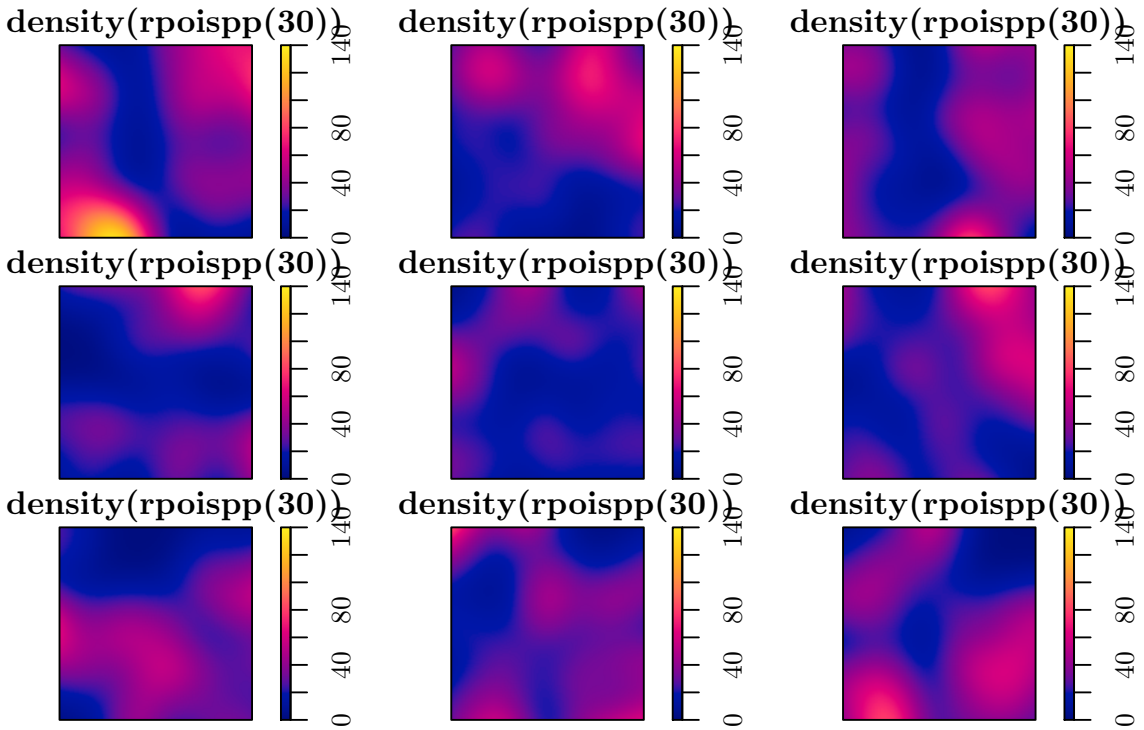
Stat 534 Homework 4

Kenny Flagg

February 10, 2017

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.*

```
library(spatstat, quiet = TRUE)
par(mfrow = c(3, 3), mar = c(0, 0, 1, 1), cex = 1)
for(i in 1:9) plot(density(rpoispp(30)), zlim = c(0, 140)) # zlim sets the color scale.
```



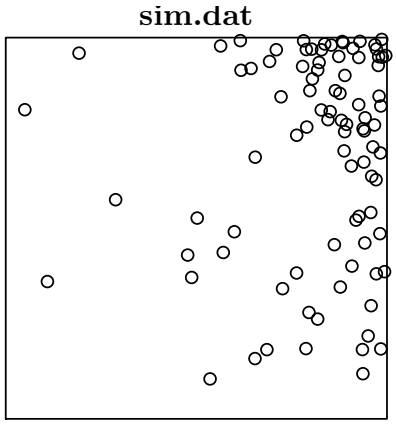
The estimated intensity functions do not look constant. Most of them have some hotspots where the estimated intensity is at least 60, and most also have some cold spots where there are few points so the estimated intensity is close to zero. This suggests we shouldn't make much out of our estimated intensity surface unless we have other information suggesting the process really is heterogeneous, and even then we should be aware that there may be a lot of uncertainty around the estimate.

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

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

(a) *Simulate a realization of the process using the following R code. Plot the results and comment.*

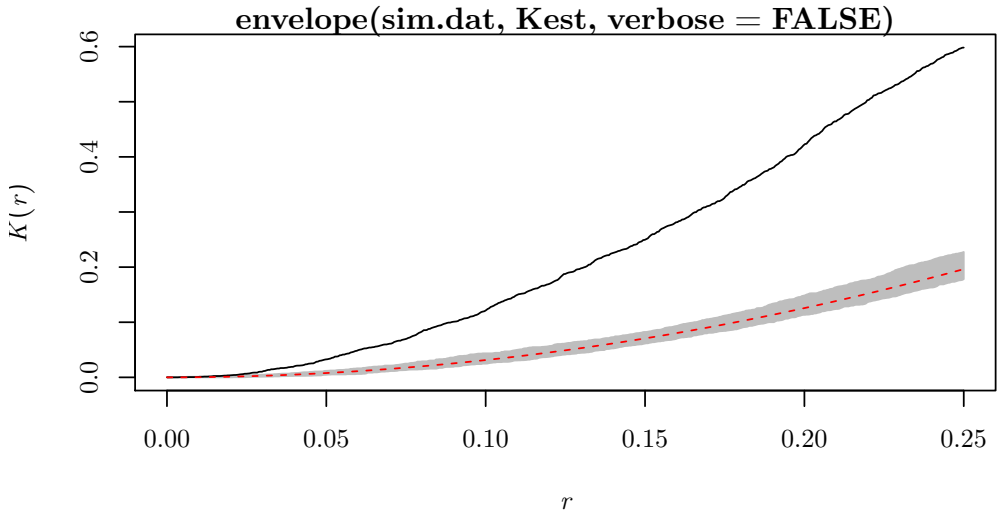
```
sim.dat <- rpoispp(function(x, y){exp(5 * x + 2 * y)})
par(mar = c(0, 0, 0.5, 0))
plot(sim.dat)
```



The plot looks about how I expected it to look. There are few events at the bottom left, many events at the top right, a noticeable increase in event intensity from left to right, and a subtle increase in intensity from bottom to top (easily seen at the right edge).

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

```
par(mar = c(4, 4, 1, 4))
plot(envelope(sim.dat, Kest, verbose = FALSE), legend = FALSE)
```



The empirical K function stays far outside of the simulation envelope for all distances except very small ones. There tend to be more events within a given distance than expected under CSR, but this is not the same pattern we've seen for clustered processes, where the K function is large for small distances but similar to the CSR K function for larger distances. This K function is close to the CSR K function at small distances, but moves away from the value expected under CSR for larger distances because, for most locations, any region other than a small neighborhood will include points with higher intensity.

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

```
sim.fit <- ppm(sim.dat, ~ x + y, correction = 'isotropic')
sim.fit

Nonstationary Poisson process

Log intensity:  ~x + y

Fitted trend coefficients:
(Intercept)           x           y
 -0.5599143   5.1565432   2.5606631

              Estimate      S.E.   CI95.lo   CI95.hi Ztest       Zval
(Intercept) -0.5599143 0.5759417 -1.688739 0.5689108       -0.9721717
x            5.1565432 0.5951682  3.990035 6.3230514   ***  8.6640102
y            2.5606631 0.4264484  1.724840 3.3964865   ***  6.0046261
```

The coefficient estimates, $\widehat{\beta}_0 = -0.56$ (SE $= 0.576$), $\widehat{\beta}_x = 5.16$ (SE $= 0.595$) and $\widehat{\beta}_y = 2.56$ (SE $= 0.426$), are pretty close to the true values, with the true values being in the 95% confidence intervals.

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

```
sim.qtest <- quadrat.test(sim.fit, method = 'MonteCarlo')
sim.qtest


Conditional Monte Carlo test of fitted Poisson model 'sim.fit' using
quadrat counts
Pearson X2 statistic

data:  data from sim.fit
X2 = 43.546, = NA, p-value = 0.098
alternative hypothesis: two.sided

Quadrats: 5 by 5 grid of tiles

par(mar = c(0, 0, 1, 0))
plot(sim.qtest)
points(sim.dat, pch = 19, col = '#00000040')
```

sim.qtest

2 0.4 2.7	0 1.2 −1.1	1 2.8 −1.1	10 9.2 0.26	28 23.5 0.92
0 0.2 −0.49	0 0.8 −0.9	0 2 −1.4	3 6.6 −1.4	15 15.6 −0.16
0 0.1 −0.35	1 0.4 0.91	4 1 3	0 3.3 −1.8	7 8.2 −0.42
1 0.1 3.1	0 0.3 −0.54	1 0.7 0.37	3 2.3 0.43	6 5.5 0.19
0 0 −0.21	0 0.1 −0.38	1 0.4 1.1	3 1.2 1.7	3 2.9 0.072

The p-value of 0.098 is small enough to make you think, but provides at best weak evidence that the model is a poor fit. The only standardized residuals with absolute value > 2 are in the leftmost column of quadrats and one right in the center. I wouldn't worry about the left edge because the expected counts are so close to zero that even 1 event would have a big residual. In the center, the cluster of points with the standardized residual of 3 isn't all that unusual because any points in a low-intensity region are going to stand out. If I encountered these data in the field I would see if a more complicated surface would fit better, but overall the model fits pretty well. With 25 quadrats, some of them are bound to be in the tails of the distribution.

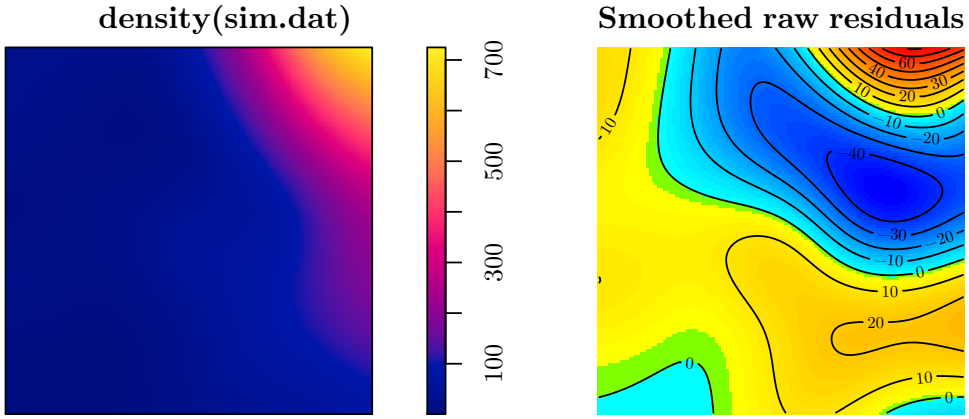
(e) *Compare these results from this model with those to a model fit under an assumption of CSR. Summarize the results. Provide me the model comparison results (AIC comparisons are fine).*

```
sim.csr <- ppm(sim.dat, ~ 1, correction = 'isotropic')
sim.csr

Stationary Poisson process
Intensity: 89
            Estimate      S.E.  CI95.lo  CI95.hi Ztest     Zval
log(lambda) 4.488636 0.1059998 4.280881 4.696392   *** 42.34571

AIC(sim.csr)

[1] -618.9773

AIC(sim.fit)

[1] -776.8786
```

The CSR model (intercept-only) assumes a constant intensity, estimated to be $\widehat{\lambda} = \exp(4.49) = 89$ with a 95% confidence interval of $\exp(4.28) = 72.3$ to $\exp(4.70) = 110$. The heterogeneous model fits much better, with an improvement in AIC of 157.9.

(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.*

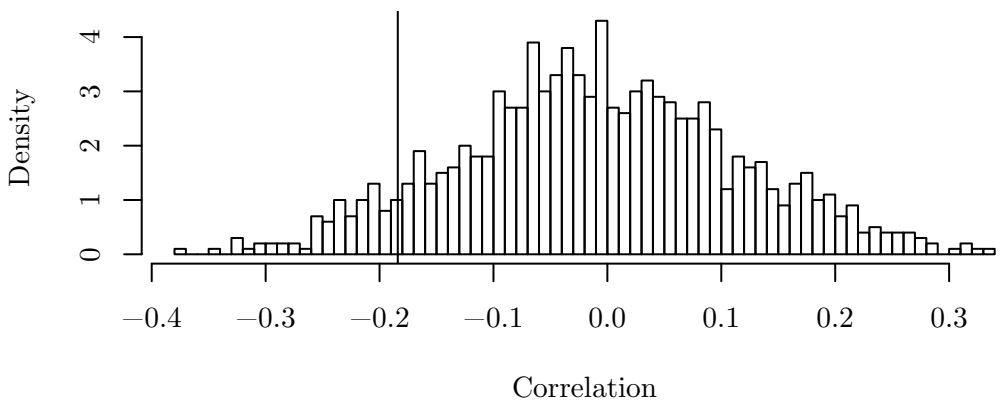
```
par(mfrow = c(1, 2), mar = c(0, 0, 1, 0))
plot(density(sim.dat))
diagnose.ppm(sim.fit, which = 'smooth')
```



The plot on the left shows a kernel density estimate of the intensity surface, and the plot on the right shows the smoothed residuals, computed as the fitted intensity surface minus a kernel density estimate. The residuals have bands alternating positive and negative, indicating that the model missed some of the patterns in the data.

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.*

```
finpines.ht <- finpines
marks(finpines.ht) <- marks(finpines)$height
obs_corr <- nncorr(finpines.ht)['correlation']
perm_corr <- c(obs_corr, replicate(999, nncorr(rlabel(finpines.ht))['correlation']))
hist(perm_corr, breaks = 100, freq = FALSE, xlab = 'Correlation',
     main = 'Permutation Distribution of Nearest Neighbor Correlation')
abline(v = obs_corr)
```




```
pval <- 2 * mean(perm_corr <= obs_corr)
pval
```

```
[1] 0.162
```

999 permutations yield an approximately symmetric permutation distributon, so doubling the left-tail p-value results in a two-sided p-value of 0.162 giving no evidence that the heights or nearest neighbors are correlated.

4. *Let's 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* λ.

   ii. *The number of offspring produced by each parent (N) is homogeneous Poisson with intensity* μ.

   iii. *The position of each offspring is determined by a bivariate normal distribution with mean* $(0, 0)$ *(i.e. it is centered over the parent) and variance-covariance matrix* $\sigma^2\mathbf{I}$. *Note that this implies that the* x *and* y *coordinates are determined independently of one another with the same variance.*

   Consider 2 offspring from the same parent located at (X_1, Y_1) *and* (X_2, Y_2).

   (a) *What is the distribution of*

   $$W = \frac{(X_1 - X_2)^2}{2\sigma^2} + \frac{(Y_1 - Y_2)^2}{2\sigma^2}?$$

   X_1, X_2, Y_1, and Y_2 are all independent $\text{N}(0, \sigma^2)$, so $X_1 - X_2$ and $Y_1 - Y_2$ are independent $\text{N}(0, 2\sigma^2)$. Then $\frac{(X_1 - X_2)^2}{2\sigma^2}$ and $\frac{(Y_1 - Y_2)^2}{2\sigma^2}$ are independent χ^2_1. Therefore, $W \sim \chi^2_2$.

(b) *Note that the Euclidean distance between the 2 points is*

$$H = \left[(X_1 - X_2)^2 + (Y_1 - Y_2)^2)\right]^{1/2} = \left(2\sigma^2 W\right)^{1/2}.$$

Derive the cdf of H.

$$\begin{aligned}
F(h) &= P(H \leq h) \\
&= P\left(\left(2\sigma^2 W\right)^{1/2} \leq h\right) \\
&= P\left(W \leq \frac{h^2}{2\sigma^2}\right) \\
&= \int_0^{\frac{h^2}{2\sigma^2}} \frac{t^0 e^{-\frac{t}{2}}}{\Gamma(1)\, 2^1} \mathrm{d}t \\
&= \int_0^{\frac{h2}{2\sigma^2}} \frac{1}{2} e^{-\frac{t}{2}} \mathrm{d}t \\
&= \left[-e^{-\frac{t}{2}}\right]_{t=0}^{\frac{h^2}{2\sigma^2}} \\
&= 1 - \exp\left(-\frac{h^2}{4\sigma^2}\right)
\end{aligned}$$

(c) *Recall that we were told that the K function for Neyman-Scott processes with homogeneous Poisson parent processes and radially symmetric $f(h)$ is*

$$K(h) = \pi h^2 + \frac{E(N(N-1))}{\lambda E(N)^2} F(h).$$

Use the results from above to find the K function for the described process.

$$\begin{aligned}
K(h) &= \pi h^2 + \frac{E(N(N-1))}{\lambda E(N)^2}\left(1 - \exp\left(-\frac{h^2}{4}\right)\right) \\
&= \pi h^2 + \frac{E(N^2) - E(N)}{\lambda E(N)^2}\left(1 - \exp\left(-\frac{h^2}{4}\right)\right) \\
&= \pi h^2 + \frac{Var(N) + E(N)^2 - E(N)}{\lambda E(N)^2}\left(1 - \exp\left(-\frac{h^2}{4}\right)\right) \\
&= \pi h^2 + \frac{\mu + \mu^2 - \mu}{\lambda \mu^2}\left(1 - \exp\left(-\frac{h^2}{4}\right)\right) \\
&= \pi h^2 + \frac{1}{\lambda}\left(1 - \exp\left(-\frac{h^2}{4}\right)\right)
\end{aligned}$$