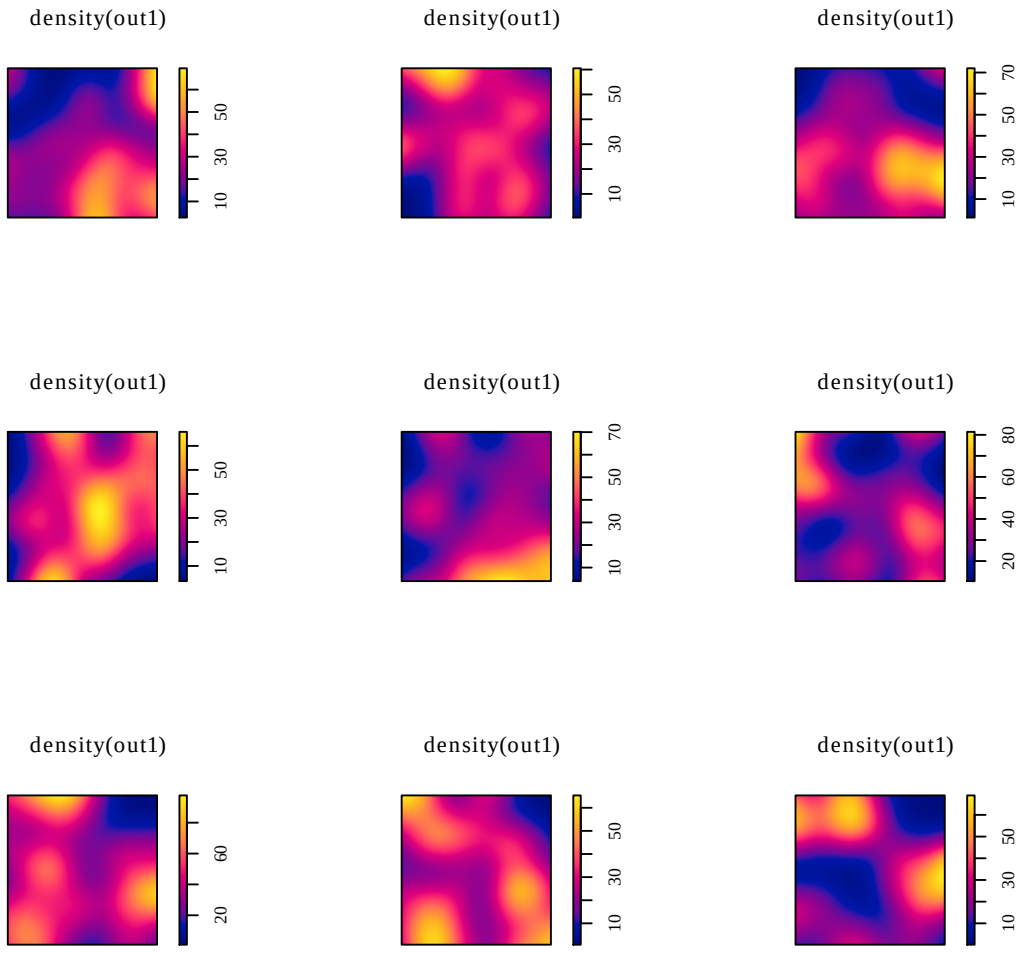
1. *For $\lambda = 30$ generate 9 realizations of CSR on the unit square. For each realization, construct a kernel estimate of $\lambda(s)$. How do the estimated intensity functions compare to the constant intensity under CSR? What precautions does this exercise suggest with regard to interpreting estimates of intensity from a single realization (or data set)? The following R code will simultaneously produce the data and plots.*

   par(mfrow=c(3,3)) for(i in 1:9) plot(density(rpoispp(30))) *Provide me with the images.*

   In four of the nine intensity functions created under CSR, there appears to be clear clusters, shown by the yellow "hot spots". Under CSR, we would not expect to see the clusters, and so the fact that these were created under CSR is concerning because a single realization under CSR may in fact violate CSR. Therefore, testing a null hypothesis of CSR using a single realization could lead to inaccurate conclusions. The inaccurate conclusions could happen many ways. If the true intensity function should be clustered, but the one CSR intensity function used to test is quite clustered, we would fail to reject CSR when we should have. If the true intensity function is CSR but the one CSR intensity function used to test the null hypothesis of CSR looks quite clustered, we may wrongly reject CSR.



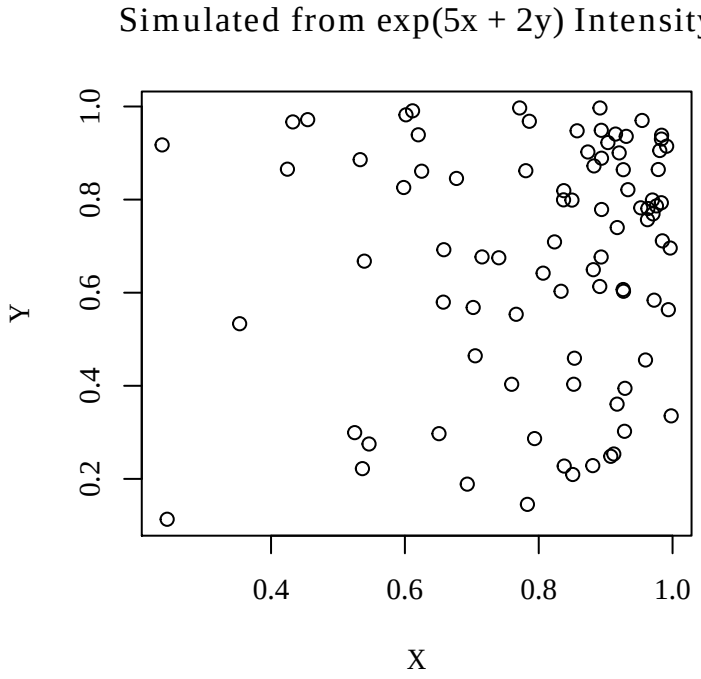
2. *In class we looked at a heterogeneous Poisson process on the unit square with with intensity function:*

$$\lambda(x, u) = exp(5x + 2y)$$

(a) *Simulate a realization of the process using the following R code.*

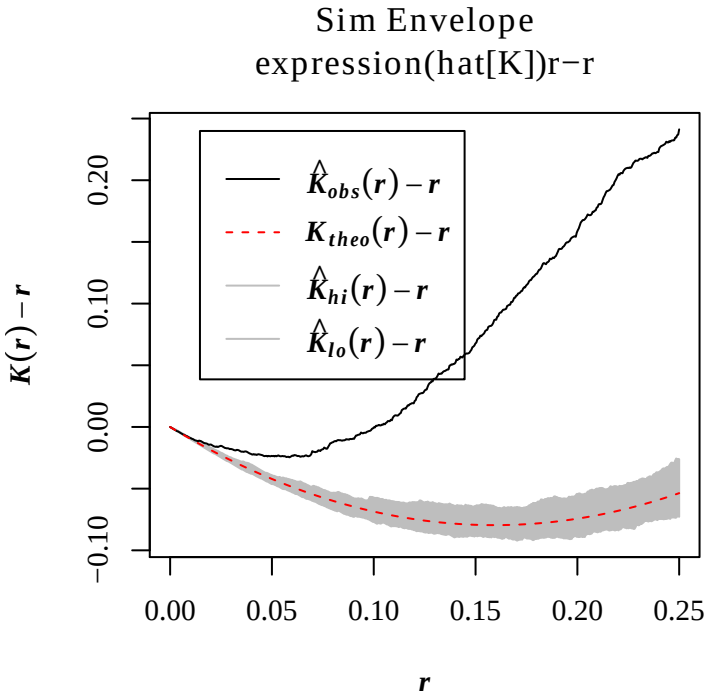
Plot the results and comment.

The intensity is highest in the upper right corner, near (1,1), and lowest when x ranges from 0 to 0.4 and y ranges from 0 to 1. The locations of the points do not look random.



(b) *Plot simulation envelopes for the K function (or some suitable modification of it) and comment.*

The plot below suggests strong evidence of clustering as hat(K(r)) - r is above the simulation envelope.



(c) *Fit a trend model to your data using ppm. Provide me with the parameter estimates and associated standard errors.*

	Estimate	S.E.
1	-0.17	0.56
2	4.89	0.58
3	2.24	0.42

ppm response is log(lambda) and explanatory variables are coordinates

Section: Lilihood methods for fitting models of spatially varying intensity surfaces – steve thinks this is neat

(d) *Check the fit using quadrat.test. Use method="MonteCarlo" instead of the large sample chi-squared test. Plot the results Discuss.*

With a p-value of 0.56, there is no evidence against CSR after accounting for the heterogenetiy in the intensity parameter across the region.

The largest discrepancies appear to be along the far right column when using the 5X5 grid, however, using Pearson's residual as a measure of the discrepancy between observed and expected show

quite low discrepancies despite the seemingly large absolute difference in that column between observed and expected.

sim.quad

0 0.4 −0.61	1 1.2 −0.18	5 2.8 1.3	8 8.7 −0.23	19 20.1 −0.25
0 0.3 −0.53	0 0.9 −0.94	1 2 −0.72	3 6.4 −1.3	21 14.9 1.6
0 0.1 −0.39	1 0.5 0.75	0 1.1 −1	5 3.5 0.77	5 8.1 −1.1
0 0.1 −0.33	0 0.4 −0.6	3 0.8 2.4	2 2.5 −0.34	9 5.9 1.3
0 0.1 −0.25	1 0.2 1.8	0 0.4 −0.66	2 1.4 0.48	0 3.2 −1.8

(e) *Compare these results from this model with those to a model fit under an assumption of CSR. Summarize the results. Provide me the model comparison results (AIC comparisons are fine).*

The default correction appears to be border.

```
sim.csr<-ppm(sim.dat,~1,correction="isotropic")
#anova(sim.csr, sim.ppm,test="Chi")

csr.aic <- AIC(sim.csr)
ppm.aic <- AIC(sim.ppm)

aic.all <- data.frame(cbind(csr.aic,ppm.aic))
colnames(aic.all) <- c("CSR", "Trend")
```

```
rownames(aic.all) <- c("AIC")

print(xtable(aic.all, align = "||l|l|l||"))
```

	CSR	Trend
AIC	-592.15	-729.31

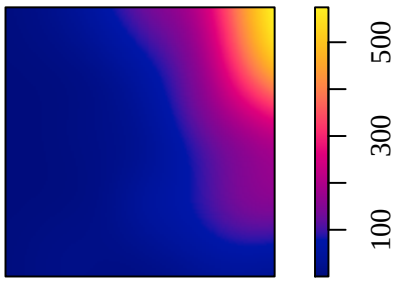
Although we failed to reject CSR after accounting for the varying intensity, the trend model has a lower AIC than the model assuming CSR and also accounts for the varying intensity and so is prefered.

(f) *Plot a nonparametric estimate of the intensity function. Compare the fitted surface you got using ppm with the nonparametric (kernel density estimate) surface in some suitable way.*

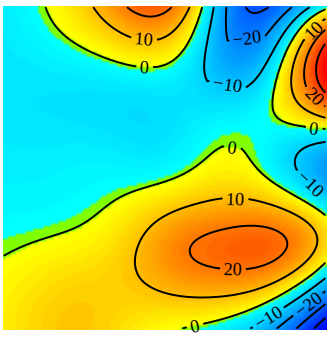
The blue represents negative residuals and the red represents positive residuals meaning that in blue regions the trend model is predicting fewer events than observed and in the red region the trend model is predicting more events than observed.

Compared to the nonparametric intensity function, the trend model is much more extreme than the intensity function and has predicted "hot spots" in areas that didn't have many observations. The trend model could be missing some predictors. The non-parametric function does much better in predicting more events in locations where more events were observed. That is, the upper right corner has the most events predicted, and the number of predicted events is slowly decreasing towards the lower left corner, which is what we observed in the data.

Nonparam Est Intensity Fn



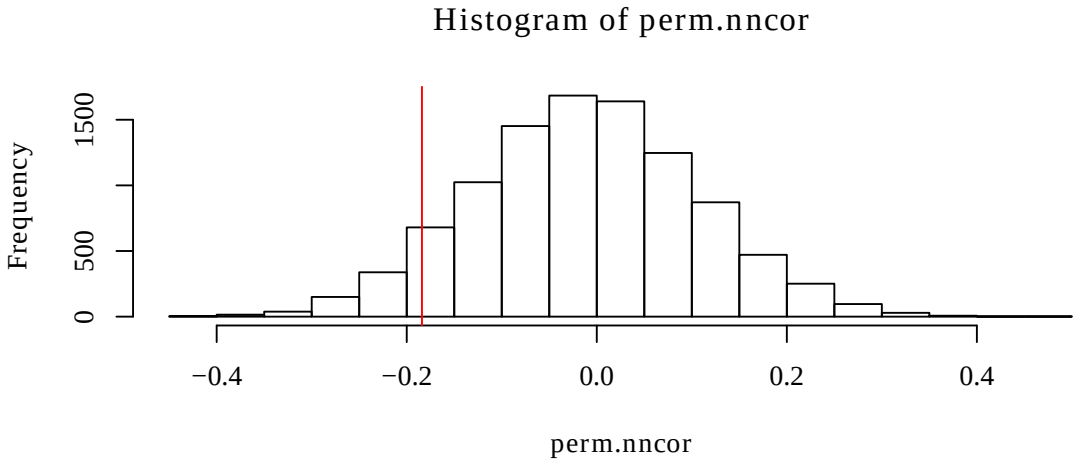
Smoothed raw residuals




```
Model diagnostics (raw residuals)
Diagnostics available:
```

```
smoothed residual field
range of smoothed field =  [-46.09, 37.24]
```

3. *Recall the use of the nncorr statistic in the Finland Pines data set. The distribution of heights (the marks) was of interest. We saw that the nearest neighbor correlation between heights was 0.1839798. We questioned whether or not this was unusual. Carry out a randomization test to assess this. You can use the rlabel command to scramble the marks if you want. Provide me with a histogram of the randomization distribution and a p-value. Discuss BRIEFLY your results. Provide me with your R-code, also. (Note - you need to extract the correlation from the nncorr output. Here is how to do that:*



H_0 : Pearson correlation = 0

H_a: Pearson correlation $\neq$ 0

The observed Pearson correlation between the nearest neighbor distance and the height was -0.184. A permutation test based on 10000 permutations of height led to a two sided p-value of 0.1207 which provides no evidence of a structured distribution of heights. Discussion with subject matter experts prior to data analysis may have lead to a one sided p-value.

4. *Lets derive a K function for something other than a CSR process. We will assume a Neyman- Scott process with the following properties. i The parent process is a homogeneous Poisson process with intensity . ii The number of off pring produced by each parent (N) is homogeneous Poisson with intensity μ. iii The position of each ospring is determined by a bivariate normal distribution with mean (0,0) (i.e. it is centered over the parent) and variance-covariance matrix 2I. Note that this implies that the x and y coordinates are determined independently of one another with the same variance.*

See attached.