Assignment Block B

Bianca Neubert and Franziska Wehrmann

Theory

Markov random fields

Let Y denote a GMRF with mean vector μ and precision matrix Q of dimension $n \in \mathbb{N}$.

i) To show:

$$Y_i | Y_{-i} \sim \mathcal{N}\left(\mu_i - \frac{1}{q_{ii}} \sum_{j \in \mathrm{ne}(i)} q_{ij}(y_j - \mu_j), \frac{1}{q_{ii}}\right)$$

Proof: Since Y is a GMRF with expectation μ and precision matrix Q, we have for the conditional density of $Y_i | Y_{-i}$

$$\pi(y_i | y_{-i}) = \pi(y_i | y_{\mathrm{ne}(i)})$$

where $y_{\mathrm{ne}(i)}$ denotes the vector with indices in $\mathrm{ne}(i)$. Because the MRF is Gaussian, we have further for every $A \subset \{1, ..., n\}$

$$Y_A \sim \mathcal{N}(\mu_A, Q_A^{-1})$$

where $Y_A = (Y_i)_{i \in A}$, $\mu_A = (\mu_i)_{i \in A}$ and $Q_A = (q_{ij})_{i,j \in A}$ for the entries q_{ij} of Q.

Combining this, we get for the density of $Y_i | Y_{-i}$, since $Y_{\mathrm{ne}(i)}$ is also normally distributed and hence has strictly positive density, that

$$\begin{aligned}
\pi(y_i | y_{-i}) &= \frac{\pi(y_{\{i\} \cup \mathrm{ne}(i)})}{\pi(y_{\mathrm{ne}(i)})} \\
&\propto \exp\left\{ -\frac{1}{2}(y_{\{i\} \cup \mathrm{ne}(i)} - \mu_{\{i\} \cup \mathrm{ne}(i)})^T Q_{\{i\} \cup \mathrm{ne}(i)} (y_{\{i\} \cup \mathrm{ne}(i)} - \mu_{\{i\} \cup \mathrm{ne}(i)}) \right. \\
&\quad \left. + \frac{1}{2}(y_{\mathrm{ne}(i)} - \mu_{\mathrm{ne}(i)})^T Q_{\mathrm{ne}(i)} (y_{\mathrm{ne}(i)} - \mu_{\mathrm{ne}(i)}) \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[q_{ii} y_{ii}^2 + 2 \sum_{j \in \mathrm{ne}(i)} y_i q_{ij} y_j + q_{ii} \mu_i^2 + 2 \sum_{j \in \mathrm{ne}(i)} \mu_i q_{ij} \mu_j \right. \right. \\
&\quad \left. \left. -2 q_{ii} \mu_i y_i - 2 \sum_{j \in \mathrm{ne}(i)} \mu_i q_{ij} y_j - 2 \sum_{j \in \mathrm{ne}(i)} y_i q_{ij} \mu_j \right] \right\} \\
&= \exp\left\{ -\frac{q_{ii}}{2} \left[(y_i - \mu_i)^2 + \frac{2}{q_{ii}} (y_i - \mu_i) \sum_{j \in \mathrm{ne}(i)} q_{ij}(y_j - \mu_j) \right] \right\} \\
&\propto \exp\left\{ -\frac{q_{ii}}{2} \left[y_i - \mu_i + \frac{1}{q_{ii}} \sum_{j \in \mathrm{ne}(i)} q_{ij}(y_j - \mu_j) \right]^2 \right\}
\end{aligned}$$

which is the kernel of a normal distribution with the given expecation and variance.

iii) To show: $\mathbb{C}\text{or}[Y_i, Y_j | Y_{-ij}] = -\frac{q_{ij}}{\sqrt{q_{ii}q_{jj}}}$

Proof:

Local charcteristics

Suppose S is a finite set equipped with a symmetry relation $\sim$. For count outcomes, a specification commonly used in disease mapping is the *auto-Poisson* model where

$$\pi(y_i | y_{S \setminus i}) = \exp(-\mu_i) \frac{\mu_i^{y_i}}{y_i}$$
$$\log(\mu_i) = - \sum_{j \in \text{ne}(i)} y_j \ i \neq j$$

for $y \in \mathbb{N}_0$, $i \in S$. Determine the canonical potential with respect to $a = 0$.

Computation

Guerry's data on social morals

This task is based on Guerry's data on moral statistics including aggregated numbers of suicide (`suicids`). Let us first import and inspect the data.

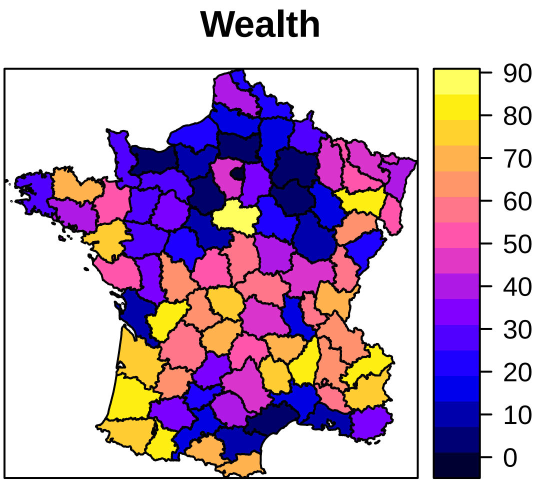
```
guerry <- readOGR("guerry/Guerry.shp")
```

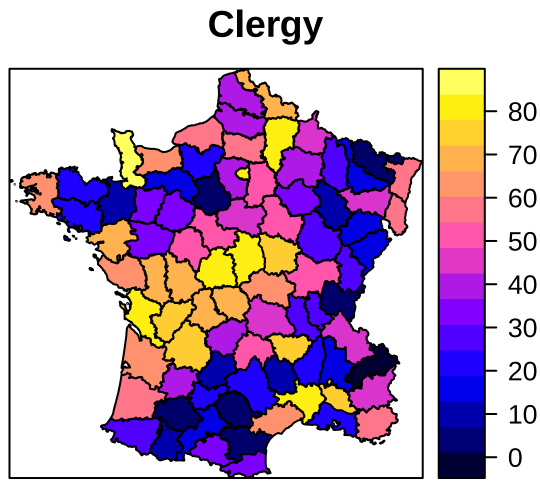
```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Bianca\Documents\Uni\Semester_10\Spatial_Statistics\Assignment_A\SpatStat\guerry\Gu
## with 85 features
## It has 26 fields
## Integer64 fields read as strings:  dept Crm_prs Crm_prp Litercy Donatns Infants Suicids MainCty Weal
```

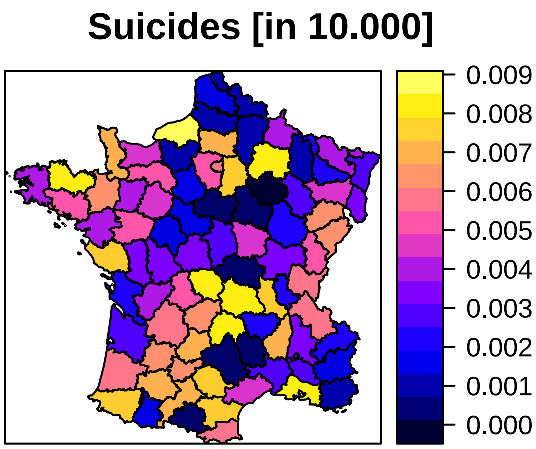
We see that the variables are stored in form of strings, so we first change the class to integer of the three variables of interest in this assignment: `Suicids`, `Wealth` and `Clergy`

```
guerry$Suicids <- as.integer(guerry$Suicids)/10000
guerry$Wealth <- as.integer(guerry$Wealth)
guerry$Clergy <- as.integer(guerry$Clergy)
```

Using `spplot` we can get a first impression on the spatial distribution of these variables.







Specify a (non-spatial) regression model of `Wealth` and `Clergy` on `Suicids`.

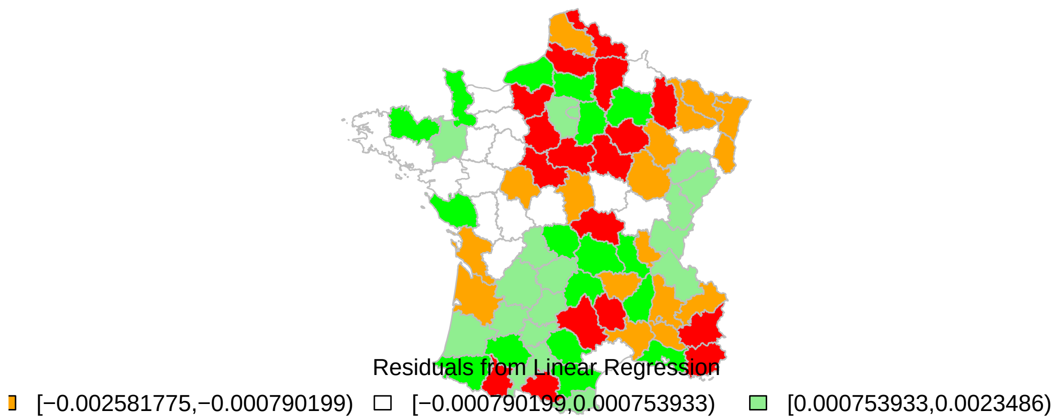
```
reg <- lm(Suicids ~ Wealth + Clergy, data=guerry)
summary(reg)
```

```
##
## Call:
## lm(formula = Suicids ~ Wealth + Clergy, data = guerry)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0043642 -0.0024202  0.0001414  0.0019298  0.0043123
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.848e-03  7.267e-04   6.671 2.77e-09 ***
## Wealth      -2.245e-06  1.098e-05  -0.204    0.839
## Clergy      -1.067e-05  1.114e-05  -0.958    0.341
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002484 on 82 degrees of freedom
## Multiple R-squared:  0.01147,    Adjusted R-squared:  -0.01264
## F-statistic: 0.4757 on 2 and 82 DF,  p-value: 0.6232
```

We can see, that in this the coefficients of Clergy and Wealth are not significant which at first does not indicate a relationship between these and Suicides. This model does not explain the variation in Suicides very well (e.g. Multiple R-squared = 0.011).

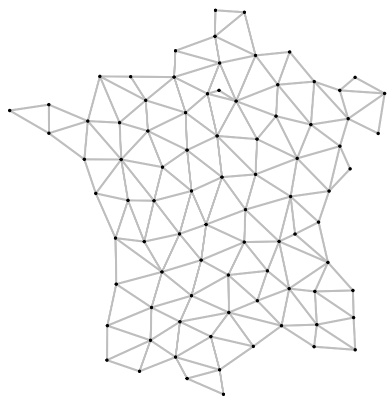
We can plot the residuals:



TODO: Moron manual ?? Meinst du Moran's I :)

To test for autocorrelation in the data, we can use Moran's I and Geary's C. For this, we need to specify a neighborhood structure in a matrix.

```
xy <- coordinates(guerry)
W_cont_el <- poly2nb(guerry, queen=FALSE)
W_cont_el_mat <- nb2listw(W_cont_el, style="B", zero.policy=TRUE)
plot(W_cont_el_mat, coords=xy, cex=0.1, col="gray")
```



For the Moran's I ad Geary's C we get the following results

```
##
##  Moran I test under randomisation
##
## data:  guerry$Suicids
## weights: W_cont_el_mat
##
## Moran I statistic standard deviate = 1.2871, p-value = 0.09903
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##        0.07436323       -0.01190476        0.00449226
```

```
##
##  Geary C test under randomisation
##
## data:  guerry$Suicids
## weights: W_cont_el_mat
##
## Geary C statistic standard deviate = 0.85148, p-value = 0.1973
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic       Expectation          Variance
##       0.938002736       1.000000000       0.005301488
```

The Moran's I test statistic and its expectation are both close to zero (0.074 and -0.011). However, the Moran's I test statistic is larger than the expectation which is an indication for a positive autocorrelation. If we look the given p-value we can reject the null of no autocorrelation for a 0.1-niveau, but not for a 0.05-niveau. On the other hand, Geary's C is close to 1 and the p-value with 0.197 suggests that the null of no autocorrelation can not be rejeceted for a 0.1-niveau. To sum this up, there seems to be very little global autocorrelation in the data, if there is any at all.

TODO: whats the difference? to moran below? , Interpretation Warum moran with residuals? mc nicht?

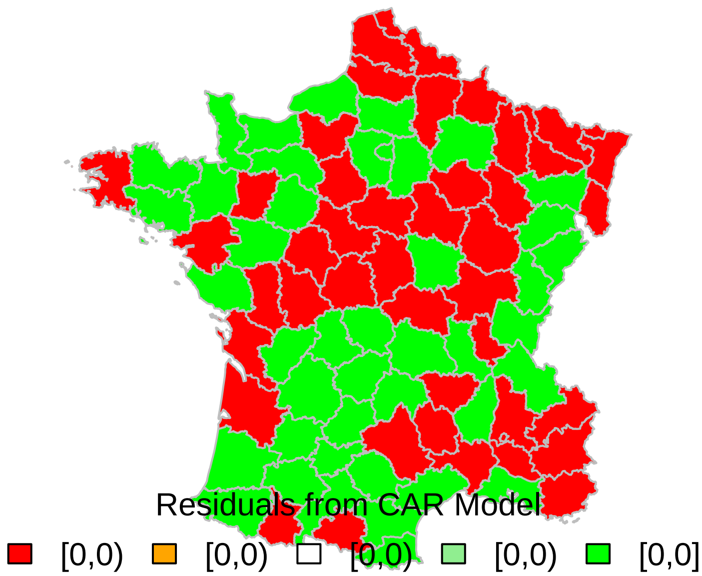
CAR

Let us compute a CAR model.

```
car.out <- spdep::spautolm(Suicids ~ Wealth + Clergy, data=guerry,
                           listw=W_cont_el_mat, family="CAR")
mod.car <- fitted(car.out)
summary(car.out)
```

```
##
## Call: spdep::spautolm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, family = "CAR")
##
## Residuals:
##         Min          1Q      Median          3Q         Max
## -0.00446476 -0.00212060  0.00026742  0.00153406  0.00454259
##
## Coefficients:
##               Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)  4.9400e-03  7.6599e-04  6.4491 1.125e-10
## Wealth      -5.6404e-06  1.1198e-05 -0.5037    0.6145
## Clergy      -9.9394e-06  1.1396e-05 -0.8722    0.3831
##
## Lambda: 0.066815 LR test value: 1.0627 p-value: 0.30261
## Numerical Hessian standard error of lambda: 0.059395
##
## Log likelihood: 391.2798
## ML residual variance (sigma squared): 5.805e-06, (sigma: 0.0024094)
## Number of observations: 85
## Number of parameters estimated: 5
## AIC: -772.56
```

Again, the coefficients for Wealth and Clergy are not significant. The plot of the residuals:



SAR

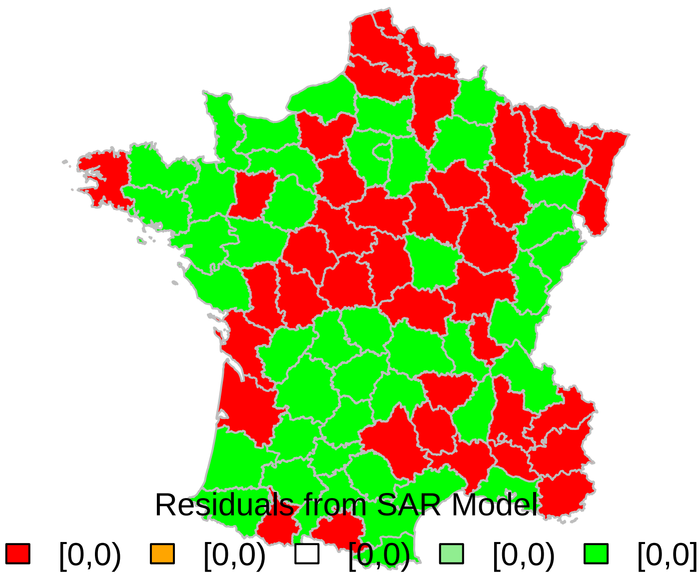
Now, the same for SAR

```
mod.sar <- lagsarlm(Suicids ~ Wealth + Clergy, data=guerry,
                    listw=W_cont_el_mat, zero.policy=T, tol.solve=1e-12)
summary(mod.sar)
```

```
##
## Call:lagsarlm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, zero.policy = T, tol.solve = 1e-12)
##
## Residuals:
##         Min          1Q      Median          3Q         Max
## -0.00456243 -0.00193496  0.00011255  0.00170948  0.00466154
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)  4.2012e-03  9.0632e-04  4.6355 3.561e-06
## Wealth      -3.0633e-06  1.0690e-05 -0.2866    0.7744
## Clergy      -1.0055e-05  1.0827e-05 -0.9288    0.3530
##
## Rho: 0.030525, LR test value: 1.3711, p-value: 0.24162
```

```
## Asymptotic standard error: 0.024869
##     z-value: 1.2274, p-value: 0.21966
## Wald statistic: 1.5066, p-value: 0.21966
##
## Log likelihood: 391.434 for lag model
## ML residual variance (sigma squared): 5.8277e-06, (sigma: 0.0024141)
## Number of observations: 85
## Number of parameters estimated: 5
## AIC: -772.87, (AIC for lm: -773.5)
## LM test for residual autocorrelation
## test value: 0.09036, p-value: 0.76372
```

Again, the coefficients for Wealth and Clergy are not significant.The plot of the residuals:



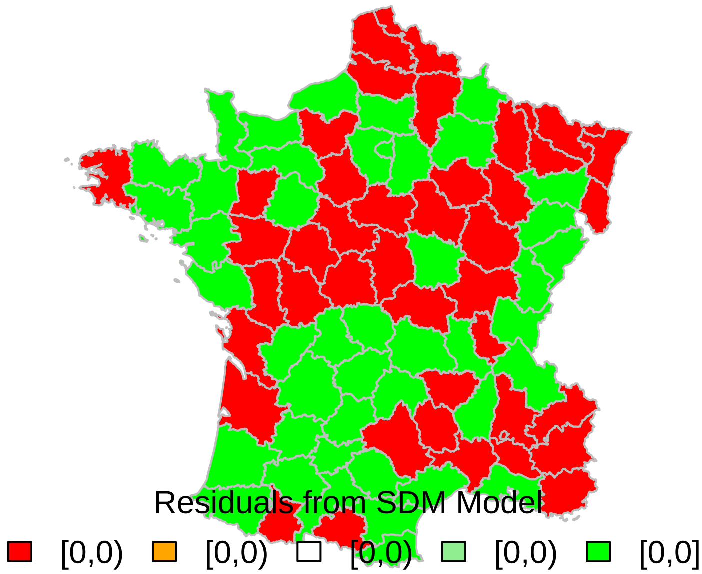
SDM

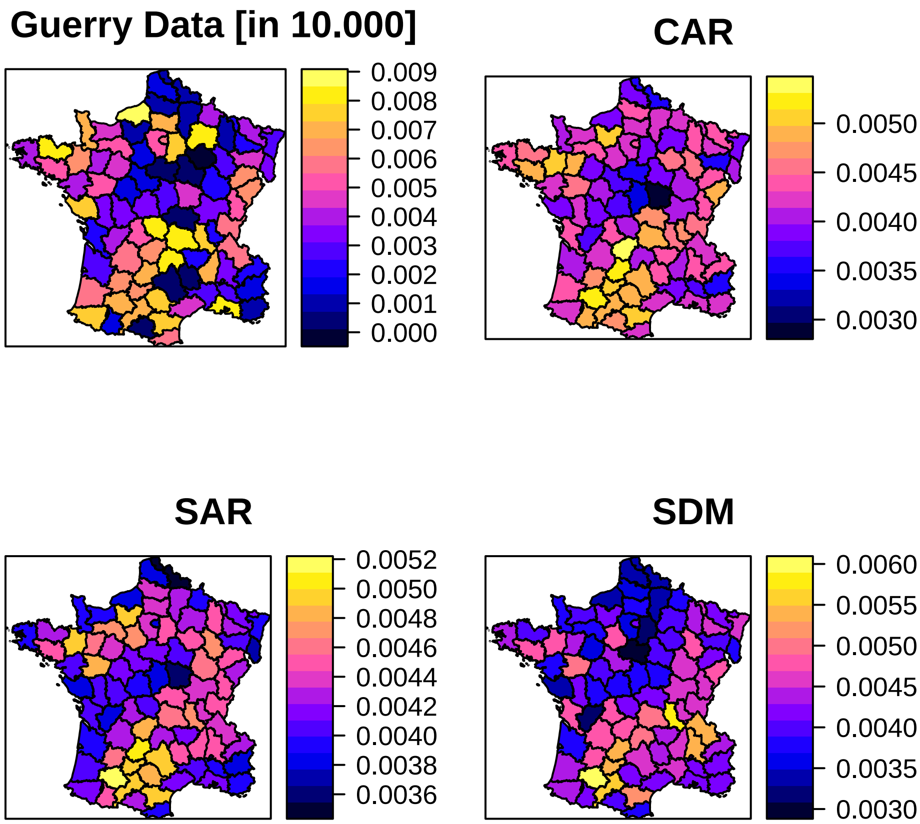
And SDM

```
mod.sdm <- lagsarlm(Suicids ~ Wealth + Clergy, data=guerry, listw=W_cont_el_mat,
                    zero.policy=T, type="mixed", tol.solve=1e-12)
summary(mod.sdm)
```

```
##
## Call:lagsarlm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, type = "mixed", zero.policy = T, tol.solve = 1e-12)
```

```
##
## Residuals:
##         Min          1Q      Median          3Q         Max
## -0.00428035 -0.00218065  0.00031401  0.00151466  0.00505583
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                   Estimate  Std. Error z value Pr(>|z|)
## (Intercept)      4.5773e-03  1.2261e-03  3.7333 0.000189
## Wealth          -1.2404e-05  1.2086e-05 -1.0263 0.304727
## Clergy          -8.2784e-06  1.2440e-05 -0.6655 0.505748
## lag.(Intercept) -1.9335e-04  3.4986e-04 -0.5526 0.580511
## lag.Wealth       7.1785e-06  4.2031e-06  1.7079 0.087651
## lag.Clergy      -2.4491e-06  4.3288e-06 -0.5658 0.571542
##
## Rho: 0.02627, LR test value: 0.59903, p-value: 0.43895
## Asymptotic standard error: 0.032206
##     z-value: 0.81569, p-value: 0.41468
## Wald statistic: 0.66535, p-value: 0.41468
##
## Log likelihood: 392.9685 for mixed model
## ML residual variance (sigma squared): 5.6282e-06, (sigma: 0.0023724)
## Number of observations: 85
## Number of parameters estimated: 8
## AIC: -769.94, (AIC for lm: -771.34)
## LM test for residual autocorrelation
## test value: 0.011826, p-value: 0.9134
```


Residuals from SDM Model
[0,0)
[0,0)
[0,0)
[0,0)
[0,0]


Guerry Data [in 10.000]
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
CAR
0.0050
0.0045
0.0040
0.0035
0.0030
SAR
0.0052
0.0050
0.0048
0.0046
0.0044
0.0042
0.0040
0.0038
0.0036
SDM
0.0060
0.0055
0.0050
0.0045
0.0040
0.0035
0.0030