Spatial Statistics; STAT 605
First half of Exam 2 (take home exam)
Due Friday, April 1, 2022, midnight.

For this exam, I make the following truthful statements:

- I have not received, I have not given, nor will I give or receive, any assistance to another student taking this exam;
- The only resources I have used while working this exam are the class notes, my homework assignments, postings that the instructor has placed on Canvas or google drive, and discussions (verbal or emailed) with the instructor.
- I have not used any other resources, including books (printed or ebooks) or papers or any resources on the internet.

Name (please print): ______________________________

Signature: ______________________________ Date: ______________

Name: ____________________

Spatial Statistics; STAT 605, Spring 2022; Exam 2 (first part of take home exam)

For problems that require computation, be sure to show your work. Answers without supporting work will not receive full credit.

1. Definitions. For each of the following, define the term and state its importance in statistics (spatial statistics if the term is specific to spatial stats). (I expect 2-3 sentences for each of these, no more.)
   (a) edge effects (for point pattern data)

   (b) CSR (complete spatial randomness)

   (c) Monte Carlo tests (also, why are they so useful when working with point pattern data?)

2. We model CSR using a spatial Poisson process (for point pattern data). Consider a rectangular region R with $0 \leq x \leq 3$ and $0 \leq y \leq 2$.

   (a) If the intensity for a (homogeneous) Poisson process in this region is given by $\lambda(x, y) = 1.4$,

      i. What is the distribution of $N(R)$, the number of events in the region?

      ii. Find $P(N(R)) = 12$, the probability that there are 12 events in the region.

   (b) If the intensity for an inhomogeneous Poisson process in this region is $\lambda(x, y) = x + y$,

      i. Calculate $\gamma = \iint_R \lambda(x, y)\, dx\, dy$

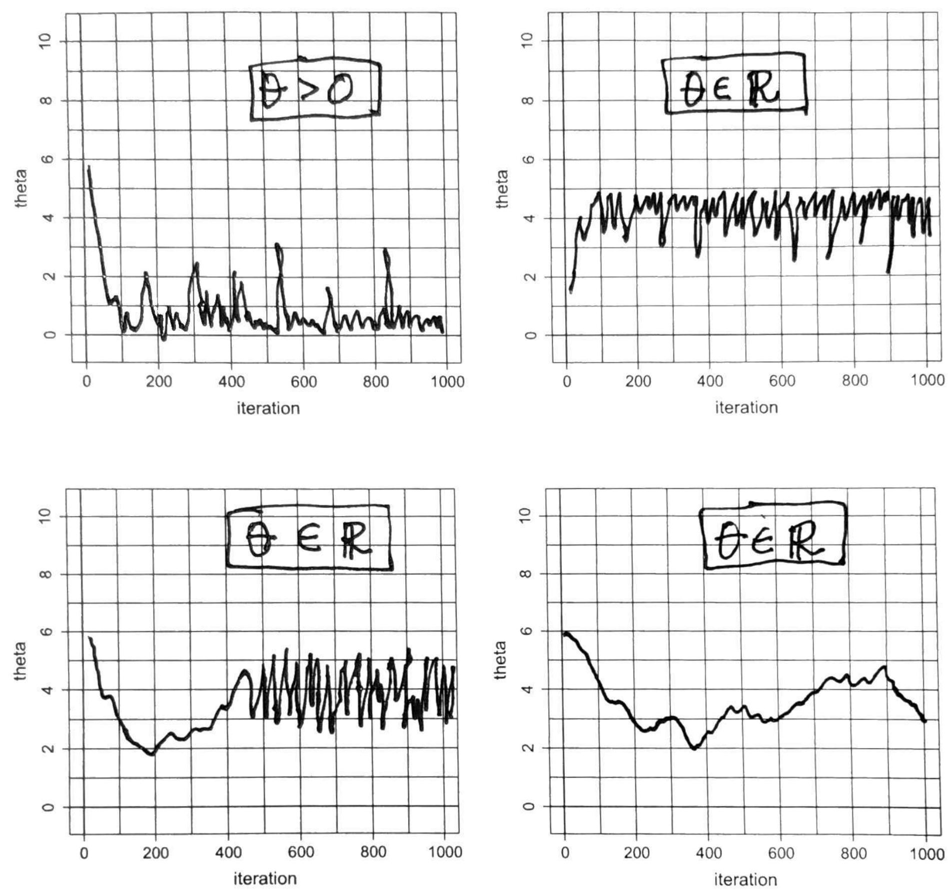
      ii. Find the distribution of $N(R)$.

      iii. What is the expected number of events in the region R?

3. For each of the following traceplots from a Bayesian analysis of a geostatistical spatial data set,

   - Identify whether the traceplot indicates that convergence has occurred.
   - State whether we should omit iterations as burn-in (**and if so, indicate how much**)
   - State whether we should use narrower or wider Metropolis proposals (or are do they seem appropriate); **If narrower, why will making them narrower help? If wider, why will making them wider help?**
   - Do you think we should run our MCMC for more iterations?

   A plot may indicate the need for two or more "fixes." If it indicates the need for a "fix", explain briefly.



4. I wish to conduct a formal test for CSR using $m = 11$ randomly selected quadrats. The counts are:

$$\{5, 6, 8, 5, 6, 5, 6, 5, 6, 6, 6\}$$

(a) I suspect from looking at these data that there is some evidence of regularity. Why? What is it about these data that seems at face value to be consistent with regularity? What would we expect to see (and why) if there were clustering?

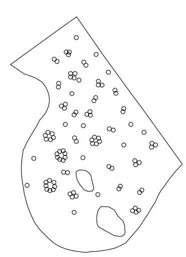
(b) For these data, the sample mean is 5.82 and the sample variance is 0.764. Carry out a hypothesis test of H_0 : CSR versus H_a: regularity, using $\alpha = 0.05$. Be sure to calculate the p-value. Be sure to state your conclusions in layman's terms.

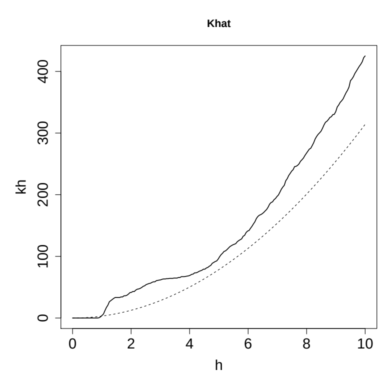
5. Use the plots below to answer the following questions about the locations of 99 people sitting in Gordon Park having a picnic lunch on a sunny day. There are two holes in the region. Distances are in meters. (This is the `gordon` data in the `spatstat` package.)

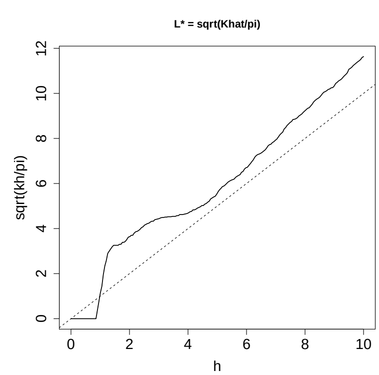
   (a) The first plot shows the locations of the picnickers. Comment on the plot; any unusual features? Do you see any strong evidence of clustering or regularity?

   (b) What is definition of "K-function"?

   (c) The second plot shows an estimated K-function (solid) line along with the K-function corresponding to CSR (dashed curve); the 3rd plot shows $L^* = \sqrt{K/\pi}$. Discuss the features of these plots; is there evidence of regularity or clustering at any scale? How can you tell? What explains the initial flat spot in the plot ($h < 1$m)? What explains the jump in the plot just to the right of $h = 1$m? What are you looking for in these plots?







6. For the picnic data on the previous page, I use a Monte Carlo test using the test statistic

$$A = 2\sqrt{\lambda}\bar{W}$$

where W = distance from a randomly selected event to another event.

(I use the Monte Carlo test because I am worried about edge effects.)

Use the R output below to test H_0 : CSR versus H_a : clustering. What is the value of my test statistic? What is the p-value? What is your conclusion?

```
> # find all nearest event distances:
> W <- nndistG( cbind(gordon2$x,gordon2$y) )$dists
> W[1:3]
[1] 4.940166 1.109053 1.141082
> min(W)
[1] 0.854578
> ( n.tot <- length(gordon2$x) )
[1] 99
> ( n.in.SRS <- round(n.tot/10) )
[1] 10
> # Here's the sample of events I'll use for the test:
> some.W <- sample(W, size=n.in.SRS, replace=FALSE)
> foo <- gordon2$window$bdry[[1]]
> ( area.of.park <- areapl(cbind(foo$x,foo$y)) )
[1] 2244.576
> ( lambda.hat <- n.tot / area.of.park )
[1] 0.04410633
> ( AA <- 2*sqrt(lambda.hat)*mean(some.W) )
[1] 0.6907354
> n.sims <- 2000
> A.results <- rep(NA,n.sims)
> for( i in 1:n.sims ) {
+   my.sim.data <- csr( mypoly, npoints=n.tot)
+   all.W <- nndistG( my.sim.data )$dists
+   some.W <- sample( all.W, size=n.in.SRS, replace=FALSE)
+   A.results[i] <- 2*sqrt(lambda.hat)*mean(some.W)
+ }
> median(A.results)
[1] 1.038397
> rank( c(AA,A.results) )[1]
[1] 37
```