression with both spatially structure and unstructured random effects**

$$log(\mu_i) = log(E_i) + \alpha_1 + \alpha_2 Proppoor + \alpha_3 HIVprev + \alpha_4 meanHHsize + \alpha_5 Altitude + \alpha_6 Urban + \alpha_7 Illiteracy + u_i$$

$$\alpha_1 \sim dflat(\ ), \qquad \alpha_j \sim N(0.0, 10^4), j = 1, \ldots, 7, u_i | \boldsymbol{u}_{-i} \sim N\left(\mu_i + \rho \sum_j W_{ij}(u_j - \mu_j), {\sigma_u}^2/d\right), and\ \tau_v \sim Gamma(0.5, 0.005)$$

DIC

**Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes**

| | Dbar | Dhat | pD | DIC |
|---|---|---|---|---|
| O2008 | 677.353 | 606.951 | 70.402 | 747.754 |
| total | 677.353 | 606.951 | 70.402 | 747.754 |

**Node statistics**

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| alpha[1] | -0.3423 | 0.05775 | 0.005523 | -0.4577 | -0.3396 | -0.2303 | 1 | 10000 |
| alpha[2] | -0.08964 | 0.05809 | 0.005671 | -0.2049 | -0.09264 | 0.03254 | 1 | 10000 |
| alpha[3] | 3.538 | 0.6392 | 0.06246 | 2.058 | 3.606 | 4.658 | 1 | 10000 |
| alpha[4] | 0.2429 | 0.09002 | 0.008863 | 0.0674 | 0.2533 | 0.3857 | 1 | 10000 |
| alpha[5] | -0.2863 | 0.09791 | 0.009653 | -0.5029 | -0.2771 | -0.1215 | 1 | 10000 |
| alpha[6] | 0.4646 | 0.2442 | 0.02435 | -0.02466 | 0.4927 | 0.8583 | 1 | 10000 |
| alpha[7] | 0.1553 | 0.05709 | 0.005477 | 0.04527 | 0.1513 | 0.2781 | 1 | 10000 |
| sigma.u | 0.05953 | 0.04818 | 0.004401 | 0.01417 | 0.04301 | 0.2032 | 1 | 10000 |
| sigma.v | 0.3944 | 0.03753 | 0.001273 | 0.3298 | 0.3918 | 0.4722 | 1 | 10000 |
| tau.u | 1040.0 | 1518.0 | 99.34 | 24.21 | 540.6 | 4985.0 | 1 | 10000 |
| tau.v | 15.64 | 779.7 | 9.052 | 4.485 | 6.515 | 9.194 | 1 | 10000 |