Homework 4
Stat 597a: Spatial Statistics
Due Tuesday, November 14

Hand in a paper copy of your solutions. You don't need to typeset them, but make sure the problems are presented in order. Any code you write that is not specifically requested in the questions should be included as an Appendix at the end.

1. Alt and Vach (1991) describe an archaeological investigation of an early medieval burial ground in Germany. One question of interest was whether grave sites tended to be placed according to family units. The archaeologists considered an inherited feature in the teeth of the excavated skeletons; each grave has a location and an indicator variable for whether the individual had this feature or not. The data for the point process is in the file `dental.reduced.dat`. The second column is the indicator variable and the last two columns are x and y location. The first column is an index you may ignore.

   (a) Load the data and create two `ppp` objects from it, one for affected and one for unaffected individuals. A key question is: what is the window? It is not available for this data. For now, take the window to be the same for each `ppp` object: use a rectangular region based on the range of x and y for both datasets. (Have a look at the help file for `owin`.)

   (b) For each dataset separately, create Monte Carlo simulation envelopes for the F and G functions and plot them. Clearly label your plots and turn in a few sentences describing any choices you made in creating the envelopes. Is their evidence against CSR in this dataset? If so, what type of violation is suggested?

   (c) Now let's consider a more likely window: that these graves represent a complete excavation of the area in which they appear, and that area is irregularly shaped. Since we don't know what it is, we can choose a rough polygon outline to surround the points. You can create such an outline by plotting the locations and using the `locator` function; again, see `help(owin)` for the details of how to specify a polygon boundary. Create two new `ppp` objects with this new window.

   (d) Repeat step (b) for the new datasets. What changes? Can you explain the reason for this, based on the form of the test statistics?

2. Simulate four datasets on the unit square, from a homogeneous Poisson process with a rate λ of your choosing. For each one, fit a kernel estimate of the intensity function and plot it with the points overlaid. What do you conclude about interpreting the results of this estimator? In other words, what can go wrong if we fit a spatial point process model without checking for CSR first?

3. Read through sections 15.2 and 15.3 of the notes by Adrian Baddeley about fitting point process models in `R` with the `spatstat` package (you can find it on Canvas).

   In 15.3, follow along with the `R` code example for the `bei` dataset. In particular, find the MLEs for inhomogeneous models with intensity functions:

   Model 1: $\lambda(x) = \exp\{\beta_0 + \beta_1 Z(x)\}$

   Model 2: $\lambda(x) = \beta Z(x)$

   Plot a kernel density estimate of $\lambda(x)$, ignoring covariates, as well as the fitted intensities under the two models. Put them all on the same color scale, and include a sentence or two comparing them. *Note: the data for this problem are available from the* `spatstat` *package.*