Spatial Analysis of Geographic Data

Julian Bernauer

Data and Methods Unit
MZES
julian.bernauer@mzes.uni-mannheim.de

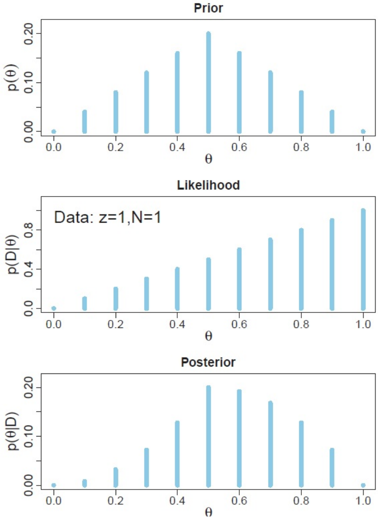
20 March 2018

Open-ended Bayesian Spatial Analysis in R

Overview

1. Bayesian analysis: Stan
2. → code
3. Interpolation
4. → code
5. Time and space
6. → code (kind of...)
7. Starting your own project: formal, ideas, problems?
8. → exercise

Bayesian estimation (Figures from Kruschke 2015)



Bayesian estimation

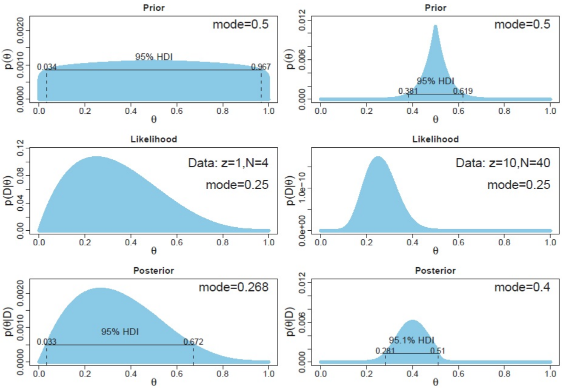
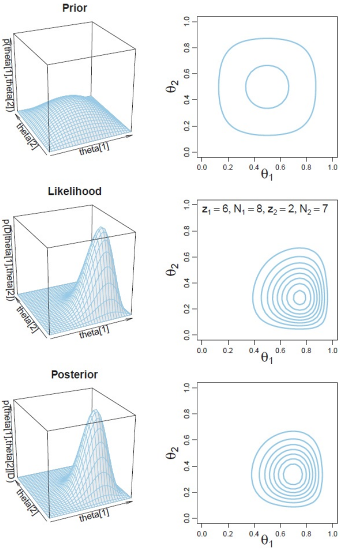


Figure 5.3: The left side is the same small sample as the left side of Figure 5.2 but with a flatter prior. The right side is the same larger sample as the right side of Figure 5.2 but with a sharper prior. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Bayesian estimation



Bayes

Theorem

Model (θ) and data (y)

$$Pr(\theta|y) = \frac{Pr(\theta, y)}{Pr(y)} \tag{1}$$

$$= \frac{Pr(\theta)Pr(y|\theta)}{Pr(y)} \tag{2}$$

- We know $Pr(y)$ as well as $Pr(y|\theta)$ and want $Pr(\theta|y)$: Probability of model given data
- Bayes works via $Pr(\theta)$: Prior for model
- Maximum likelihood avoids $Pr(\theta)$ and looks for relative likelihood of models

Bayesian estimation basics

See Lunn et al. (2013)

- Bayesian statistics and simulation: friends for life
- Monte Carlo: Calculate single values for integral to find its shape
- Markov Chain Monte Carlo (MCMC): Splits whole integral into single pieces which are easier to work with
- Values depend on previous values and converge by "trying" solutions which might have a higher probability
- Gibbs sampling: special case of Metropolis-Hastings, updates sub-vectors of quantities of interest
- Hamiltonian Monte Carlo: "knows" whether it is in the tails of a distribution or not and adjusts accordingly
- Maximum likelihood has to find a numerical solution for the log likelihood by iteration, abstains from use of priors

Hamiltonian Monte Carlo

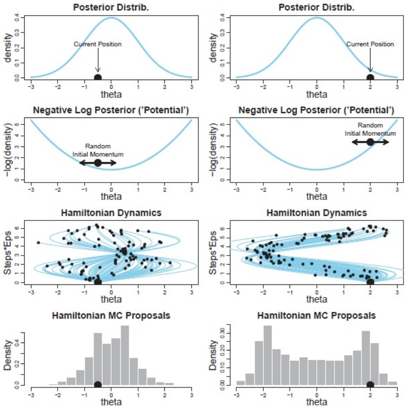


Figure 14.2: Examples of a Hamiltonian Monte Carlo proposal distributions for two different current parameter values, marked by the large dots, in the two columns. For this figure, a large range of random trajectory lengths (Steps*Eps) is sampled. Compare with Figure 14.1. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Languages

- WinBUGS → can be called from R, some spatial features
- JAGS → well implemented in R, but little on spatial analysis
- Stan → in R, Hamiltonian Monte Carlo, promises speed and flexibility, under development, spatial models feasible
- Surely you could do it in Python as well (Stan can be used there, too), use Stata or program a maximum likelihood estimator...

→ I mostly used JAGS in the last few years, great!

Some spatial models in WinBUGS and Stan

- Conditional autoregressive model (CAR) in WinBUGS (Lunn et al. 2013: 263): "spatially structured random effects distribution in a hierarchical model"
- Some other variants in GeoBUGS... see documentation
- ICAR in Stan: `http://mc-stan.org/users/documentation/case-studies/icar_stan.html`
- Explicit implementation of a spatial error model: `https://rpubs.com/chrisbrunsdon/carstan`
- Spatial lag model surely also feasible, found no example

WinBUGS: ICAR model code

```
v[1:N] ~ car.normal(nb2[], weight[], num[], vtau)
```

Stan: ICAR model code

```
data {
  int<lower=0> N;
  int<lower=0> N_edges;
  int<lower=1, upper=N> node1[N_edges];
  int<lower=1, upper=N> node2[N_edges];
}
parameters {
  vector[N] phi;
}
model {
  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  sum(phi) ~ normal(0, 0.01 * N);
}
```

Stan: getting started

- Installing: `https://github.com/stan-dev/rstan/wiki/Installing-RStan-on-Windows`
- A range of models in Stan code: `https://github.com/stan-dev/example-models/wiki/ARM-Models-Sorted-by-Type`
- We will implement the ICAR model in Stan and extend it towards the running example (attacks on refugees)

→ Code "SAGD_4_open_stan.R"

Interpolation / small area estimation

- Also for sparse data, see Selb and Munzert (2011, PA), Bernauer and Munzert (2012, Representation)
- Borrowing strength from general mean (multilevel in Bayes) and spatial neighbours
- Missing data: extreme case, predictions can be based on spatial information
- Bayes $\rightarrow$ priors provide some information

Small Area Estimation Applications

- Selb and Munzert (2011): Bayesian hierarchical and spatial (CAR in WinBUGS) modelling to obtain estimates of political preferences in small units
- Bernauer and Munzert (2012): Using geographically stabilized survey estimates of political positions in electoral districts

Selb and Munzert (2011)

- 2009 Bundestag election
- party preferences from survey data: sparse for districts and often just off the mark
- Use of spatial information provides considerable improvement
- Compare and combine their approach with the gold standard of post-stratification
- Geographic information useful especially given a lack of structural data

Selb and Munzert (2011)

$$\phi_j \mid \phi_k \sim \mathrm{N}\left(\frac{\sum_{k\neq j} w_{jk}\phi_k}{\sum_{k\neq j} w_{jk}}, \frac{\sigma_\phi^2}{\sum_{k\neq j} w_{jk}}\right), \tag{7}$$

where the w_{jk} are elements of a $J \times J$ adjacency matrix assuming a value of 1 if units j and k are neighbors, that is, have a common border or vertex, otherwise 0. Hence, the expected conditional mean of ϕ in j corresponds to the average value of ϕ in the neighborhood of j, with its variance parameter, σ_ϕ^2, controlling how similar ϕ_j is to its neighbors. Deviation υ_j, on the other hand, is assumed to vary independently and identically across districts according to a normal distribution,

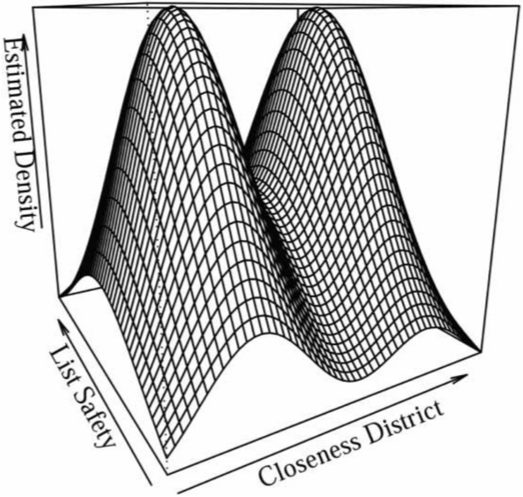
$$\upsilon_j \sim \mathrm{N}(0, \sigma_\upsilon^2). \tag{8}$$

Including both a spatially structured and an independent random component into the model will, in effect, pull the directly (but, due to small N_j, inaccurately) observed proportion of respondents holding the preference in constituency j toward both its neighborhood and the overall sample mean, with the amount of shrinkage increasing with decreasing N_j. That is, inferences for the district-level parameters, $\pi_j = \mathrm{logit}^{-1}(\alpha^0 + \phi_j + \upsilon_j)$, reflect not just the direct survey information in district j, but also draw on relevant information in the neighboring districts (which will normally host more respondents than j), as well as in all the other districts (i.e., the whole survey sample). The relative amount of local versus global smoothing is then determined by the estimated variance σ_ϕ^2 in proportion to σ_υ^2. Further, by exploiting the conditional distribution of ϕ_j, the model equally informs estimates of constituency preferences for areas not covered by the survey, provided a constituency is not an island (i.e., it has neighbors to draw information from).[6]

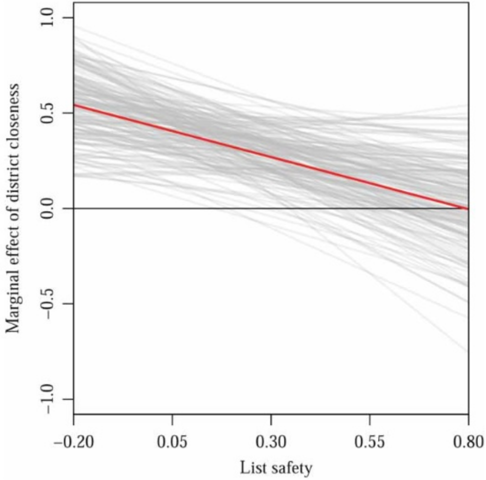
Bernauer and Munzert (2012)

- Germany mixed electoral system, 2009 BTW election
- Continuous measures of closeness of district race and list safety
- Measure of positioning of candidates between list and party
- District voter positions stabilized using spatial information
- Finding: Closeness of the district race leads to a positioning closer to the constituency rather than the party – given low levels of list safety

Bernauer and Munzert (2012)



Bernauer and Munzert (2012)



Example: sunshine hours

- Longtime sunshine hours per year in electoral districts
- Some missing data
- Implemented in WinBUGS as it handles missing data more easily

→ Code "SAGD_4_open_interpol.R"

Basics time

Analogy between space and time by autocorrelation

It's a violation - too!

- Dependency due to serial autocorrelation as t influenced by $t-1$
- Dependency due to "shocks" across j at specific t
- Dependency due to context as t's for single j are similar (regardless of order)

Space and Time

Differences (Schabenberger and Gotway 2005: 27)

- Time is directed: past, present, future
- No anisotropy (directional dependency): "spatial dependencies may develop differently in various directions"

→ Individual development of time series and spatial models
→ Spatial adaptions of time series models tend to be difficult

Integration with W

Difficulties (Schabenberger and Gotway 2005: Chapter 9)

- "[S]tatistical tools for the analysis of spatio-temporal processes are not [...] fully developed" (432)
- "This creates a dizzying array of spatio-temporal data structures" (432)

Integration with W

Schabenberger and Gotway (2005: Chapter 9)

- Recommend the joint analysis of spatio-temporal data
- Derive spatio-temporal covariance functions
- Separable for purely spatial and temporal components; non-separable for spatio-temporal interactions
  $\rightarrow$ advanced material
- Spatio-temporal process (442): earthquakes, explosions and survival
- What about Bayes?

Integration with W

Practical

- Parallels: lag model (SLM) and autocorrelation (AR); error model (SEM) an moving averages (MA)
- Can be integrated into an ARMA model (Lunn et al. 2013: 258)

Bayesian implementation AR (1) – no W (following example sunspots Lunn et al. 2013: 257ff)

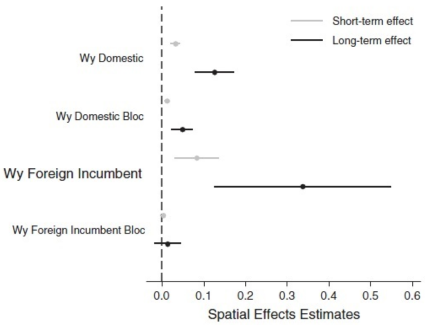
```
for (t in 1:n) {
y[t]      ~ dnorm(m[t], tau)
}
for (t in 2:n) {
m[t]      <- c + theta*y[t-1]
eps[t]    <- y[t] - m[t]
}
m[1]         <- y[1] - eps[1]
eps[1]        ~ dnorm(0, 0.0001)
theta         ~ dnorm(0, 0.0001)
c             ~ dnorm(0, 0.0001)
tau          <- 1/pow(sigma, 2)
sigma         ~ dunif(0, 100)
```

Example: Böhmelt et al. (2016, APSR)

- Type of model: spatial lag
- Temporally and **spatio-temporally lagged dependent variable**
- W: successful foreign parties, actually not based on geography
- Data: 200+ parties in 26 countries 1977-2010
- Estimation: spatial OLS and spatial maximum likelihood
- Findings: Parties respond to domestic competitors and successful foreign incumbents

Example: Böhmelt et al. (2016, APSR)

FIGURE 1. Short-Term and Asymptotic Long-Term Spatial Effects of Spatial-Lag Variables



Notes. The horizontal bars are 90 percent confidence intervals and the vertical dashed line represents a spatial effect of 0. Estimates are based on models in Table 2.

Application attempt: running example

- Outcome: daily attacks (0/1)
- Jäckle and König (2017, WEP) controls for time: multilevel states and (admin.) districts, accumulation of attacks per district, total attacks in Germany in last week
- Jäckle and König (2017, WEP) find clear effects of e.g. accumulated attacks -> no direct model of AR, no W
- Here: integration with W, multilevel states and (electoral) districts, AR-process
- Implementation: Half-way, issue of sparse data and integration of the specifications

→ See code for a start, "SAGD_4_open_time.R"

→ Might switch to monthly data

Your own project: getting started

- Have you thought about your own spatial analysis project?
- What about your W?
- Spatial mechanisms?
- Questions?

Your own project: formal stuff

- GESS: 5000 words, deadline 6 weeks after semester ends
- Sketches/drafts minimum 3 pages of text until 17 April: research question, relevance, some state of the art, argument, research design, open questions
- Presentation 24 April: 10-15 min.
  → Preliminary code and results are encouraged!
- Send me a message / visit me at the MZES to talk!

Thank you for your attention!