Week 7.

Stat 222, Spatial Statistics. Lecture MWF 9am, Math-Sci 5203.
Professor: Rick Paik Schoenberg, frederic@ucla.edu, www.stat.ucla.edu/∼frederic

DAY FIFTEEN. Monday, 5/14/01.

1) Midterm exam reviewed.

2) Some Gaussian processes.

a) Brownian Motion (BM):
In one dimension (i.e. a process $Z(t)$ on the line), BM is defined as follows:
• increments $Z(t_j) - Z(t_i)$ are Gaussian, with mean 0 and variance $|t_j - t_i|$.
• disjoint increments are independent. That is, if $t_1 < t_2 < t_3 < t_4$, then $Z(t_2) - Z(t_1)$ is independent of $Z(t_4) - Z(t_3)$.
• $Z(t)$ is continuous as a function of t.
b) Brownian Sheet. In 2 or more dimensions, the same process described above is sometimes called Brownian *sheet*.
c) Brownian Bridge (BB):
In one-dimension, BB is Brownian motion on the segment $[0, 1]$ but *tied down* so that $Z(1)$ must equal 0. Alternatively, may be defined as follows. Let Z be BM, and let $Y(s) = Z(s) - sZ(1)$. Then Y is BB.
In two-dimensions, the definition is the same except the process is constrained so that $Z(1, 1) = 0$.
3) Kriging.

From Krige (1951), a South African mining engineer who studied ore-grade distributions. Matheron (1963) coined the term "kriging" to mean predicting the value (or more generally, some function of a set of values) of a spatial process by minimizing mean-squared-error.

Can predict a data point $Z(s)$, or some numerical summary $f\{Z(1, 1), Z(1, 2), \ldots\}$ of the data. Note that, like regression, kriging is typically described as a tool for *prediction* of values that are at present unobserved but will be collected in the future. However, it may be applied to data collected in the past that will never be observed, and kriging in this context is usually called *smoothing* or *filtering* rather than *predicting*. But it's really all the same thing.

DAY SIXTEEN. Wednesday, 5/16/01.

1) Projects.

I want you to choose some *spatial point process* data for your projects. Basically any mapped collection of points will do. Need not be on an academic subject.

a) MUST have at least 30 points. Ideally 100 or more.

b) RELATIVE locations should be of interest. Question: are the points uniformly spread, or does there seem to be some clustering or inhibitory behavior? These should be interesting questions.

c) Try to avoid datasets that are only interesting if you have marks or covariates.

d) It is completely acceptable for YOU to approximate the locations yourself. That is, if you have a *map* of the locations but not the numerical coordinates, estimate the coordinates yourself, and in the end show me the original map along with the map of your coordinates, and see if they match up well. Also, feel free to rescale and truncate the map however you wish.

e) It is completely appropriate for you to use data that you already have, or that is related to other research you are doing.

2) Examples of Point Process (PP) data.

a) Occurrances of earthquakes in Southern California since 1990.

b) Houses currently for sale in Los Angeles (or those that cost over 1 million dollars).

c) Walk around the UCLA campus, and plot the locations of tall (over 6'2") people.

d) Birthplaces of US presidents (or candidates).

e) Places where Allen Iverson shoots from.

f) Reported sightings of UFOs in the United States.

g) Locations of Starbucks stores in Los Angeles County.

Note that X and Y don't have to be locations: can have a map of points indicating unemployment and interest rate when the Federal Reserve has lowered interest rates, for example, but generally this is discouraged because the methods we'll be discussing are primarily of interest for truly *spatial* data. Basically any **map** with points on it will do!

3) Kriging.

a) Ordinary (and simple) kriging.

Say we want to predict an observation $Z(s_0)$, given $Z(s_1), Z(s_2), \ldots$.

In ordinary kriging, we assume $Z(s) = \mu + \delta(s)$, where μ is a constant and $\delta(s)$ is an intrinsically stationary, mean 0 process, with some correlation, and some known (or estimable) variogram $2\gamma(h)$. In addition, we confine ourselves to an estimate of the form

$\hat{Z}(s) = \sum_{i=1}^{n} \lambda_i Z(s_i)$, where $\sum \lambda_i = 1$. That is, we are requiring ourselves to estimate $Z(s)$ using some linear combination of the surrounding observations, and the λ_i are to be

chosen so that we minimize mean squared error. If μ is *known*, this is called *simple* kriging; otherwise it's *ordinary* kriging.

Given the matrix Γ consisting of variogram terms, see Cressie pp121-122 for formulas for ordinary kriging estimates in terms of Γ. Using the resulting kriging predictors λ_i, one can predict or smooth to obtain $\hat{Z}(s)$. See p126 for an example.

b) Universal kriging.

Now suppose that the mean μ is not constant, but changes with s. Set some collection of functions $f_1(s), f_2(s), \ldots, f_p(s)$, which are fixed, for all s. Suppose

$\mu(s) = \sum_{i=1}^{p} \beta_j f_j(s)$, where the β_j are to be estimated.

So $Z(s) = \sum \beta_j f_j(s) + \delta(s)$, where δ is as in ordinary kriging. Again, predict using

$\hat{Z}(s) = \sum_{i=1}^{n} \lambda_i Z(s_i)$, where the λ_i are chosen to minimize mean squared error.

DAY SEVENTEEN. Friday, 5/18/01.

1) More on kriging.

Let me emphasize that after doing kriging, you end up with some estimated optimal values of λ_i. These are optimal for prediction, given your variogram $2\gamma(h)$. Page 122 of Cressie gives the formula for deriving the λs from $2\gamma(h)$:

$\lambda_0 = \Gamma_0^{-1}\gamma_0$,

where λ_0 = the vector $(\lambda_1, \ldots \lambda_n, m)'$,

$\gamma_0 = (\gamma(s_0 - s_1), \ldots, \gamma(s_0 - s_n), 1)'$,

Γ_0 = an $(n+1) \times (n+1)$-matrix where $\Gamma_0(i,j) = \gamma(s_i - s_j)$, for $i, j = 1, 2, \ldots, n$, and $\Gamma_0(i, n+1) = 1$, $\Gamma_0(n+1, j) = 1$, and $\Gamma_0(n+1, n+1) = 0$.

Thus Γ_0 has zeros along the diagonal. Given γ, then, one obtains λ_0 and uses that for prediction. The basic idea is that you have to use the variogram since observations are correlated, so if you're trying to predict a value $Z(s_i)$, using a weighted average of nearby values, then you should weight values $Z(s_j)$ more if they are likely to be very similar to $Z(s)$, i.e. if the variogram $2\gamma(s_i - s_j)$ is smaller.

2) Point processes.

Point processes are highly discontinuous, and the basic goal in their analysis is simply to smooth them. Can think of the points as peas, and we want to smash them, resulting in a smoother picture that emphasizes certain basic features about how the peas are laid out, and which is more easily interpreted than simply the image of the points themselves.

a) Nonparametric estimation.

(i) Kernel estimation.

$\lambda_h(s) = \frac{1}{p_h(s)} \sum_{i=1}^{n} h^{-d} K\left(\frac{s - s_i}{h}\right)$, for some kernel K.

d is just the dimension of the process (so $d = 2$ for planar processes).

$p_h(s)$ is an edge correction factor $= \int h^{-d} K(\frac{s-u}{h}) du$.

Kernel estimation is like smashing the peas, and this gives an estimate of the overall intensity λ_1 of the point process. However, it doesn't tell you about clustering or other second-order (interactive) behavior of the point process. So you might also look at...

(ii) K-function.

The theoretical K-function is given by

$K(h) = \frac{1}{h} E[\#$ of points within h of a given point$]$.

How is it estimated? We'll discuss this next time....