Homework 2 - Stat 534
Due Friday January 26, 2017

This assignment will require you to use `spatstat`*. We will look at some of the quadrat tests we went over in class. Install* `spatstat` *if you have not already done so.*

1. It was pointed out in class that, conditional on n events, event locations are uniformly distributed for a homogeneous Poisson process. We will consider the simplest example of this. Consider a one-dimensional process on a transect of length L, $(0, L]$. Given that one event has occurred on the interval $(0, L]$ what is the probability that it occurred in the subinterval $(0, s]$ for $s < L$?

2. We looked at an example of `quadrat.test` on the amacrine data set in class. We will use it to analyze another data set, called `redwood`. You can read about this in the `spatstat` help material. You will be using the `quadrat.test` function. You can also read about this function in the help material.

   (a) Read the help pages on the `quadrat.test` function. What null hypothesis do they claim to be testing?

   ```
   require(spatstat)
   data(redwood)
   help(redwood) # optional information on the data set
   help(quadrat.test)
   ```

   (b) Use `quadrat.test` on the redwood data set.

   ```
   plot(redwood) # you do not need to turn this in but it may help you below
   redwood.fit<-quadrat.test(redwood)
   redwood.fit
   ```

   The default partitioning of the grid is 5×5. Does that appear appropriate here? Justify your answer.

   (c) Redo the analysis using a 3×3 grid.

   ```
   redwood.fit<-quadrat.test(redwood,nx=3,ny=3)
   redwood.fit
   ```

   (d) Is the value of the test statistic X^2 indicative of clustering, CSR, or a regular pattern? Justify your answer. Note that I am only asking you to compare the observed value of X^2 to what you would expect under each of these three patterns. You do not need to calculate a P-value just yet.

   (e) The investigator suspected a clustered pattern and the plot would seem to be consistent with this. Rerun the test with `alternative="c"` in the argument list. Give the p-value for the test and interpret the results. Does the test provide evidence against CSR and for clustering? Justify your answer.

   (f) We can plot the results of the fit.

   ```
   plot(redwood.fit)
   ```

You will see a plot of the 3×3 grid. There are 3 numbers in each cell: the observed count (upper left), expected count under CSR (upper right), and a scaled residual (lower number). The sum of the scaled residuals is the X^2 statistic. Give the results of the test and using the plot indicate where CSR seems to break down, if it does.

(g) Quadrat size can be important. Repeat the analysis using a 2×2 grid. Give the results and compare to what we saw with the 3×3 grid.

```
redwood.fit<-quadrat.test(redwood,nx=2,ny=2)
redwood.fit
```

3. We will compare results from Monte Carlo procedures based on Poisson sampling and based on conditioning on the number of observed points. We will use the `cells` data set. The R code to accomplish that is shown below. Compare the two procedures. What do they indicate about the spatial pattern and why? Which procedure do you like best for this data set and why?

```
data(cells)
hbar<-mean(nndist(cells))
hbar
hbar.pois<-rep(0,1000)
hbar.cond<-rep(0,1000)
hbar.pois[1]<-hbar
hbar.cond[1]<-hbar
for(i in 2:1000){
# Poisson Monte Carlo
dat.pois<-rpoispp(42)
hbar.pois[i]<-mean(nndist(dat.pois))
# Conditional Monte Carlo
dat.cond<-runifpoint(42)
hbar.cond[i]<-mean(nndist(dat.cond))}
par(mfrow=c(2,1))
hist(hbar.pois,prob=T,main="Poisson Monte Carlo")
abline(v=hbar)
hist(hbar.cond,prob=T,main="Conditional Monte Carlo")
abline(v=hbar)
# Poisson P-value
2*sum(hbar.pois>=hbar)/1000
# Conditional P-value
2*sum(hbar.cond>=hbar)/1000
```

4. Below is the frequency distribution of the number of trees per quadrat in a sample of 100 quadrats each of radius 6 m.

Trees per quadrat	0	1	2	3	4	≥ 5
Count	34	33	17	7	3	6

The data were pooled for counts ≥ 5 to meet the assumptions of the method. Carry out a Poisson goodness-of-fit test based on an assumption of CSR. Discuss the results. The sample mean of the observed counts was 1.43.

5. Suppose we have a realization of a spatial point process consisting of N event locations $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N\}$. Let H_i denote the distance between the ith event and the nearest neighboring event. The cumulative distribution function of H (the nearest event-event distance) is the G function. (This problem will be continued on the next homework assignment).

   (a) What is the G function if the point process is CSR; i.e. what is $G(h) = P(H \leq h)$?

   (b) Find the pdf of H.

   (c) Find $E[H]$ and $Var(H)$. Hint: you found the pdf but before you start evaluating a gnarly integral take a close look at that pdf and see if you cannot identify the family of distributions it belongs to. If you can do that then you can use that knowledge to find the mean and variance.