Lab 3 Pattern Analysis-Hypothesis Testing

Dominic LaRoche

September 16, 2014

1 Assigment 1

Based on the inclusion of simulation results (fig. 1) we can conclude that the japanese pine sapling data is not different from complete spatial randomness (CSR).

```
rm(list=ls())
library(spatstat)

##
## spatstat 1.38-1        (nickname:  'Le Hardi')
## For an introduction to spatstat, type 'beginner'

data(japanesepines)
jp<-japanesepines
```
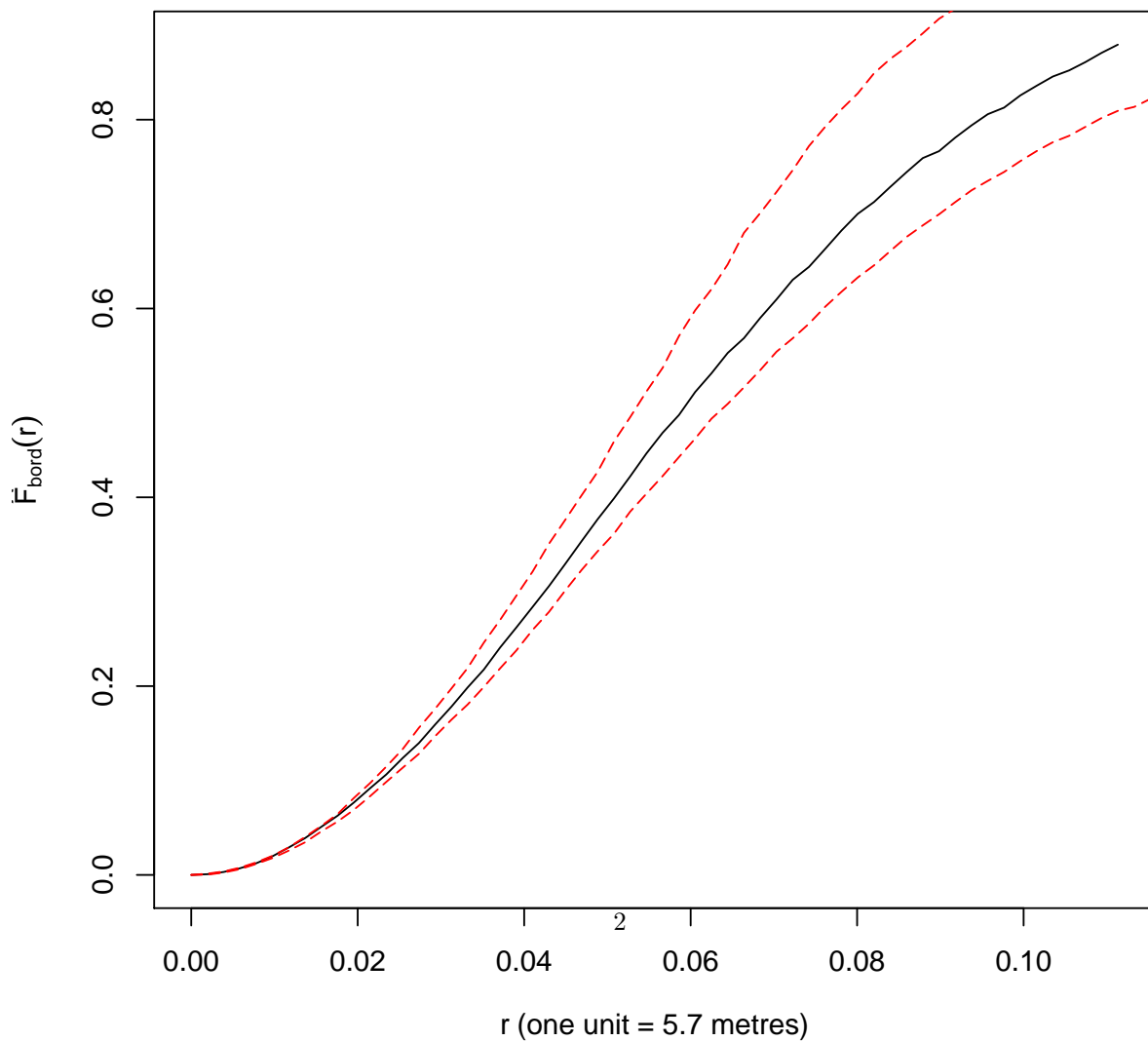
```
fhat.env<-function(n, s, r, win=owin(c(0,1),c(0,1))){
  #function to create upper and lower CSR bounds on Fest
  hold<-matrix(0,s, length(r))
  for (i in 1:s){
    hold[i,]<-Fest(runifpoint(n,win=win), r=r)$rs
  }
  mn<-apply(hold, 2, mean)
  Up<-apply(hold,2,max)
  Down<-apply(hold,2,min)
  return(data.frame(mn,Up, Down))
}

jp.fhat<-Fest(jp)
jp.win<-window(jp)
plot(jp.fhat, rs~r, main="F estimates")
jp.fenv<-fhat.env(n=jp$n, s=100, r=jp.fhat$r, win=jp.win)
#upper and lower envelopes
lines(jp.fhat$r, jp.fenv$Up,lty=5, col=2)
lines(jp.fhat$r,jp.fenv$Down,lty=5, col=2)
```

F estimates

2   Assignment II

Here I define a function to get upper and lower bounds for the K function:

```r
khat.env<-function(n, s, r, win=owin(c(0,1),c(0,1))){
  #function to create upper and lower CSR bounds on Fest
  hold<-matrix(0,s, length(r))
  for (i in 1:s){
    hold[i,]<-Kest(runifpoint(n,win=win), r=r)$border
  }
  mn<-apply(hold, 2, mean)
  Up<-apply(hold,2,max)
  Down<-apply(hold,2,min)
  return(data.frame(mn,Up, Down))
}
```

Based on the plot of the K-function with upper and lower bounds created from simulated CSR data (fig. 2), we can conclude that the Japanese pine saplings are not significantly different from CSR.

```
jp.khat<-Kest(jp)
plot(jp.khat, cbind(border, theo)~r, main="K function for JP data")

##        lty col    key            label                            meaning
## border   1   1 border hat(K)[bord](r) border-corrected estimate of K(r)
## theo     2   2   theo     K[pois](r)          theoretical Poisson K(r)

jp.kenv<-khat.env(n=jp$n, s=100, r=jp.khat$r, win=jp.win)
#upper and lower envelopes
lines(jp.khat$r, jp.kenv$Up,lty=5, col=3)
lines(jp.khat$r,jp.kenv$Down,lty=5, col=3)
```
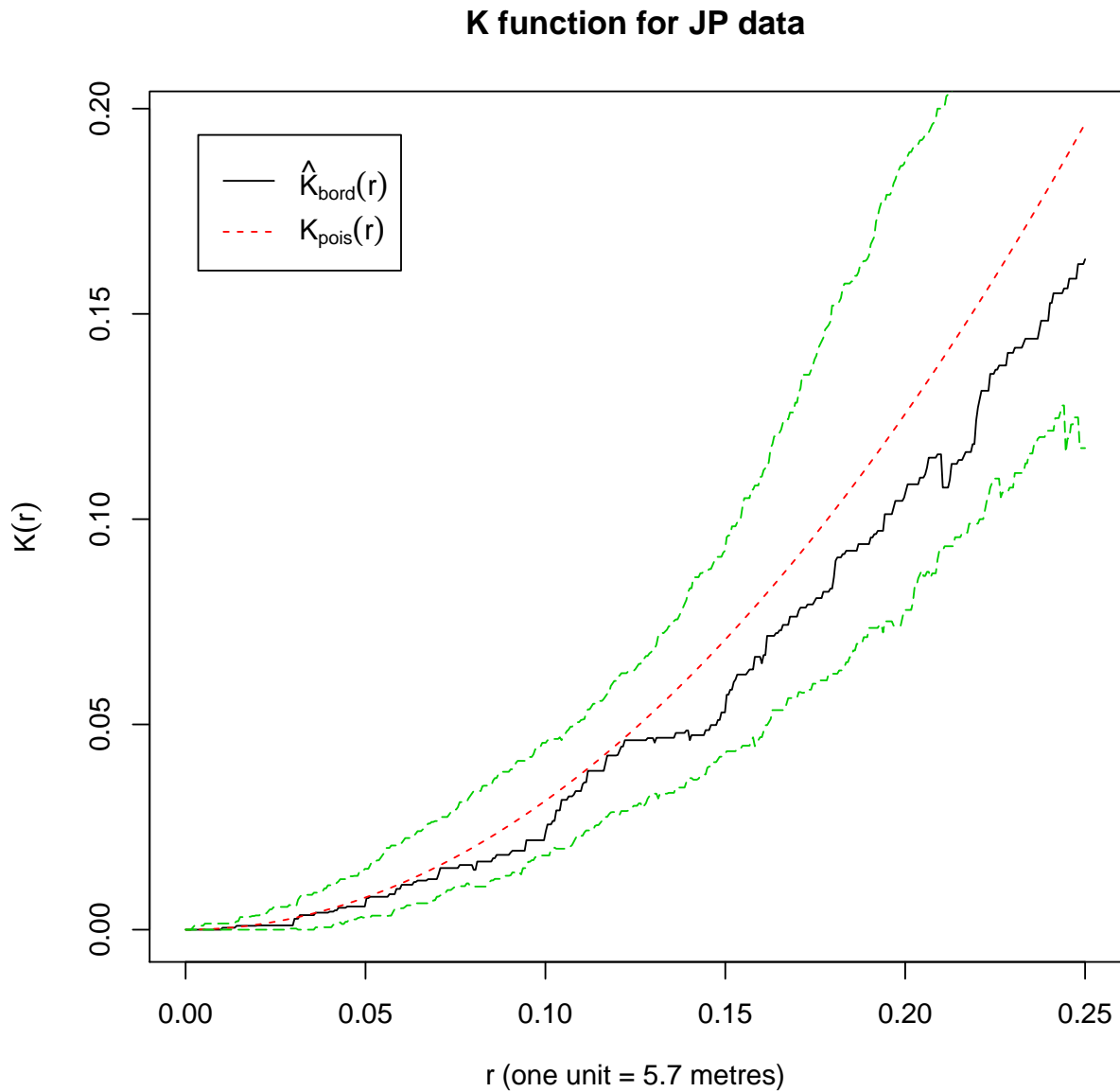

Figure 2: Plot of the $\hat{K}$ function with upper and lower bounds created from simulated CSR data. The red dashed line represents the theoretical K-function for a poisson CSR pattern.

4

3 Assignment III

```
data(redwood) #load the dataset
# summary(redwood)
regular<-rsyst(nx=10) #generate regular pattern
# summary(regular)
rdwd<-redwood
```

3.1 Quadrat Tests

We will test for a first order effect using two different quadrat tests, one with 4 cells (2x2) and one with 16 (4x4). The two grids are shown in figure 3.

We cannot reject the null hypothesis of complete spatial randomness with the 2x2 quadrat tests ($\alpha \geq 0.1$). However, the 4x4 quadrat test produced a highly significant result (p-value = 0.0004). Figures 4 and 5 show the counts and expected values for each cell.

```
qt.rdwd2<-quadrat.test(rdwd,2,2)
qt.rdwd2

##
##  Chi-squared test of CSR using quadrat counts
##  Pearson X2 statistic
##
## data:  rdwd
## X2 = 6.516, df = 3, p-value = 0.1781
## alternative hypothesis: two.sided
##
## Quadrats: 2 by 2 grid of tiles

qt.rdwd4<-quadrat.test(rdwd,4,4)

## Warning:  Some expected counts are small; chi^2 approximation may be inaccurate

qt.rdwd4

##
##  Chi-squared test of CSR using quadrat counts
##  Pearson X2 statistic
##
## data:  rdwd
## X2 = 42.26, df = 15, p-value = 0.0004102
## alternative hypothesis: two.sided
##
## Quadrats: 4 by 4 grid of tiles
```
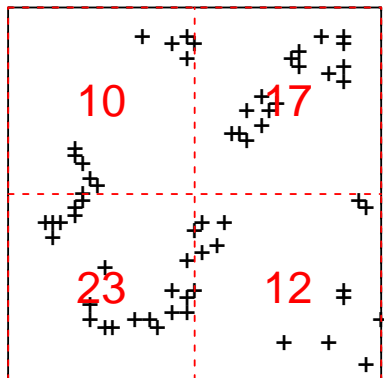
```
par(mfrow=c(1,2))
#2x2
q.rdwd2<-quadratcount(rdwd,nx=2,ny=2)
plot (rdwd, pch="+", main = " 2x2 Quadrat for redwood data")
plot (q.rdwd2, add=TRUE, col="red", cex=1.5, lty=2)
#4x4
q.rdwd4<-quadratcount(rdwd,nx=4,ny=4)
plot (rdwd, pch="+", main = "4x4 Quadrat for rdwd data")
plot (q.rdwd4, add=TRUE, col="red", cex=1.5, lty=2)
```
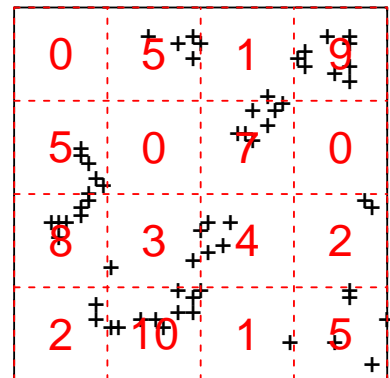

Figure 3: Quadrat counts for the California redwood data.

```
plot(qt.rdwd2, main="Quadrat Test: 2x2 rdwd data")
plot(rdwd, pch="+", col="red", add=T)
```
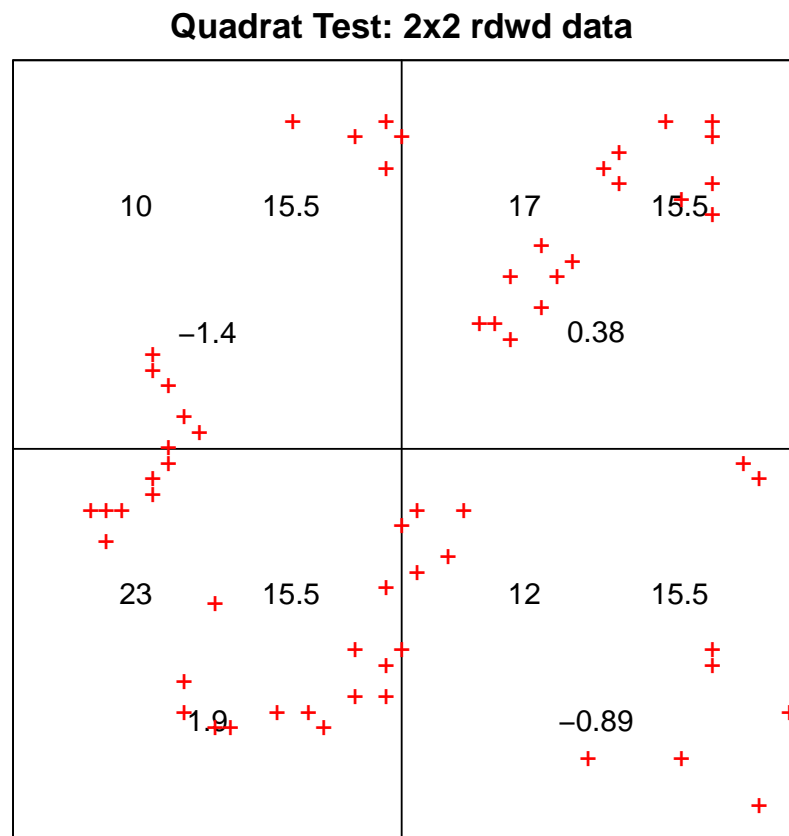
Quadrat Test: 2x2 rdwd data

Figure 4: A plot of the test for the 2x2 quadrat test.

```
plot(qt.rdwd4, main="Quadrat Test: 4x4 rdwd data")
plot(rdwd, pch="+", col="red", add=T)
```
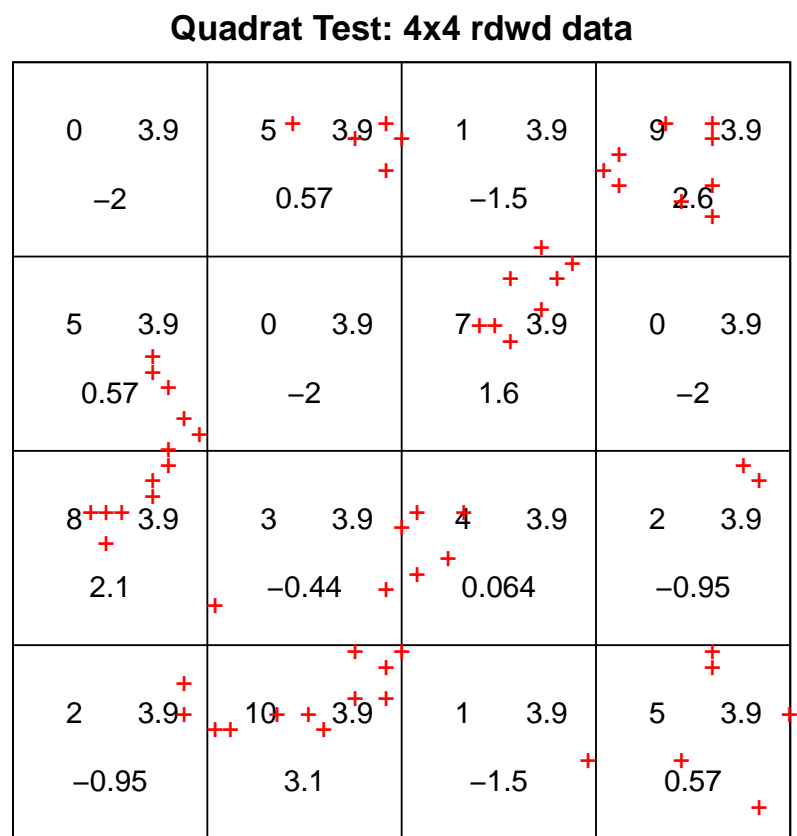
Quadrat Test: 4x4 rdwd data

0   3.9
 	
−2
	5 + 3.9 +
	       +
	        +
	0.57
		1 3.9
		
		−1.5
			9+ +3.9
			+
			+ +
			 + 2.6 +
			            +

5   3.9
   +
0.57 +
      +
       +
	0 3.9
	
	−2
		+
		 + +
		7++ 3.9
		   +
		1.6
			0 3.9
			
			−2

8+++ 3.9
  +
2.1
	3 3.9
	
	−0.44 +
	+
		+
		 4 3.9
		  +
		0.064
			+
			  +
			2 3.9
			
			−0.95

+
2   3.9+
++
−0.95
	+ +
	  +
	10+ + 3.9 +
	3.1
		1 3.9
		      +
		−1.5
			+
			5 3.9 +
			+
			0.57
			    +

Figure 5: A plot of the test for the 4x4 quadrat test.

3.2 Kernel estimation

Figure 6 shows two different realizations of kernel density estimation with bandwidths equal to 0.05 (top) and 0.1 (bottom) for both the Japanese pine sapling data and the California redwood data. Both bandwidth values appear to indicate clustering of the points in the redwood data (bottom) whereas the pine data (top) is not as well captured by either kernel.

```r
# ?density.ppp
par(mfrow=c(2,2))
jp.Z.0.5<-density.ppp(jp, 0.05)
plot(jp.Z.0.5, main="JP: sigma=0.05")
points(jp, pch="+", col="6")

jp.Z.1<-density.ppp(jp, 0.1)
plot(jp.Z.1, main="JP: sigma =0.1")
points(jp, pch="+", col="6")

rdwd.Z.0.5<-density.ppp(rdwd, 0.05)
plot(rdwd.Z.0.5, main="RdWd: Sigma=0.05")
points(rdwd, pch="+", col="6")

rdwd.Z.1<-density.ppp(rdwd, 0.1)
plot(rdwd.Z.1, main="RdWd: Sigma =0.1")
points(rdwd, pch="+", col="6")
```
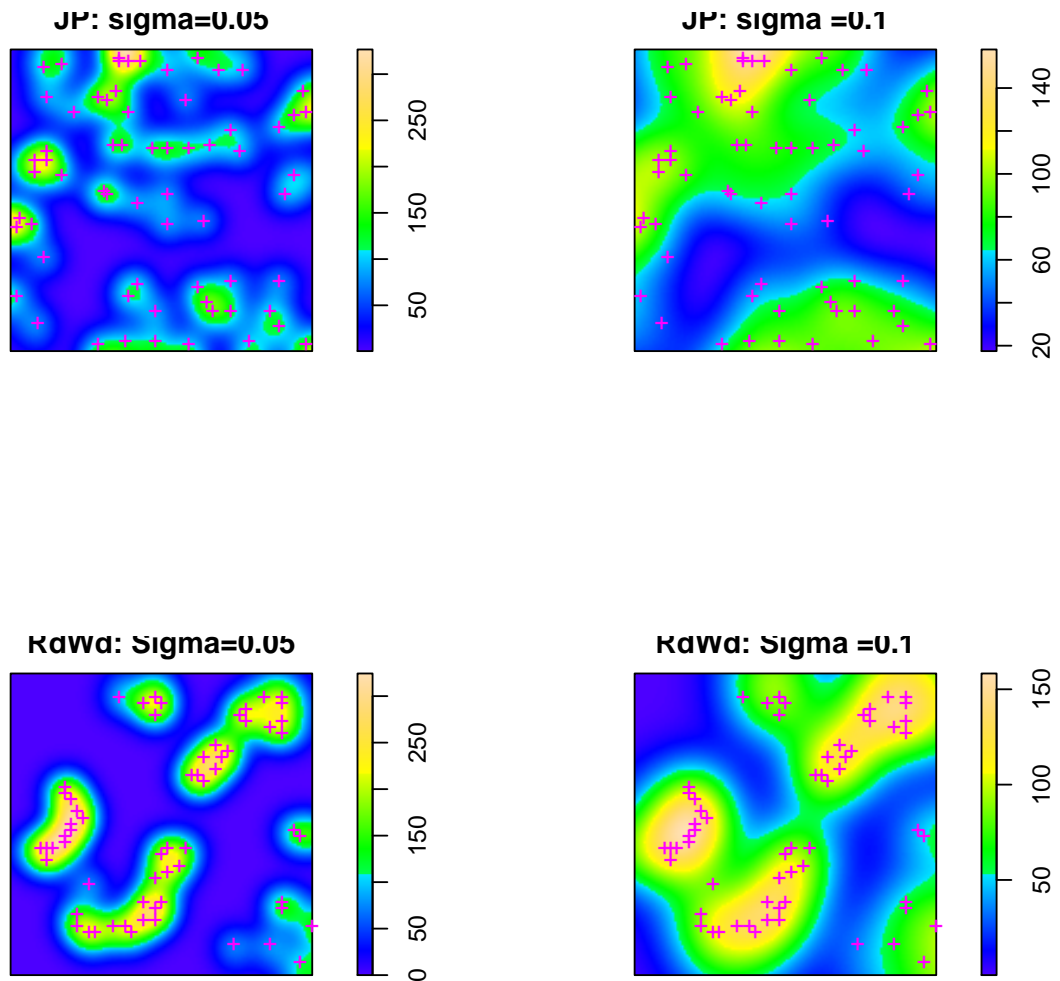
Figure 6: Two different kernel density estimation with sigma =0.05 and 0.1 for the Japanese pine sapling data (top) and the redwood data (bottom).

3.3 Second Order Effect

The plots of $\hat{G}$ (fig. 7) and $\hat{F}$ (fig. 8) both suggest that the pattern is not random for the California redwood data. In the $\hat{G}$ plot the ECDF lies above the upper and lower bounds generated from simulated CSR data. This suggests that the redwood saplings are clustered in space. The same inference is made from the $\hat{F}$ plot in which the distance between a random point and an observed point is larger than one would expect with CSR points. The $\hat{K}$ function plot (fig. 9) and the $\hat{L}$ function both Also indicate clustering in the California redwood data. For all plots references above the Japanese pine sapling data lies within the CSR bounds calculated from simulated CSR data. This indicates that we cannot reject the null hypothesis of CSR.

```
rdwd.ghat<-Gest(rdwd)
rdwd.win<-window(rdwd)
plot(rdwd.ghat, rs~r, main="G estimates")
plot(jp.ghat, rs~r, add=T, col=4)

rdwd.genv<-ghat.env(n=rdwd$n, s=100, r=rdwd.ghat$r, win=rdwd.win)
#upper and lower envelopes
lines(rdwd.ghat$r, rdwd.genv$Up,lty=5, col=2)
lines(rdwd.ghat$r,rdwd.genv$Down,lty=5, col=2)
legend("topleft",legend=c("Redwood","Pine","CSR Bounds"),col=c(1,4,2),lty=c(1,1,5))
```
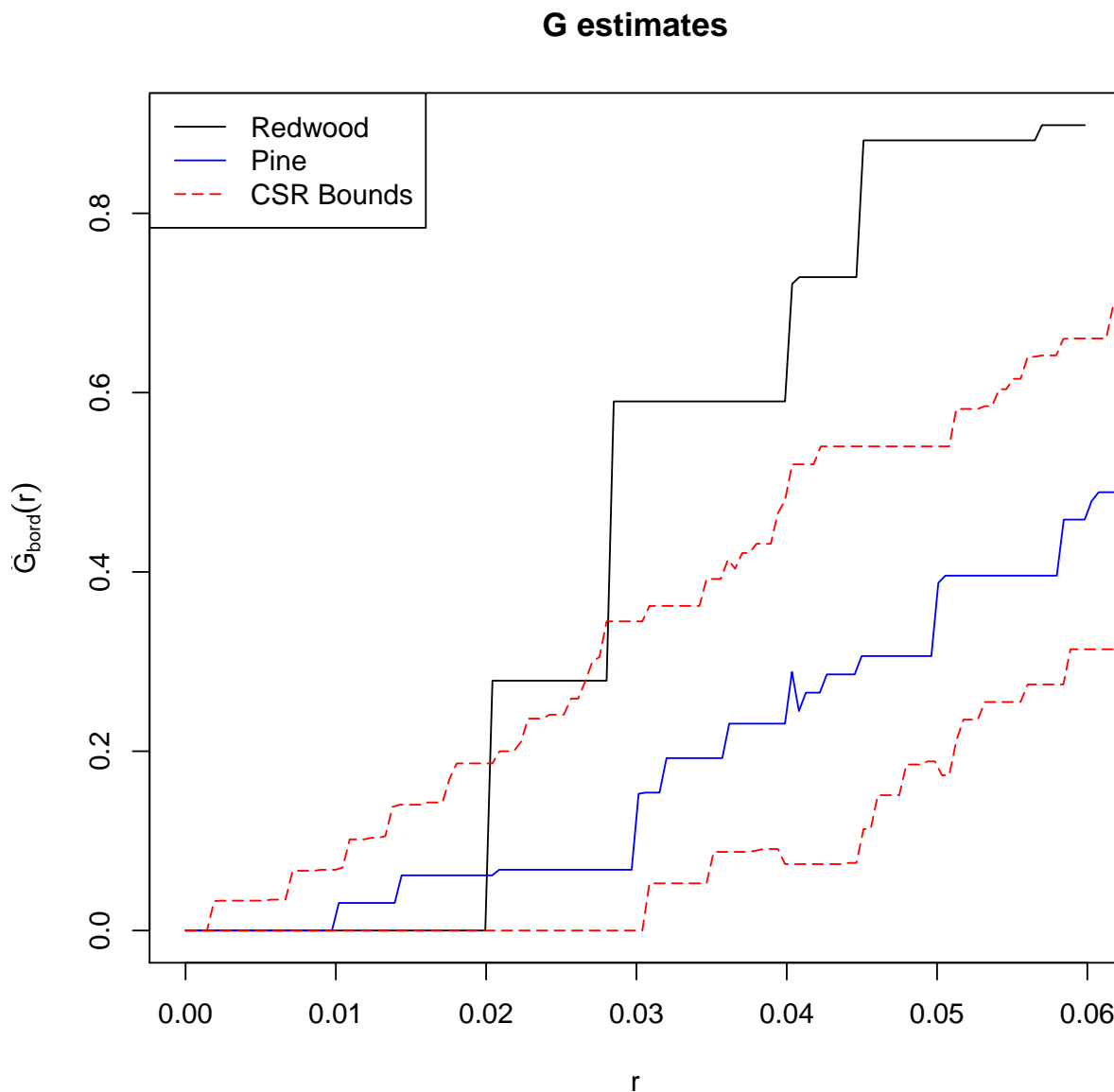
G estimates

Figure 7: Estimation of the G function using nearest neighbor distances and comparison with simulated upper and lower bounds of poisson CSR data. The black line represents the redwood data and falls outside the CSR bounds whereas the blue line represents the Japanese pine data and falls inside the CSR bounds

```
rdwd.fhat<-Fest(rdwd)
plot(rdwd.fhat, rs~r, main="F estimates")
plot(jp.fhat, rs~r, add=T, col=4)
rdwd.fenv<-fhat.env(n=rdwd$n, s=100, r=rdwd.fhat$r, win=rdwd.win)
#upper and lower envelopes
lines(rdwd.fhat$r, rdwd.fenv$Up,lty=5, col=2)
lines(rdwd.fhat$r,rdwd.fenv$Down,lty=5, col=2)
legend("topleft",legend=c("Redwood","Pine","CSR Bounds"),col=c(1,4,2),lty=c(1,1,5))
```
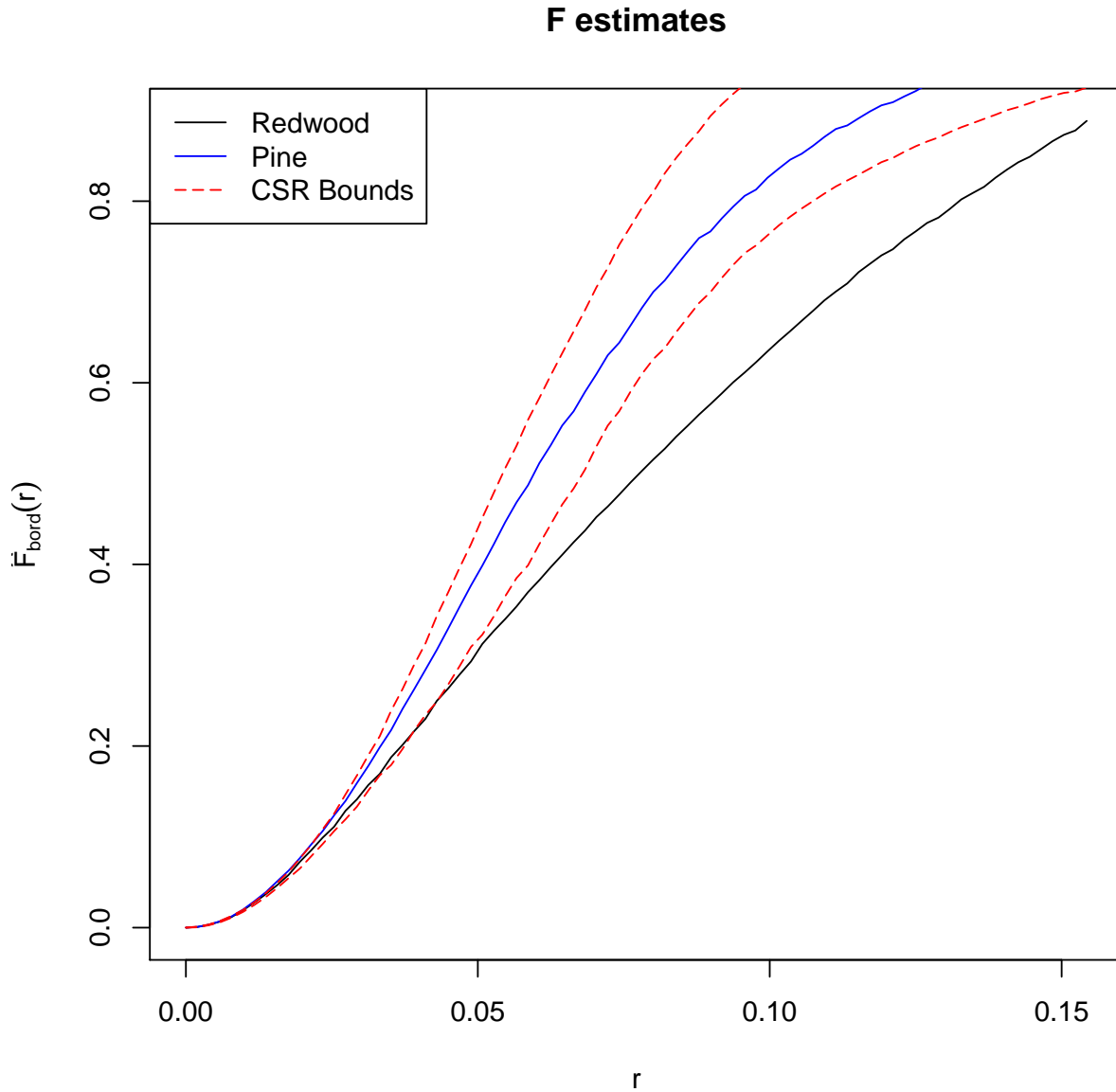
F estimates

Figure 8: Estimation of the F function using point-to-event distances and comparison with upper and lower bounds generated from CSR data for both the Japanese pine sapling data and the California redwood data. As before the redwood data lies outside the CSR bounds whereas the pine data lies inside.

```
rdwd.khat<-Kest(rdwd)
plot(rdwd.khat, border~r, main="K function for rdwd data")
plot(jp.khat, border~r, add=T, col=4)
rdwd.kenv<-khat.env(n=rdwd$n, s=100, r=rdwd.khat$r, win=rdwd.win)
#upper and lower envelopes
lines(rdwd.khat$r, rdwd.kenv$Up,lty=5, col=2)
lines(rdwd.khat$r,rdwd.kenv$Down,lty=5, col=2)
legend("topleft",legend=c("Redwood","Pine","CSR Bounds"),col=c(1,4,2),lty=c(1,1,5))
```
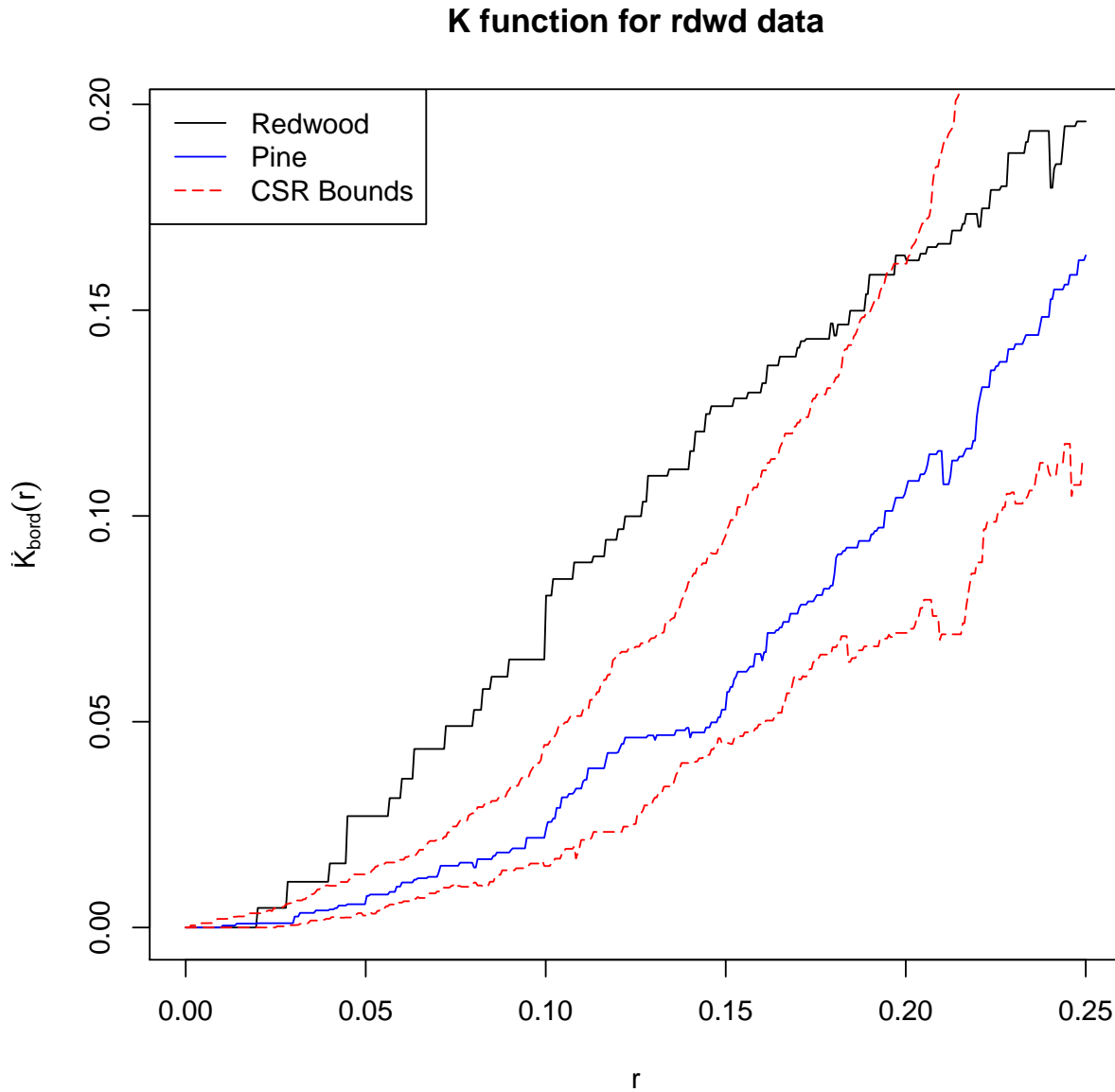

Figure 9: Plot of the K function for the redwood data showing the same pattern as the $\hat{G}$ and $\hat{F}$ plots, i.e. the redwood data appears clustered and the pine data lies within what we would expect from CSR data.

```
plot(rdwd.khat, sqrt(cbind(border, theo)/pi)-r~r,ylab="L(r)", main="L function",ylim=c(-.07,.07),lty=c(
plot(jp.khat, sqrt(border/pi)-r~r,add=TRUE, col=4)
lines(rdwd.khat$r, sqrt(rdwd.kenv$Up/pi)-rdwd.khat$r,lty=5, col=2)
lines(rdwd.khat$r, sqrt(rdwd.kenv$Down/pi)-rdwd.khat$r,lty=5, col=2)
legend("topleft",legend=c("Redwood","Pine","CSR Bounds","Pois."),col=c(1,4,2,3),lty=c(1,1,5,2))
```
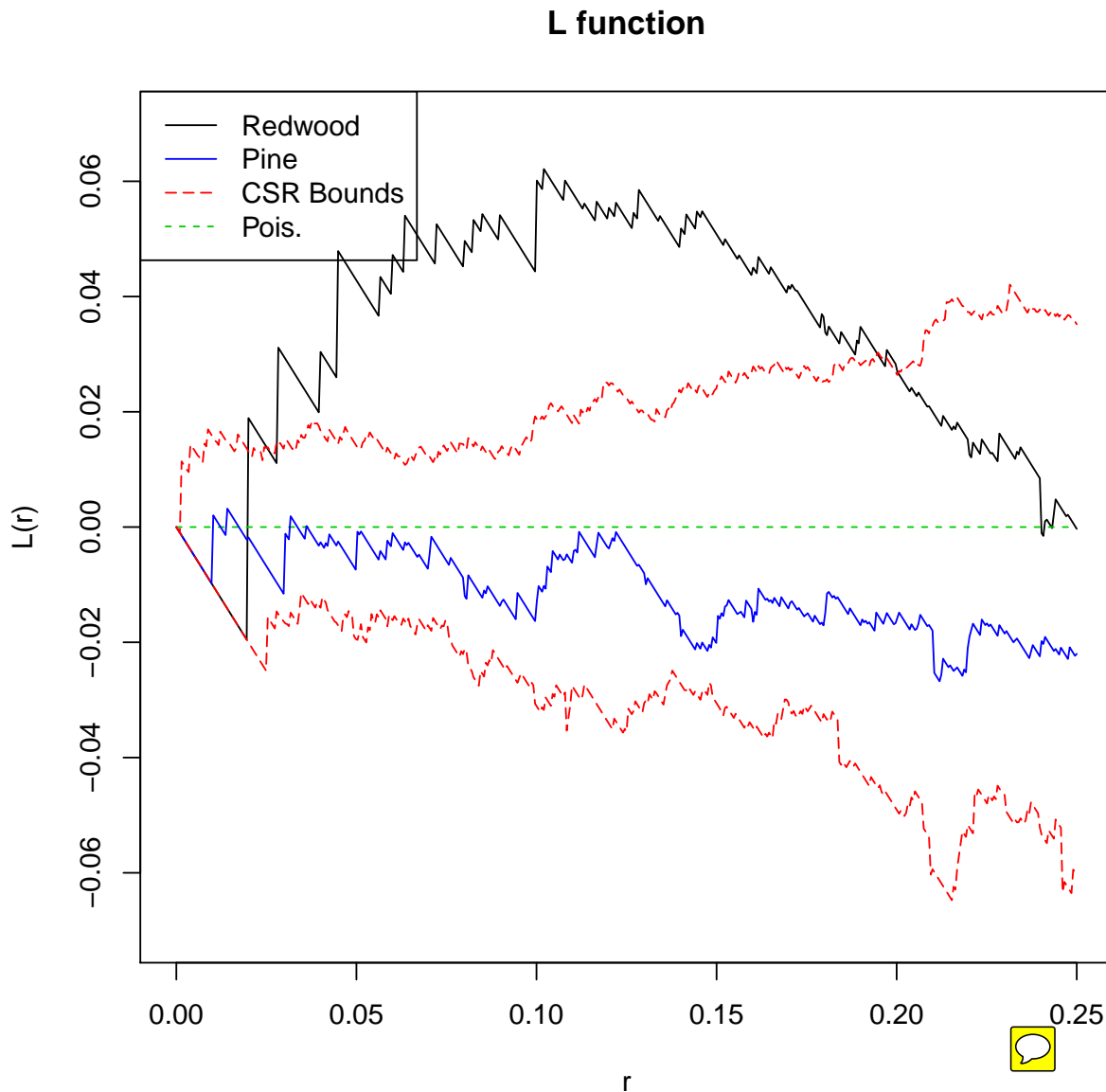

L function

Figure 10: The $\hat{L}$ plot of the california redwood and Japanese pine data indicating a cluster pattern in the redwood data and a random pattern in the pine data. The horizontal green lines shows the expected value for a poisson CSR realization.