Exam 1 - Stat 534
Due Friday March 10, 2017

You may use the class notes from this semester, the text, homeworks and associated keys **from this semester only**. *Do not use any other material. Do discuss the exam with anyone but me. Pay close attention to your email the next couple of days as I will be providing clarification/corrections in that way. It is due by 4 on Friday but I have a meeting starting at 3 so you will want to talk to me before then.*

1. Suppose we have an intrinsically stationary process with semivariogram

$$\left(\frac{1}{2}\right) Var\left(Z\left(\mathbf{s}_i\right) - Z\left(\mathbf{s}_j\right)\right) = \gamma\left(\mathbf{s}_i - \mathbf{s}_j\right) = \gamma\left(\mathbf{h}_{ij}\right).$$

On an earlier homework you showed 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 any constants $a_i, i = 1, \cdots, n$ with $\sum_{i=1}^{n} a_i = 0$ but you did it under an assumption of second order stationarity. We will now establish it in general.

   (a) First show that

$$-\left(\frac{1}{2}\right)\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \left(Z\left(\mathbf{s}_i\right) - Z\left(\mathbf{s_j}\right)\right)^2\right\} = \left\{\sum_{i=1}^{n} a_i Z\left(\mathbf{s}_i\right)\right\}^2$$

   (b) Now take expectations of both sides to establish the result.

2. Let $X_0 \sim Gamma(\alpha, \beta)$ with the parameterization

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-x/\beta\right);\ x > 0$$

and 0 elsewhere. Let $X_i \sim Gamma(\alpha_i, \beta)$ for $i = 1, \cdots, n$. We construct a one-dimensional regularly spaced random field at locations $i = 1, \cdots, n$

$$Z(s_i) = X_0 + X_i;\ i = 1, \cdots, n$$

You can assume that $X_0, X_1, \cdots, X_n$ are independent.

   (a) What is the distribution of $Z(s_i)$.

   (b) Find $E(Z(s_i))$ and $Var(Z(S_i))$. You can use known properties of the Gamma distribution to answer this question, i.e. you can just write down the answer if you know it or can find it.

   (c) Find $Cov(Z(S_i), Z(s_j))$ for $i \neq j$.

   (d) Is this a second-order stationary process? Justify your answer.

3. The `lansing` data set in the `spatstat` package contains spatial locations of several different species of trees. We will be looking and comparing the distributions of black oaks and maples.

```
require(spatstat)
data(lansing)
blackoak<-split(lansing)$blackoak
maple<-split(lansing)$maple
```

The rectangular region is a unit square. Use the isotropic edge corrected version when applicable below. Answer the following questions. You will be computing several simulation envelopes below. Be patient and keep nsim=99, the default.

(a) What does the K function measure?

(b) It is often easier to interpret the L function than the K function. Based on the L function do the black oaks appear to be clustered or do they appear to be regularly distributed? Do the maples appear to be clustered or do they appear to be regularly distributed? Justify your answer. Simulation envelopes will help you give a better answer to this question.

(c) Compare the two L functions and discuss whether or not the 2 processes appear to be the same. You can use the results from (a) but you should also look at the difference more formally using the following also provided in an attached script file.

```
require(splancs)
# specify radii
h<-seq(0,.5,l=100)
# get coordinates
tree.poly<-list(x=c(blackoak$x,maple$x),y=c(blackoak$y,maple$y))
# recompute the K functions
kblackoak<-khat(as.points(blackoak),bboxx(bbox(as.points(tree.poly))),h)
kmaple<-khat(as.points(maple),bboxx(bbox(as.points(tree.poly))),h)
# get the differences
k.diff<-kblackoak - kmaple
# generate the envelope
env<-Kenv.label(as.points(blackoak),as.points(maple),
bboxx(bbox(as.points(tree.poly))),nsim=99,s=h)
# plot the results
plot(h,seq(-0.15,0.05,l=length(h)),type="n",ylab="Kdiff",
main="Envelopes for Kdiff")
lines(h,k.diff)
lines(h,env$low,lty=2)
lines(h,env$up,lty=2)
abline(h=0)
```

(d) Plot $L_{ij} - h$ versus h for black oaks and maples.

```
Kplot<-envelope(lansing,Kcross,i="blackoak",j="maple")
plot(Kplot,sqrt(./pi)-r~r,ylab="Lij - h",main="Cross L Function",legend=F)
```

What type of relationship between the point patterns of the two species of trees is indicated by this plot? Justify your answer.

(e) Based on the above, comment on the null hypotheses of independence and random labeling.

4. You were sent the `wheat` data set on a previous homework assignment. You want to predict the value of Z (yield) at an arbitrary location. Assume a pure nugget effect model.

   (a) What are the kriging weights and what is the predicted value?

   (b) What is the estimate of the sill?

   (c) What is the kriging standard error (note that this is a *prediction* error)?

5. Carbon-Nitrogen data example: We looked at estimating the semivariogram of the residuals from a simple linear regression model of total carbon on total nitrogen in class. We used `gls` to do this (as part of incorporating a spatial covariance structure into the regression) specifying an exponential covariance model and estimating the parameters using both maximum likelihood and REML. Lets check to see what the `lik.fit` function in the `geoR` package would return as parameter estimates (nugget, practical range, and partial sill) and see if the results are comparable. Some of the relevant `R` code is included in the attached script file. Compare the estimates on page 11 of the Spatial Regression notes and the estimates you get out of `lik.fit`. Use the same starting values.