Assignment 4

Spatial Point Patterns & Kriging

Juan Carlos Villaseñor-Derbez

2018-03-14

```r
suppressPackageStartupMessages({
  library(sf)
  library(spdep)
  library(spatstat)
  library(cowplot)
  library(maptools)
  library(splancs)
  library(magrittr)
  library(tidyverse)
})
```

Part 1

Define a couple of fucntions I will use

```r
# Extract the geometry feature in a sfc_POINT object and add them as columns to the data.frame
sfc_as_cols <- function(x, names = c("x","y")) {
  stopifnot(inherits(x,"sf") && inherits(sf::st_geometry(x),"sfc_POINT"))
  ret <- sf::st_coordinates(x)
  ret <- tibble::as_tibble(ret)
  stopifnot(length(names) == ncol(ret))
  x <- x[ , !names(x) %in% names]
  ret <- setNames(ret,names)
  dplyr::bind_cols(x,ret)
}

# Converte kilometers to degree
km2d <- function(km){
  (km/1.852)/60
}

# Calculate the envelope for G or F estimations
env_wrapper <- function(x, fun){
  r <- seq(0,km2d(10),length.out=10000)

  x %>%
    st_set_geometry(NULL) %>%
    as.matrix() %>%
    ppp(x = .[,1], y = .[,2], window = owin(xrange = bbox(.)[1,], yrange = bbox(.)[2,])) %>%
    envelope(fun = fun, r = r, nsim=99, nrank=2, verbose = F) %>%
    as.data.frame()
}
```

```r
# Standard plot for the envelopes
env_plot <- function(x, ylab){
  ggplot(x, aes(x = r, y = theo)) +
    geom_ribbon(aes(ymin = lo, ymax = hi), alpha = 0.5) +
    geom_line(linetype = "dashed", color = "red") +
    geom_line(aes(y = obs), color = "black") +
    labs(y = ylab)
}
```

Load CA and h2kb,rds

```r
CA <- us_states %>%
  filter(NAME == "California")

proj <- st_crs(CA)

data <- readRDS("h2kb.rds") %>%
  st_as_sf(coords = c("XCORD", "YCORD"), crs = proj) %>%
  sfc_as_cols() %>%
  select(x, y)
```

Sample 100 and 500 datapoints

```r
set.seed(43)
SAMPLE1 <- data[sample(100, replace=F),]

SAMPLE2 <- data[sample(500, replace=F),]
```

Visualize the sampled plots

```r
s1p <- ggplot() +
  geom_sf(data = CA) +
  geom_sf(data = SAMPLE1, color = "blue")

s2p <- ggplot() +
  geom_sf(data = CA) +
  geom_sf(data = SAMPLE2, color = "red")

plot_grid(s1p, s2p, labels = "AUTO", ncol = 1)
```

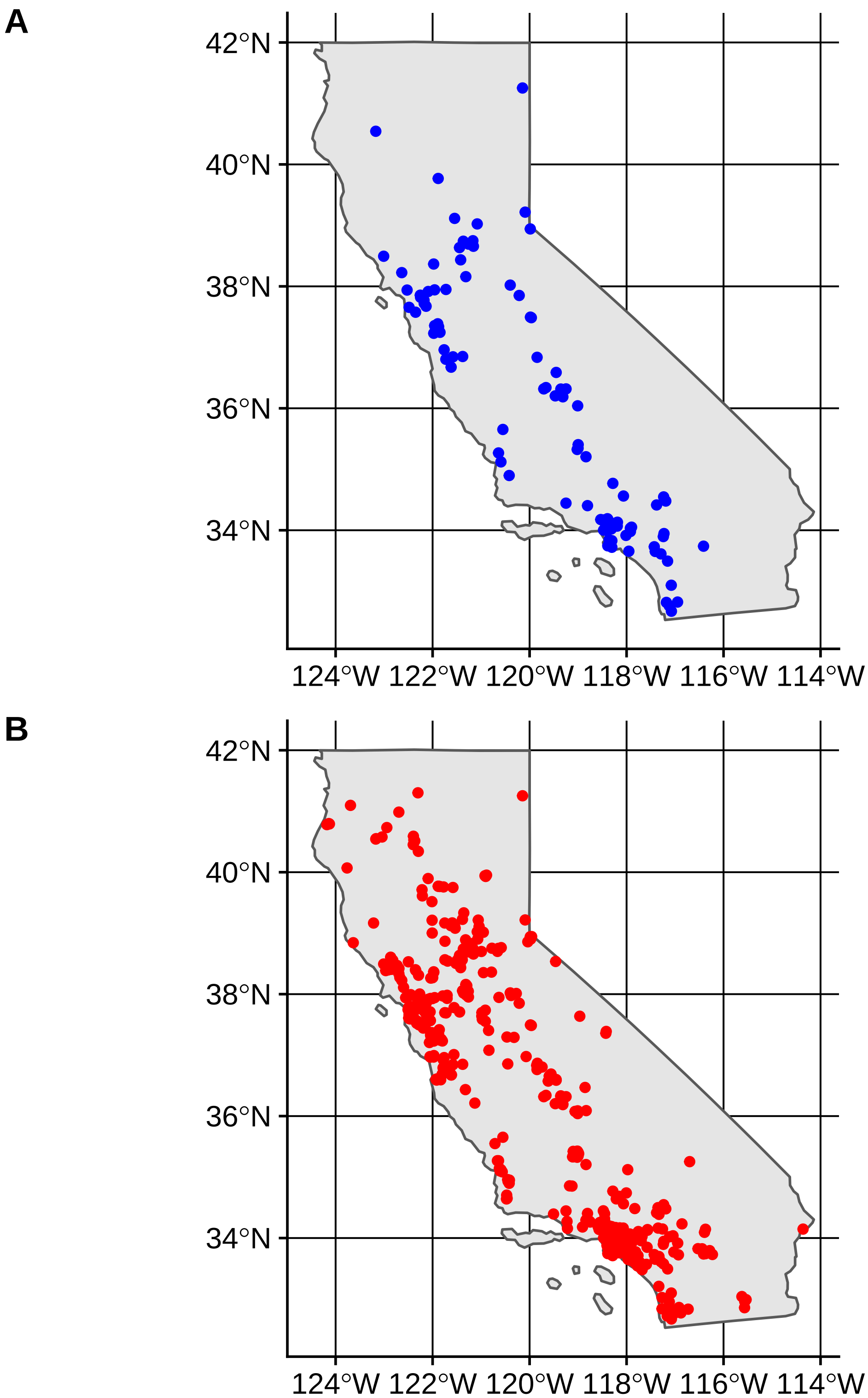


Figure 1: Random sample points obtained in SAMPLE1 (A) and SAMPLE2 (B).

Calculate G and F functions for each sample and plot them

Figure 2 shows the G anf F functions for the SAMPLE1 and SAMPLE2 subsamples. At a low sample size the G function (Fig. 2A) appears stepped and not as concave as the case with a larger sample size (Fig. 2B). In both cases, however, the deviation from the theoretical simulations is clear. For the F function, however, the deviation from the theoretical pathway is not as evident for the small sample size (Fig. 2C) as for the large sample size (Fig. 2D).

```
GS1 <- SAMPLE1 %>%
  env_wrapper(Gest) %>%
  env_plot(ylab = "G(r)")

GS2 <- SAMPLE2 %>%
  env_wrapper(Gest) %>%
  env_plot(ylab = "G(r)")

FS1 <- SAMPLE1 %>%
  env_wrapper(Fest) %>%
  env_plot(ylab = "F(r)")

FS2 <- SAMPLE2 %>%
  env_wrapper(Fest) %>%
  env_plot(ylab = "F(r)")

plot_grid(GS1, GS2, FS1, FS2, labels = "AUTO")
```

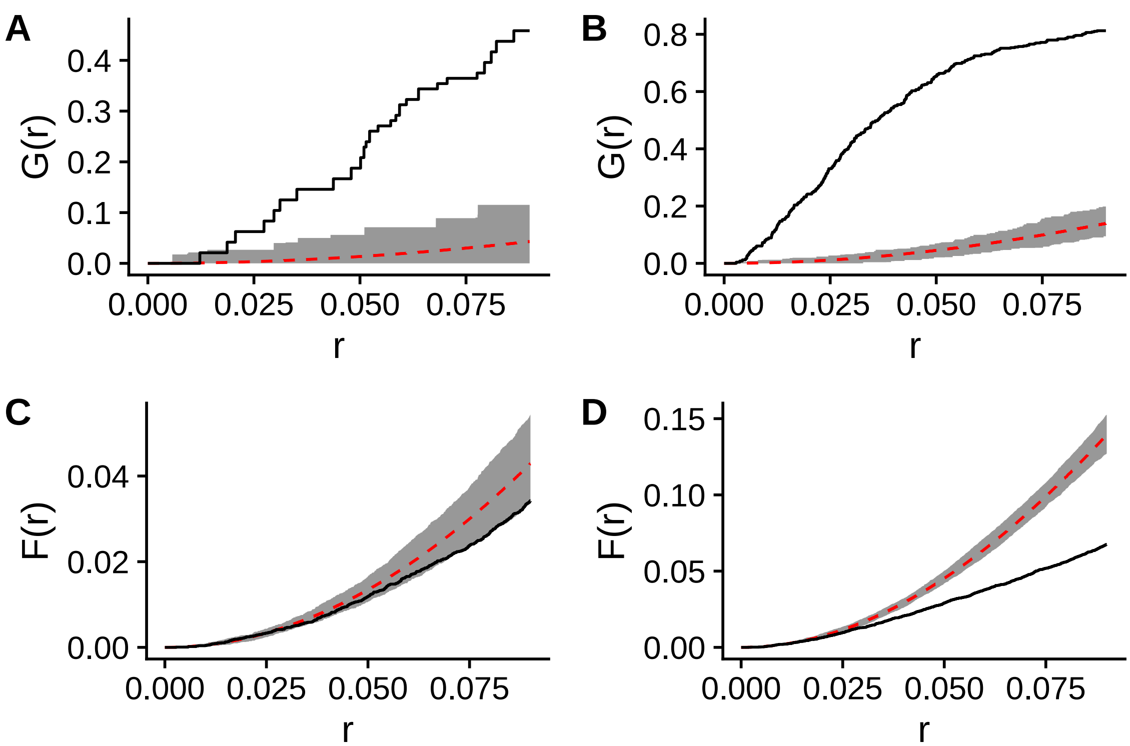


Figure 2: G and F functions for SAMPLE1 (A, C) and SAMPLE2 (B, D).

Part 2

Define functions

```r
# Get points of the bounding box and create a points object
get_poly <- function(x){
  sp_point <- x%>%
    st_set_geometry(NULL) %>%
    as.matrix()

  as.points(c(min(sp_point[,1]),
              max(sp_point[,1]),
              max(sp_point[,1]),
              min(sp_point[,1])),
            c(max(sp_point[,2]),
              max(sp_point[,2]),
              min(sp_point[,2]),
              min(sp_point[,2])))
}

# Get the binwith for which MSE is minimized
get_bw <- function(x){

  poly <- get_poly(x)

  x %>%
    st_set_geometry(NULL) %>%
    as.matrix() %>%
    mse2d(poly=poly, nsmse=100, range=0.1) %>%
    as.data.frame()
}

# Get a GridTopology object for each sample
get_grd <- function(x){
  grd <- x%>%
    as("Spatial") %>%
    Sobj_SpatialGrid(maxDim=100) %$%
    SG %>%
    summary() %$%
    grid

  GridTopology(cellcentre.offset = grd$cellcentre.offset,
                    cellsize = grd$cellsize,
                    cells.dim = grd$cells.dim)
}
```

Get the optimal binwidth for each sample

```r
S1bw <- get_bw(SAMPLE1) %>%
  filter(mse == min(mse))
```

```r
S2bw <- get_bw(SAMPLE2) %>%
  filter(mse == min(mse))
```

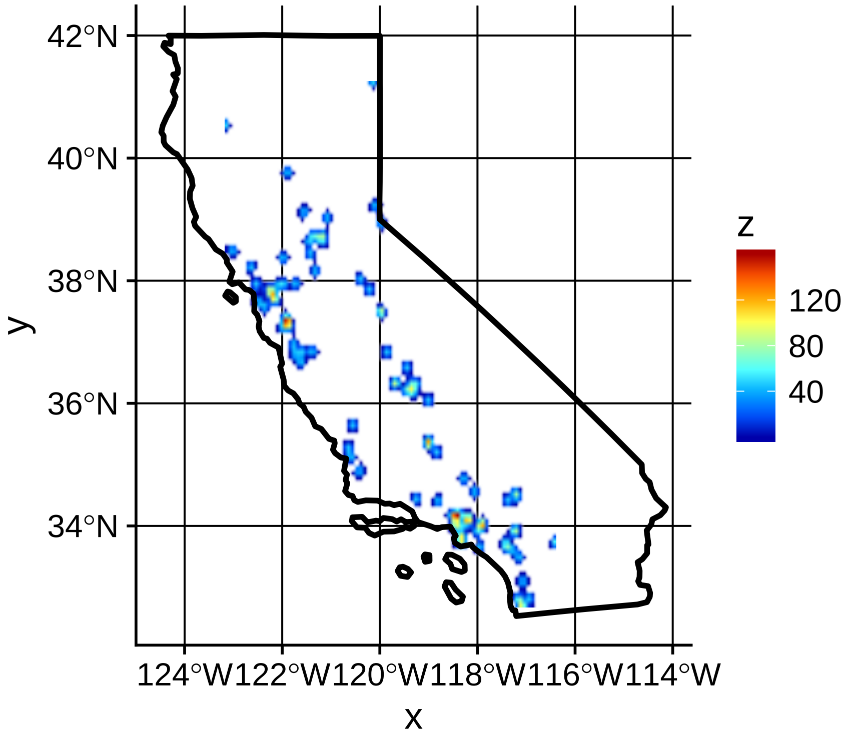
Table 1: Optimal binwidths and associated MSEs

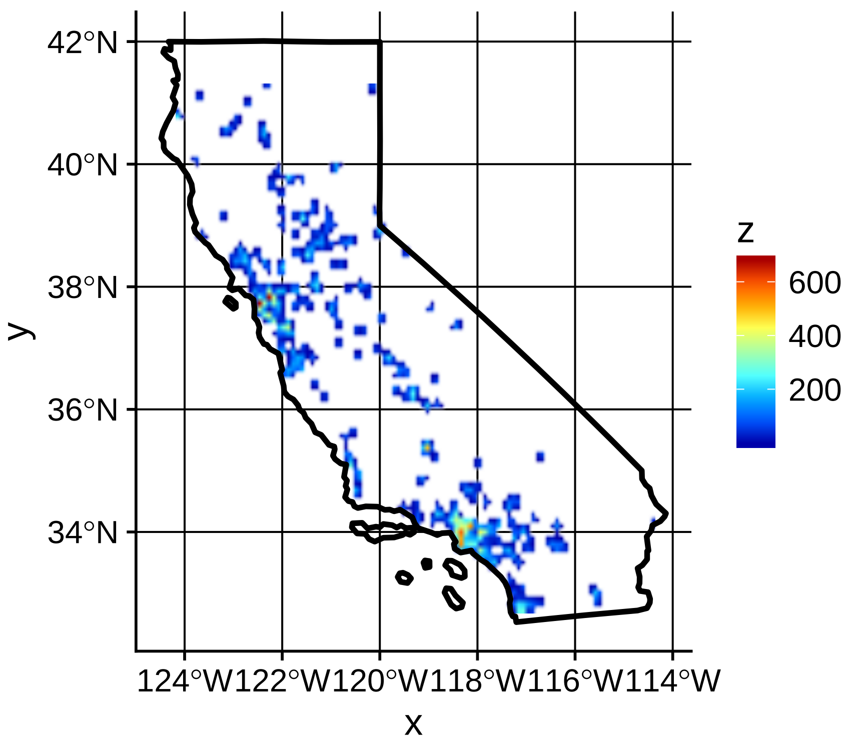
h	mse
0.100	0.1107084
0.077	-21.1328938

Each of the spatial kernels use a different optimal binwith size, shon in Table 1. Due to the random sampling performed on the original data, SAMPLE1 contains a smaller portion that clearly does not represent the data. For example, tha largest observed density was 160. When repeting the exercise with SAMPLE2, we observe that the maximum value reaches 680. Howver, the patterns remain the same; Two regions of high clustering (San Francisco and Los Angeles) are identified.

```r
S1grd <- get_grd(SAMPLE1)

k1 <- SAMPLE1 %>%
  st_set_geometry(NULL) %>%
  as.matrix() %>%
  spkernel2d(poly = get_poly(SAMPLE1), h0 = S1bw$h, grd = S1grd) %>%
  as.data.frame() %>%
  magrittr::set_colnames(value = "z") %>%
  SpatialGridDataFrame(grid = S1grd, data = .) %>%
  as("SpatialPointsDataFrame") %>%
  st_as_sf() %>%
  sfc_as_cols() %>%
  st_set_geometry(NULL) %>%
  filter(z > 0) %>%
  ggplot() +
  geom_raster(aes(x = x, y = y, fill = z), interpolate = T) +
  scale_fill_gradientn(colors = colorRamps::matlab.like(50)) +
  geom_sf(data = CA, fill = "transparent", color = "black", size = 1)

S2grd <- get_grd(SAMPLE2)

k2 <- SAMPLE2 %>%
  st_set_geometry(NULL) %>%
  as.matrix() %>%
  spkernel2d(poly = get_poly(SAMPLE2), h0 = S2bw$h, grd = S2grd) %>%
  as.data.frame() %>%
  magrittr::set_colnames(value = "z") %>%
  SpatialGridDataFrame(grid = S2grd, data = .) %>%
  as("SpatialPointsDataFrame") %>%
  st_as_sf() %>%
  sfc_as_cols() %>%
  st_set_geometry(NULL) %>%
  filter(z > 0) %>%
  ggplot() +
  geom_raster(aes(x = x, y = y, fill = z), interpolate = T) +
  scale_fill_gradientn(colors = colorRamps::matlab.like(50)) +
  geom_sf(data = CA, fill = "transparent", color = "black", size = 1)
```

```
plot_grid(k1, k2, ncol = 1)
```





Part 3

Part 3: Compare the Kriging examples in Isaacs and Srivastava and the class notes of the week of March 5 – describe differences.

Part 4

Then, compare them to the examples in "Phaedon's Examples of variograms" document on Gaucho space. What type of spatial distribution and continuity does the Isaacs&Srivastava represent and how is that different from the data analyzed in the Lecture Notes of March 6?