R Notebook

TASK 4: Case-control studies

Using the Primary Biliary Cirrhosis data, perform a case-control point pattern analysis. Point locations and marks are stored in *PBCdata.txt" file. Coordinates of the window are in "PBCpoly.txt" file.

The data are:

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

pbc.data <- read.delim("T4/PBCdata2.txt")
head(pbc.data)
```

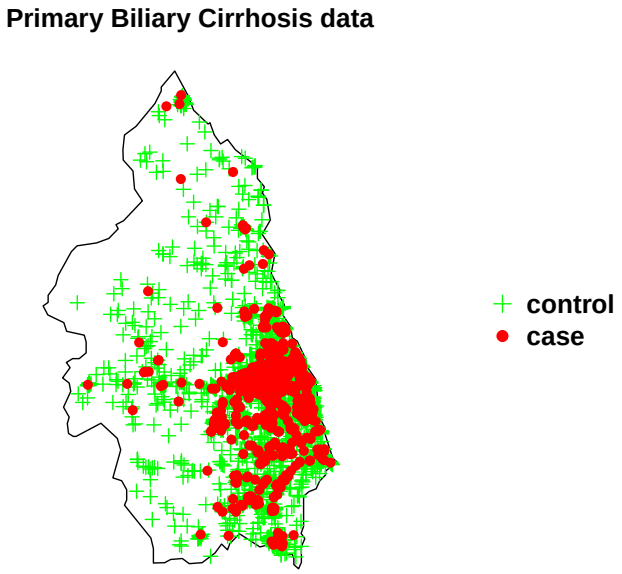
```
##       x     y marks
## 1 43830 56150  case
## 2 42840 56510  case
## 3 43740 54320  case
## 4 42480 56240  case
## 5 42280 56930  case
## 6 42610 55850  case
```

1) Create the ppp object.

```
poligon = read.delim("T4/PBCpoly.txt")
min.pX = min(poligon$x)
max.pX = max(poligon$x)
min.pY = min(poligon$y)
max.pY = max(poligon$y)

pol.illa <- list(x = poligon$x, y = poligon$y)

pbc.data$marks = factor(pbc.data$marks)

pbc.p = ppp(pbc.data$x, pbc.data$y, poly = pol.illa, range(poligon$x), range(poligon$y),
    marks = pbc.data$marks)
```

```
cases = split(pbc.p, pbc.data$marks)$case
controls = split(pbc.p, pbc.data$marks)$control

par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    2.5, 0))
```

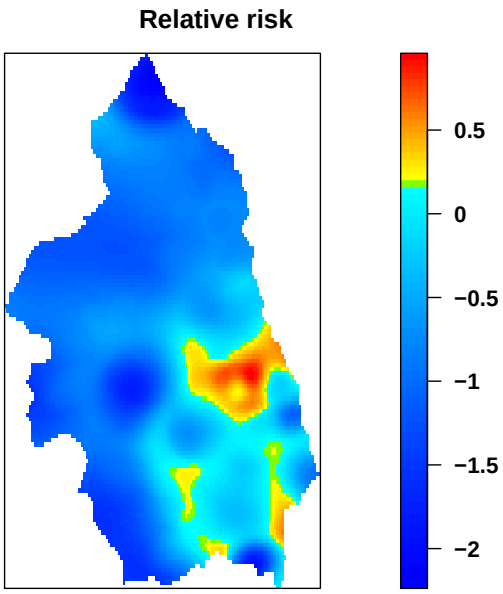
```
plot(pbc.p$window, main = "Primary Biliary Cirrhosis data")
points(controls, pch = 3, col = "green")
points(cases, pch = 16, col = "red")
legend("right", c("control", "case"), pch = c(3, 16), col = c("green", "red"), bty = "n",
    cex = 1.2)
```



2) Assess the spatial variation of the risk

We plot the relative risk in logarithmic scale, so the locations colored in red have a Relative risk greater than 1, while locations colored in blue have a Relative risk smaller than 1.

```
chp <- risk(cases, controls, adapt = T)

M = max(chp$rr)
m = min(chp$rr)

par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 0))
plot(chp$rr, gamma = 1.3, main = "Relative risk", col = beachcolours(c(m, M)))
```



Let's use a permutation test to see if there is a significant difference between cases and controls risk, computing on a regular grid the statistic:

$$\hat{T} = |c| \sum_{i=1}^{p} (\hat{\rho}(s_i) - \hat{\rho}_0)^2$$

where c is the cell of the grid and (under the null hypothesis H_0) $\rho_0 = 0$.

```
cellsize <- chp$rr$xstep * chp$rr$xstep
rho0 <- 0
```

The value of the statistic is:

```
ratiorho <- cellsize * sum((chp$rr$v - rho0)^2, na.rm = T)
ratiorho
```

```
## [1] 35687966
```

```
# Permutation function
perm_rr <- function() {
    new_ch <- rlabel(pbc.p)

    num = length(pbc.p$x)
    indices = sample(1:num)

    new_cases <- split(new_ch, f = pbc.p$marks[indices])$case
    new_controls <- split(new_ch, f = pbc.p$marks[indices])$control
    new_chp <- risk(new_cases, new_controls)
    cellsize <- new_chp$rr$xstep * new_chp$rr$ystep
    ratio_perm <- cellsize * sum((new_chp$rr$v - rho0)^2, na.rm = T)
    ratio_perm
}
```

And its p-value (obtained through a permutation approach) is:

```
(sum(rperm > ratiorho) + 1)/(nsim + 1)
```

```
## [1] 1
```

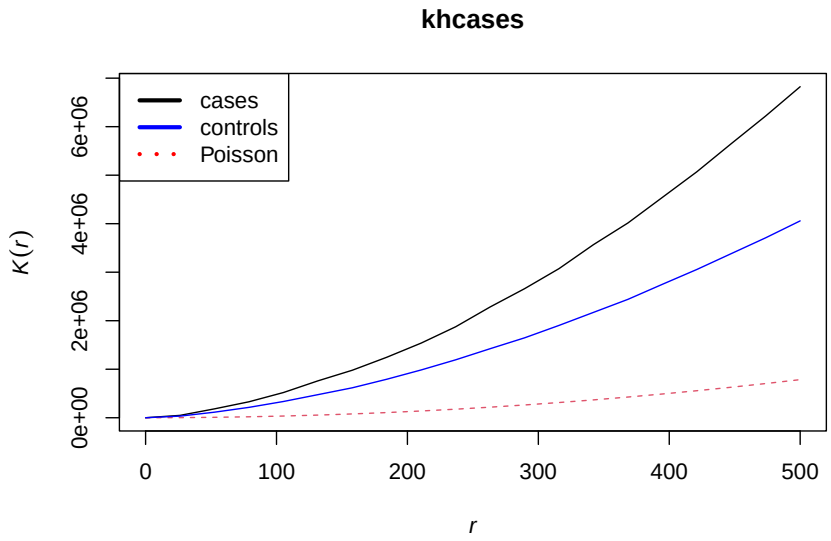
So we don't have enough evidence to reject the null hypothesis.

3) Compare the interaction patterns.

We draw the K-functions for cases and controls:

```
s = seq(0, 500, length = 20)
khcases <- Kest(cases, r = s, correction = "best")
khcontrols <- Kest(controls, r = s, correction = "best")
```

```
plot(khcases, legend = F)
lines(khcontrols$r, khcontrols$iso, lty = 1, col = "blue")
legend("topleft", legend = c("cases", "controls", "Poisson"), lty = c(1, 1, 3), col = c("black",
    "blue", "red"), lwd = 3)
```



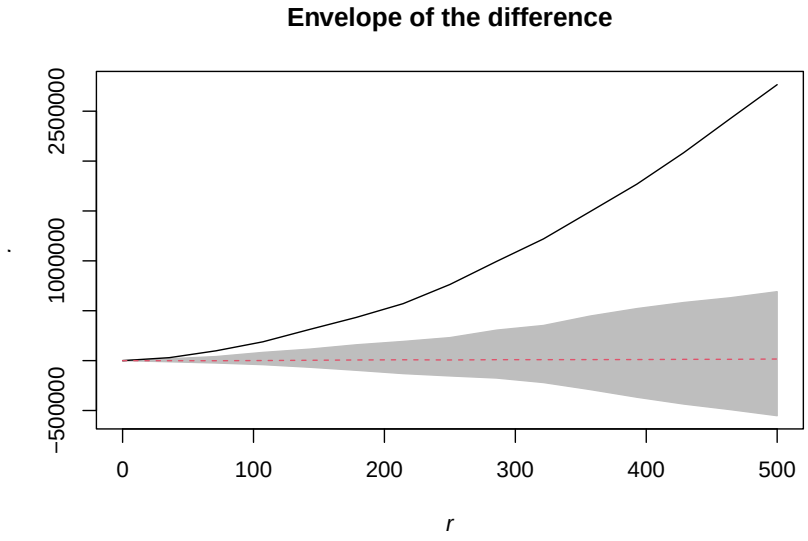
The K-functions for both cases and controls are above the Poisson curve, indicating a cluster interaction. To test the null hypothesis that the two K-functions are equal, we will use a permutational approach.

```
Kdif <- function(X, r, cr = "iso") {
    k1 <- Kest(X[marks(X) == "case"], r = r, correction = cr)
    k2 <- Kest(X[marks(X) == "control"], r = r, correction = cr)
    D = k1[[cr]] - k2[[cr]]
    res <- data.frame(r = r, D = D)
    return(fv(res, valu = "D", fname = "D"))
}
```

```
nsim <- 39
envKdif <- envelope(pbc.p, Kdif, r = s, nsim = nsim, savefuns = TRUE, simulate = expression(rlabel(pbc.p
```

```
## Generating 39 simulations by evaluating expression  ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
##
## Done.
```

```
plot(envKdif, legend = F, main = "Envelope of the difference")
```



As shown in the plot, the difference between the K-functions falls outside the envelopes for all ranges, indicating that the difference is not significant. The p-value for this test is as follows:

```
simfuns <- as.data.frame(attr(envKdif, "simfuns"))[, -1]
khcovdiag <- apply(simfuns, 1, var)
T0 <- sum(((khcases$iso - khcontrols$iso)/sqrt(khcovdiag))[-1])
T_pm <- apply(simfuns, 2, function(X) {
    sum((X/sqrt(khcovdiag))[-1])
})
pvalue <- 2 * (sum(abs(T_pm) > abs(T0)) + 1)/(nsim + 1)
pvalue
```

```
## [1] 0.05
```

The p-value is less than 5, so we have sufficient evidence to reject the null hypothesis and conclude that the interaction pattern differs between cases and controls.