Homework 5 - Stat 534
Due Tuesday February 17, 2017

1. Let $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \gamma(\mathbf{h}_{ij})$ be a semivariogram for a second-order stationary spatial process.

   (a) Show that $\gamma(\mathbf{h}_{ij}) = C(\mathbf{0}) - C(\mathbf{h}_{ij})$.

   (b) Show that
   $$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(\mathbf{h}_{ij}) \leq 0$$
   for any sites $\mathbf{s}_i, i = 1, \cdots, n$ and for constants $a_i, i = 1, \cdots, n$ with $\sum_{i=1}^{n} a_i = 0$.

2. Matheron's semivariogram estimator is
   $$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2$$
   where $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j\}$ and $|N(\mathbf{h})|$ is the number of pairs in the set $N(\mathbf{h})$.

   (a) Let $Z(\mathbf{s}) = \mu + e(\mathbf{s})$ where $E[e(\mathbf{s})] = 0$ with $\gamma_Z(\mathbf{h}) = \gamma_e(\mathbf{h})$ (adding a constant to the random error terms does not change the variance/covariance properties of the process). Show that $\widehat{\gamma}(\mathbf{h})$ is unbiased for $\gamma_Z(\mathbf{h})$. That is, show $E[\widehat{\gamma}(\mathbf{h})] = \gamma_Z(\mathbf{h})$. Hint: Show that under an assumption of a constant mean
   $$\gamma_Z(\mathbf{h}) = \frac{1}{2}E[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2.$$

   (b) We pointed out in class that Matheron's Estimator is biased in the presence of trend. Let $Z(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s})$ with $E[e(\mathbf{s})] = 0$ and $\gamma_Z(\mathbf{h}) = \gamma_e(\mathbf{h})$. Show
   $$E[\widehat{\gamma}(\mathbf{h})] = \gamma_e(\mathbf{h}) + \frac{1}{2|N(\mathbf{h})|}\sum_{N(\mathbf{h})}[\mu(\mathbf{s}_i) - \mu(\mathbf{s}_j)]^2.$$

   (c) Consider the (very) simple model $Z_i = 10 + e_i$ where the e_i's are independent normally distributed error terms with variance $\sigma^2 = 81$. We have a pure nugget effect model $\gamma_Z(h) = \gamma_e(h) = 81$. Simulate 100 observations of Z_i and calculate the empirical semivariogram assuming the observations are on a one-dimensional transect.

   ```
   library(geoR)
   Zdat<-10+rnorm(100,0,9)
   i<-1:100
   xycoord<-cbind(c(rep(1,100),i))
   Zvgram<-variog(coords=xycoord,data=Zdat)
   plot(Zvgram)
   ```
   Compare what you see in the plot to the true $\gamma_Z(\mathbf{h})$. Is the result consistent with part (a) above? Why or why not?

(d) Redo the above calculations based on the (still) simple model $Z_i = 10 + 10i + e_i$, i.e. there is now a linear trend and the process is no longer stationary.

```
Zdat2<-10 +10*i +rnorm(100,0,9)
Zvgram2<-variog(coords=xycoord,data=Zdat2)
plot(Zvgram2)
```

Compare the empirical semivariogram `Zvgram2` to `Zvgram`. Are the results consistent with part (b) above? Justify your answer.

(e) Fit a linear model to the data in (d), extract the residuals, and compute the empirical semivariogram for the residuals. Note that what you are doing is removing the trend.

```
e.resid=residuals(lm(Zdat2~i))
evgram=variog(coords=xycoord,data=e.resid)
plot(evgram)
```

Compare the 3 empirical semivariograms.

3. Attached is a data set containing the carbon nitrogen values used in the carbon/nitrogen data set. The first 2 columns contain the coordinates, total nitrogen is in the third column, total carbon is in the 4th column and the ratio is in the last column. We will work with the total carbon data. Use `geoR` for the analysis. It will be easiest if you convert the data into a `geodata` object as follows.

```
library(geoR)
# I called the data set CN.dat
TC.geodata<-as.geodata(CN.dat,coords.col=1:2,data.col=4)
```

(a) Calculate the empirical semivariogram. Give initial *eyeball* estimates of the nugget effect, sill, and (effective) range.

(b) Fit an exponential semivariogram to the carbon data using OLS, WLS, MLE, and $REML$ methods. Specify a nugget effect in each case, i.e. you do not need to consider models without a nugget. Plot the fitted functions and comment on which one you like best.

(c) Redo part (b) by fitting a spherical model.

(d) Summarize your results in table format. Compare the results and discuss.

Watch those range parameters you enter initial estimates for in the two models. You are entering 1/3 the range for the exponential model and the range itself for the spherical.