Random fields on a lattice

We look at the dataset of 49 neighbourhoods in Columbus, Ohio, from the year 1980. The data include various information for each neighbourhood including a measure of criminality (residential burglaries and vehicle thefts per thousand households in the neighborhood). The points correspond to the centroids of the respective neighbourhoods. The line segments join the centroids of neighbourhoods which share a common piece of boundary. This relation of being neighbours iff a common piece of boundary exists is symmetric.

```
> library(spdep)
> data(columbus)
> ?columbus
```
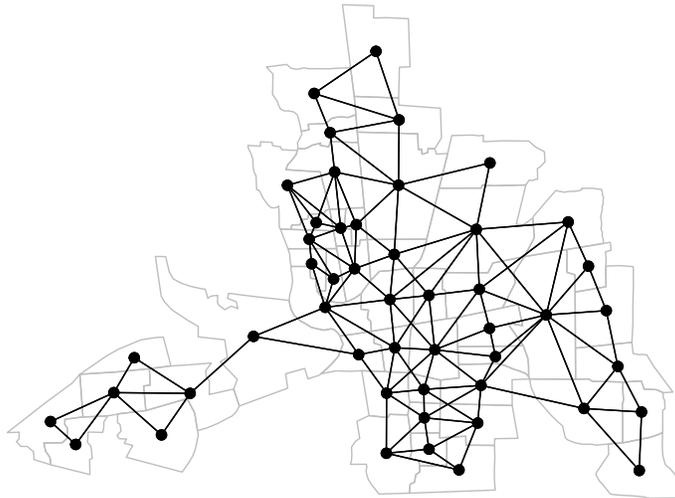

The following object contains the neighbourhood information.

```
> col.gal.nb
```

```
Neighbour list object:
Number of regions: 49
Number of nonzero links: 230
Percentage nonzero weights: 9.579342
Average number of links: 4.693878
```

Now we attach the binary weights to each link and also the normalized binary weights (which sum up to 1 for each point).

```
> col.b <- nb2listw(col.gal.nb,style="B")    # binary weights
> col.w <- nb2listw(col.gal.nb,style="W")    # normalized binary weights
```

Continuous data

We look now at the measure of criminality. It contains both high values and low values. It is of interest to find out if the values in neighbouring sites are correlated.

```
> columbus$CRIME

 [1] 15.725980 18.801754 30.626781 32.387760 50.731510 26.066658  0.178269
 [8] 38.425858 30.515917 34.000835 62.275448 56.705669 46.716129 57.066132
[15] 48.585487 54.838711 36.868774 43.962486 54.521965  0.223797 40.074074
[22] 33.705048 20.048504 38.297871 61.299175 40.969742 52.794430 56.919785
[29] 60.750446 68.892044 17.677214 19.145592 41.968163 23.974028 39.175053
[36] 14.305556 42.445076 53.710938 19.100863 16.241299 18.905146 16.491890
[43] 36.663612 25.962263 29.028488 16.530533 27.822861 26.645266 22.541491

> summary(columbus$CRIME)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1783 20.0500 34.0000 35.1300 48.5900 68.8900
```

Next we perform the tests of "no spatial autocorrelation" hypothesis. The tests are based on the Moran index (Geary index), the binary weigths (normalized binary weigths) and the randomization assumption (normality assumption). Here we report only a few of the 8 possible combinations. Note that the different tests may yield quite different results. We assume a constant mean value and constant variance of the values in individual sites.

```
> # Moran index for normalized weights,
> # test based on the randomization assumption
> # (all permutations of assigning the observed values
> # to the lattice points are equally probable)
> moran.test(columbus$CRIME,listw=col.w)

        Moran I test under randomisation

data:  columbus$CRIME
weights: col.w

Moran I statistic standard deviate = 5.3427, p-value = 4.578e-08
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation         Variance
     0.485770914      -0.020833333      0.008991121

> # Moran index for normalized weights,
> # test based on the normality assumption
> # (uncorrelated random variables with Gaussian distribution)
> moran.test(columbus$CRIME,listw=col.w,randomisation=FALSE)
```

```
        Moran I test under normality

data:  columbus$CRIME
weights: col.w

Moran I statistic standard deviate = 5.3818, p-value = 3.687e-08
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
     0.485770914       -0.020833333        0.008860962

> # Geary index for normalized weights,
> # test based on the randomization assumption
> geary.test(columbus$CRIME,listw=col.w)

        Geary C test under randomisation

data:  columbus$CRIME
weights: col.w

Geary C statistic standard deviate = 4.5669, p-value = 2.475e-06
alternative hypothesis: Expectation greater than statistic
sample estimates:
Geary C statistic       Expectation          Variance
     0.547803377        1.000000000        0.009804108
```

All the tests resulted in rejection of the null hypothesis.

Binary data

We now transform the numeric crime values into binary variables indicating if the criminality is low (below 35, blue points) or high (above 35, red points) in a given neighbourhood, see the figure below. We perform a test of "no spatial autocorrelation" using the BB statistic, $BB = \frac{1}{2} \sum_{i,j} w_{ij} Z_i Z_j$, where we assume that the two possible values are 0 and 1. We use two versions of the test, using either binary or normalized weights. Both tests are performed under the assumption of nonfree sampling (hypergeometric sampling), i.e. the number of red and blue regions is fixed.

```
> HICRIME <- cut(columbus$CRIME, breaks=c(0,35,80), labels=c("low","high"))
> joincount.test(HICRIME, listw=col.w)

        Join count test under nonfree sampling

data:  HICRIME
weights: col.w
```
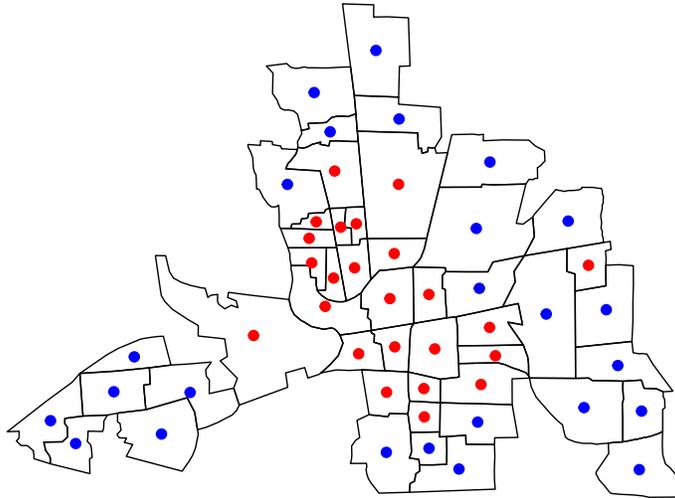
```
Std. deviate for low = 4.6176, p-value = 1.941e-06
alternative hypothesis: greater
sample estimates:
Same colour statistic           Expectation            Variance
        9.4833333             6.2500000           0.4903158



        Join count test under nonfree sampling

data:  HICRIME
weights: col.w

Std. deviate for high = 4.9963, p-value = 2.921e-07
alternative hypothesis: greater
sample estimates:
Same colour statistic           Expectation            Variance
         9.206349              5.750000            0.478553

> joincount.test(HICRIME, listw=col.b)

        Join count test under nonfree sampling

data:  HICRIME
weights: col.b

Std. deviate for low = 1.0802, p-value = 0.14
```

```
alternative hypothesis: greater
sample estimates:
Same colour statistic            Expectation              Variance
          34.00000                 29.33673              18.63846


          Join count test under nonfree sampling

data:  HICRIME
weights: col.b

Std. deviate for high = 5.9534, p-value = 1.313e-09
alternative hypothesis: greater
sample estimates:
Same colour statistic            Expectation              Variance
           52.0000                  26.9898               17.6484
```

Again, both versions of the test result in rejection of the null hypothesis.