Spatio-temporal statistics (MATH4341) Michaelmas term, 2024

Lecture notes part 3: Aerial unit data / spatial data on lattices

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim. To introduce Aerial unit data modeling: the basic building models.

Reading list & references:

[1] Cressie, N. (2015; Part II). Statistics for spatial data. John Wiley & Sons.
[2] Kent, J. T., & Mardia, K. V. (2022). Spatial analysis (Vol. 72). John Wiley & Sons.
[3] Gaetan, C., & Guyon, X. (2010; Ch 3). Spatial statistics and modeling (Vol. 90). New York: Springer.

Part 1. Basic stochastic models & related concepts for model building

Note 1. Recall from Section 2.2 of "Lecture notes part 1: Types of spatial data" that modeling aerial unit / lattice data types involves the use of random field models with a discrete index set. Such data are collected over areal units such as pixels, census districts or tomographic bins. Often, there is a natural neighborhood relation or neighborhood structure.

Note 2. This means we need to introduce suitable basic building models able to represent the characteristics of the underline data generating mechanisms. These as the "Discrete Random Fields".

1. Discrete Random Fields

Note 3. We re-introduce the definition of the random field with regards to the aerial unit data framework.

Definition 4. A random field $Z = (Z_s; s \in \mathcal{S})$ on a set of indexes $\mathcal{S}$ taking values in $\mathcal{Z}^{\mathcal{S}}$ is a family of random variables $\{Z_s := Z_s(\omega); s \in \mathcal{S}, \omega \in \Omega\}$ where each $Z_s(\omega)$ is defined on the same probability space $(\Omega, \mathfrak{F}, \mathrm{pr})$ and taking values in $\mathcal{Z}$.

Note 5. In aerial unite data modeling, the (spatial) set of sites $\mathcal{S}$, at which the process is defined, is discrete, it can be finite or infinite (e.g. $\mathcal{S} \subseteq \mathbb{Z}^d$), regular (e.g. pixels of an image) or irregular (states of a country).

Note 6. The general state space $\mathcal{Z}$ of the random field can be quantitative, qualitative or mixed. E.g., $\mathcal{Z} = \mathbb{R}_+$ in a Gamma random field, $\mathcal{Z} = \mathbb{N}$ in a Poisson random field, $\mathcal{Z} = \{0, 1\}$ in a binary random field.

Note 7. If $\mathcal{Z}$ is finite or countably infinite, the (joint)distribution of Z has a PMF

$$\mathrm{pr}_Z(z) = \mathrm{pr}(Z = z) = \mathrm{pr}(\{Z_s = z_s;\ s \in \mathcal{S}\}),\ \forall z \in \mathcal{Z}^{\mathcal{S}}$$

otherwise if $\mathcal{Z} \subseteq \mathbb{R}^d$ and Z continuous we will use the joint PDF.

Definition 8. The discrete set of sites $\mathcal{S} = \{s_i; i = 1, ..., n\}$ is often called lattice of sites.

Notation 9. Often we will use the notation Z_s instead of $Z(s)$ or Z_i instead of $Z(s_i)$. Hence, since $\mathcal{S} = \{s_i; i = 1, ..., n\}$, we can consider a more convenient notation

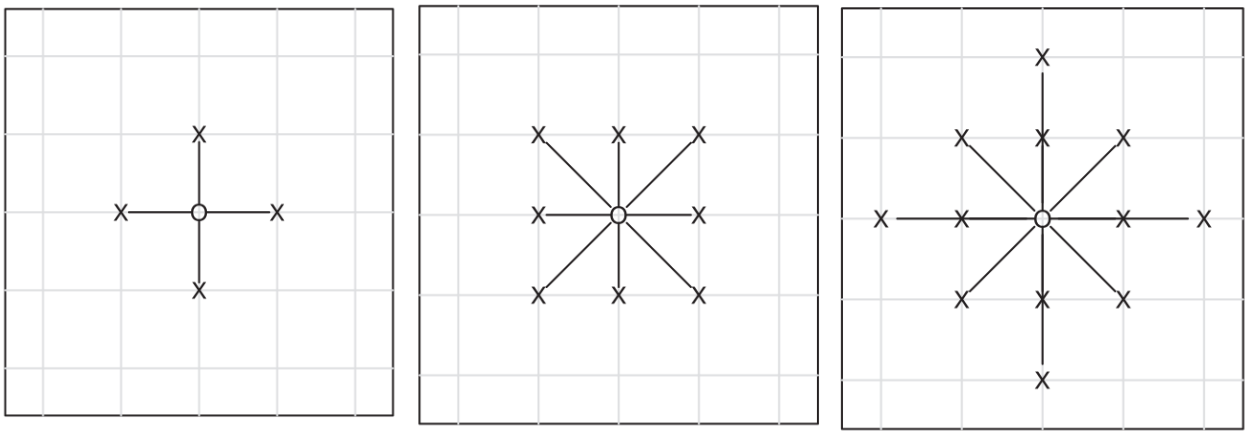
$$Z = (Z_s; s \in \mathcal{S})^\top = (Z_i = Z(s_i); i = 1, ...n)^\top.$$

Note 10. Modeling aerial unit data often requires the specification of a neighborhood relation or neighborhood structure.

Notation 11. The notation $i \sim j$ between two sites $i, j \in \mathcal{S}$ means that "sites i and j are neighboring" according to a "neighborhood relation" $\sim$.

Definition 12. Given a lattice of sites $\mathcal{S}$ and "neighborhood relation" $\sim$, we can define the neighborhood $\mathcal{N}_s$ of $s \in \mathcal{S}$ as

$$\mathcal{N}_s = \{s' \in \mathcal{S} : s \sim s'\}$$



Definition 13. Proximity matrix W is called a matrix W which aims at spatially connecting unites i and j in some fashion given some symmetric neighborhood relation $\sim$ on $\mathcal{S}$. Usually $[W]_{i,i} = 0$.

Note 14. Proximity matrix W may be such that it represents the neighborhood relation $\sim$ in a binary fashion e.g.

$$[W]_{i,j} = \begin{cases} 1 & \text{if } i \sim j \text{ and } i \neq j \\ 0 & \text{if } i \not\sim j \text{ or } i = j \end{cases}$$

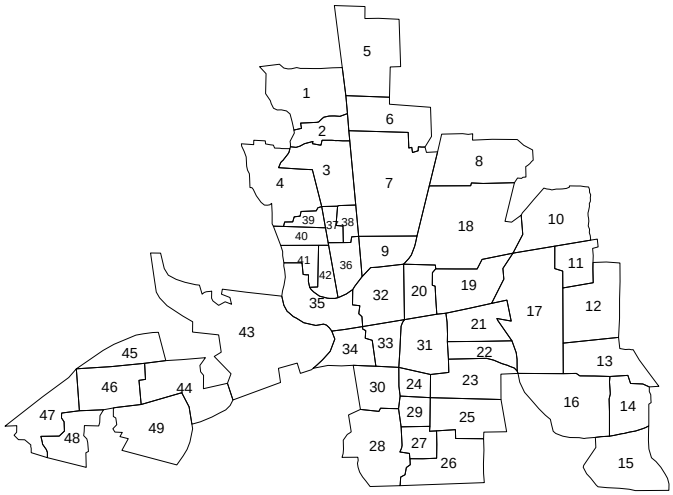


FIGURE 1.1. Lattice of spatial sites for Columbus dataset. Each neighborhood is a site. Each site is labeled. The collection of sites is the lattice of sites.

or how close site i is to site j based on some distance $d(i,j)$, e.g.

$$[W]_{i,j} = \begin{cases} 1/d(i,j) & \text{if } i \sim j \text{ and } i \neq j \\ 0 & \text{if } i \not\sim j \text{ or } i = j \end{cases}$$

Note 15. Proximity matrix W does not necessarily need to be symmetric, some times it is standardized as $[W]_{i,j} \leftarrow [W]_{i,j} / \sum_j [W]_{i,j}$.

Example 16. Consider the Columbus OH dataset which concerns spatially correlated count data arising from 49 districts/neighborhood in Columbus, OH in 1980. This is the R dataset `columbus{spdep}`. Figure 1.1 presents the sites and the lattice of sites. Each neighborhood is a site. Each site is label. The collection of sites is the lattice of sites coded with a unique labeled according to some order. One may define the "neighborhood relation $i \sim j$ considering counties that share common boarders (adjacent). Then for site $i = 43$, $i \sim j$ involves any $j \in \{44, 35, 34\}$ and for site $i = 20$, $i \sim j$ involves any $j \in \{32, 9, 18, 19, 31, 33\}$. Here $\mathcal{N}_{43} = \{44, 35, 34\}$ and $\mathcal{N}_{20} = \{32, 9, 18, 19, 31, 33\}$. The proximity matrix based on binary scheme will contain elements $W_{43,35} = 1$, $W_{43,43} = 0$, and $W_{43,33} = 0$.

Example 17. (Logistic/Ising model) Let variable Z_i denote the presence of a characteristic as $Z_i = 1$ or absence of it as $Z_i = 0$ on a site labeled by $i \in \mathcal{S}$. Then $\mathcal{Z} = \{0, 1\}$. The Ising model is defined by the (joint) PMF

$$\text{pr}_Z(z) \propto \exp\left(\alpha \sum_{i \in \mathcal{S}} z_i + \beta \sum_{\{i,j\}: i \sim j} z_i z_j\right), \ \forall z \in \mathcal{Z}^{\mathcal{S}} \tag{1.1}$$

E.g., it can model a black & white noisy image, where $\mathcal{S}$ denotes the labels of the image pixels, and Z_i denotes the presence of a black pixel ($Z_i = 1$) or its absence ($Z_i = 0$). Under Ising model (1.1), the characteristic is observed with probability $\mathrm{pr}_{Z_i}(z_i = 1) = \frac{\exp(\alpha)}{1+\exp(\alpha)}$ when $\beta = 0$. The characteristic's presence is encouraged in neighboring sites when $\beta > 0$, and discouraged when $\beta < 0$.

Notation 18. We use notation, for $\mathcal{A} \subset \mathcal{S}$

$$\mathrm{pr}_{\mathcal{A}}\left(z_{\mathcal{A}}|z_{\mathcal{S}\setminus\mathcal{A}}\right) = \mathrm{pr}\left(Z_{\mathcal{A}} = z_{\mathcal{A}}|Z_{\mathcal{S}\setminus\mathcal{A}} = z_{\mathcal{S}\setminus\mathcal{A}}\right)$$

Definition 19. Local characteristics of a random field Z on $\mathcal{S}$ with values in $\mathcal{Z}$ are the conditionals

$$\mathrm{pr}_i\left(z_i|z_{\mathcal{S}-i}\right) = \mathrm{pr}_{\{i\}}\left(z_{\{i\}}|z_{\mathcal{S}\setminus\{i\}}\right),\ i \in \mathcal{S},\ z \in \mathcal{Z}^{\mathcal{S}}$$

Example 20. (Cont. Example 17) The local characteristics of the Ising model in (1.1) are

$$\mathrm{pr}_i\left(z_i = 1|z_{\mathcal{S}-i}\right) = \frac{\exp\left(\alpha + \beta\sum_{\{i,j\}:i\sim j} z_j\right)}{1 + \exp\left(\alpha + \beta\sum_{\{i,j\}:i\sim j} z_j\right)}$$

2. Lattice Random Fields (background)

Note 21. (Recall basic properties Fourier transform) Let $\{\beta_s : s \in \mathcal{S}\}$, $\mathcal{S} \subseteq \mathbb{Z}^d$ be a set of real coefficients

- The Fourier transform
$$\tilde{\beta}(\omega) = \sum_{s\in\mathcal{S}} \beta_s e^{is^\top\omega},\ \omega \in (-\pi,\pi]^d.$$
- The inverse Fourier transform is
$$\beta_s = \int_{(-\pi,\pi]^d} e^{-is^\top\omega}\tilde{\beta}(\omega)\,\mathrm{d}\omega$$
- Regularity conditions for the Fourier transform to be a well-defined function are
  (1) If $\{\beta_s\}$ are summable, $\sum_{s\in\mathcal{S}}|\beta_s| < \infty$, then $\tilde{\beta}(\omega)$ is bounded and continuous function of ω.
  (2) If $\{\beta_s\}$ are square summable, $\sum_{s\in\mathcal{S}}|\beta_s|^2 < \infty$ then $\tilde{\beta}(\omega)$ is square-integrable over $(-\pi,\pi]^d$ and (visa versa).

Note 22. Let $(Z_s : s \in \mathcal{S})$ be a (weakly) stationary random field where $\mathcal{S} \subseteq \mathbb{Z}^d$. The following is a tool (similar to Bochner's theorem) for specifying covariance function of stationary lattice random field. It results from Herglotz's theorem.

Proposition 23. *(Herglotz's theorem) Let $c : \mathbb{Z}^d \to \mathbb{R}$ be a real-valued function on integers for $d \geq 1$. Then $c(\cdot)$ is positive semidefinite (stationary covariance function) if and only if*

it can be represented as

$$c(h) = \int_{(-\pi,\pi]^d} \exp\left(i\omega^\top h\right) dF(\omega)$$

where F is a *symmetric positive bounded finite measure on* $(-\pi,\pi]^d$ *and* $F(-\pi) = 0$. F *is called spectral measure of* $c(h)$. f *is called spectral density of* $c(h)$ *if*

$$dF(\omega) = f(\omega)\, d\omega$$

Proposition 24. *If* $c(\cdot)$ *is integrable, the spectral density* $f(\cdot)$ *can be computed by inverse Fourier transformation*

$$f(\omega) = \left(\frac{1}{2\pi}\right)^d \sum_h \exp\left(-i\omega^\top h\right) c(h)$$

Note 25. Let $\left(Y(s) : s \in \mathbb{R}^d\right)$ be a stationary random field with spectral measure $F_Y(\omega)$, $\omega \in \mathbb{R}^d$ and let $\left(Z_s : s \in \mathbb{Z}^d\right)$ with $Z_s = Y(s)$ for $s \in \mathbb{Z}^d$, then Z_s has spectral measure

$$F_Z(\omega) = \sum_{k\in\mathbb{Z}^d} F_Y(2\pi k + d\omega),\ \omega \in (-\pi,\pi]^d$$

where frequencies separated by a lag $2\pi k$, $k \in \mathbb{Z}^d$, are aliased together in the construction of$F_Z(\omega)$. Hence, there are infinitely many ways to interpolate a stationary process on $\mathbb{Z}^d$ to give a stationary random field $\mathbb{R}^d$.

Note 26. Let $(U_s : s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$, be a stationary random field covariance function

$$c_U(h) = \int_{(-\pi,\pi]^d} e^{i\omega^\top h} f(\omega)\, \mathrm{d}\omega$$

and spectral density $f(\omega)$ over $(-\pi,\pi]^d$. Let $(V_s : s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$, be a random field such as

$$V_s = \sum_{h\in\mathbb{Z}^d} U_{s+h}\beta_h,\ s \in \mathcal{S} \subseteq \mathbb{Z}^d$$

where $\left\{\beta_h : h \in \mathbb{Z}^d\right\}$ are summable functions, i.e. $\sum_h |\beta_h| < \infty$. The covariance function of $(V_s : s \in \mathcal{S})$ is

$$c_V(h) = \mathrm{Cov}(V_s, V_{s+h}) = \sum_{t\in\mathbb{Z}^d}\sum_{t'\in\mathbb{Z}^d} \beta_t\beta_{t'} c_U(h + t - t')$$

with spectral measure

$$\mathrm{d}F_V(\omega) = \left|\tilde{\beta}(\omega)\right|^2 \mathrm{d}F_U(\omega),\ \omega \in (-\pi,\pi]^d \tag{2.1}$$

where

$$\tilde{\beta}(\omega) = \sum_{h\in\mathbb{Z}^d} \beta_h e^{ih^\top\omega},\ \omega \in (-\pi,\pi]^d \tag{2.2}$$

is the Fourier transform of β_h.

Proof. This is straightforward from Proposition 23 (Herglotz's theorem) □

3. Compatibility of conditional distributions

Note 27. Here, we discuss how to represent a joint probability distribution via its full conditionals. We need this for model building purposes.

Definition 28. Let random vector $Z = (Z_1, ..., Z_n)$ with joint distribution $\pi(Z_1, ..., Z_n)$. The set of distributions $\{\pi_i(\cdot|Z_{-i}); i = 1, ...n\}$ is called compatible to the joint distribution $\pi(Z_1, ..., Z_n)$ if the joint distribution $\pi(Z_1, ..., Z_n)$ has conditionals $\{\pi_i(Z_i|Z_{-i}); i = 1, .., n\}$.

Note 29. To specify suitable building models representing spatial dependency of a random field $(Z_i)_{i\in\mathcal{S}}$, it is often easier to visualize the joint distribution pr_z in terms of conditional distributions $\{\pi_i(Z_i|Z_{\mathcal{S}-i}); i \in \mathcal{S}\}$ rather than directly.

Note 30. Thus, instead of specifying a joint model for $(Z_i)_{i\in\mathcal{S}}$, a researcher may propose putative families of conditional distributions $\{\pi_i(Z_i|Z_{\mathcal{S}-i}); i \in \mathcal{S}\}$. However, an arbitrary chosen set of conditional distributions $\{\pi_i(\cdot|\cdot); i \in \mathcal{S}\}$ is not generally compatible, in the sense that there exists a joint distribution for $(Z_i)_{i\in\mathcal{S}}$, and hence we need to impose conditions.

Note 31. In what follows, we discuss necessary and sufficient conditions regarding compatibility.

Proposition 32. *(Compatibility condition) Let F be a joint distribution with $dF(x, y) = f(x, y)\, d(x, y)$ on $\mathcal{S}_x \times \mathcal{S}_y$. Let candidate condition distributions*

$$G \text{ with } dG(x|y) = g(x|y)\, dx, \text{ on } x \in \mathcal{S}_x$$
$$Q \text{ with } dQ(y|x) = q(y|x)\, dy, \text{ on } y \in \mathcal{S}_y$$

and let $N_g = \{(x, y) : g(x|y) > 0\}$ and $N_q = \{(x, y) : q(y|x) > 0\}$. A distribution F with conditionals exists iff

(1) $N_g = N_q = N$
(2) there exist functions u and $v$where $g(x|y)/q(y|x) = u(x)\, v(y)$ for all $(x, y) \in N$ and $\int u(x)\, dx < \infty$

Proof. Omitted[1]. □

[1]See Arnold, B. C., & Press, S. J. (1989). Compatible conditional distributions. Journal of the American Statistical Association, 84(405), 152-156.

Note 33. Essentially the above conditions guarantee that

$$k(y)\, g(x|y) = f(x,y) = h(x)\, q(y|x)$$

where k, g, h, q are densities.

Example 34. The conditionals $x|y \sim \mathrm{N}(a+by, \sigma^2+\tau^2 y^2)$ and $y|x \sim \mathrm{N}(c+dx, \tilde{\sigma}^2+\tilde{\tau}^2 x^2)$ are compatible if $\tau^2 = \tilde{\tau}^2 = 0$, $d/\tilde{\sigma}^2 = b/\sigma^2$, and $|db| < 1$.

Solution. See Exercise 29 in the Exercise sheet.

Note 35. Proposition 32 can be extended to more dimensions. For more info see (Arnold, B. C., & Press, S. J. (1989). in footnote 1)

Note 36. The following theorem shows that local characteristics can determine the entire distribution in certain cases.

Theorem 37. *(Besag's factorization theorem; Brook's Lemma) Let Z be a $\mathcal{Z}$ valued random field taking values in $\mathcal{Z}^{\mathcal{S}}$ where $\mathcal{S} = \{1, ..., n\}$ with $n \in \mathbb{N}$, and such as $pr_Z(z) > 0$, $\forall z \in \mathcal{Z}^{\mathcal{S}}$. Then for all*

$$\frac{pr_Z(z)}{pr_Z(z^*)} = \prod_{i=1}^{n} \frac{pr_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{pr_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}, \quad \forall z, z^* \in \mathcal{Z}^{\mathcal{S}} \tag{3.1}$$

Proof. I will show that

$$\mathrm{pr}_Z(z) = \prod_{i=1}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)} \mathrm{pr}_Z(z^*)$$

It is

$$\mathrm{pr}_Z(z_1, ..., z_n) = \frac{\mathrm{pr}_n(z_n|z_1, ..., z_{n-2}, z_{n-1})}{\mathrm{pr}_n(z_n^*|z_1, ..., z_{n-2}, z_{n-1})} \mathrm{pr}_Z(z_1, ..., z_{n-1}, z_n^*)$$

Let proposition P_j be

$$\mathrm{pr}_Z(z) = \prod_{i=n-j}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)} \mathrm{pr}_Z\left(z_1, ..., z_{n-j-1}, z_{n-j}^*, ..., z_n^*\right)$$

Proposition P_0 is true

$$\mathrm{pr}_Z(z) = \frac{\mathrm{pr}_n(z_n|z_1, ..., z_{n-1})}{\mathrm{pr}_n(z_n^*|z_1, ..., z_{n-1})} \mathrm{pr}_Z(z_1, ..., z_{n-1}, z_n^*) \tag{3.2}$$

Proposition P_1 is true

$$\mathrm{pr}_Z(z_1, ..., z_{n-1}, z_n^*) = \frac{\mathrm{pr}_{n-1}(z_{n-1}|z_1, ..., z_{n-2}, z_n^*)}{\mathrm{pr}_{n-1}(z_{n-1}^*|z_1, ..., z_{n-2}, z_n^*)} \mathrm{pr}_Z\left(z_1, ..., z_{n-2}, z_{n-1}^*, z_n^*\right)$$

Assume that P_j is true. Then proposition $\mathrm{P}_{j+1}$is true as well, because

$$\begin{aligned}
\mathrm{pr}_Z(z) &= \prod_{i=n-j}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)}{\mathrm{pr}_i\left(z^*_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)} \mathrm{pr}_Z\left(z_1,...,z_{n-j-1},z^*_{n-j},...,z^*_n\right) \\
&= \prod_{i=n-j}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)}{\mathrm{pr}_i\left(z^*_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)} \\
&\quad \times \frac{\mathrm{pr}_{n-j-1}\left(z_{n-j-1}|z_1,...,z_{n-j-2},z^*_{n-j},...,z^*_n\right)}{\mathrm{pr}_{n-j-1}\left(z^*_{n-j-1}|z_1,...,z_{n-j-2},z^*_{n-j},...,z^*_n\right)} \mathrm{pr}_Z\left(z_1,...,z_{n-j-2},z^*_{n-j-1},...,z^*_n\right) \\
&= \prod_{i=n-(j+1)}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)}{\mathrm{pr}_i\left(z^*_i|z_1,...,z_{i-1},z^*_{i+1},...,z^*_n\right)} \mathrm{pr}_Z\left(z_1,...,z_{n-(j+1)-1},z^*_{n-(j+1)},...,z^*_n\right)
\end{aligned}$$

Then (3.1) is correct according to the induction principle. □

Note 38. Theorem 37 shows that the joint $\mathrm{pr}_Z(\cdot)$ can be constructed from its conditionals $\{\mathrm{pr}_i(\cdot|\cdot)\}$ if distributions $\{\mathrm{pr}_i(\cdot|\cdot)\}$ are compatible for $\mathrm{pr}_Z(\cdot)$, under the requirement that this construction is invariant wrt the coordinate permutation $\{1,...,n\}$ and the reference state z^*– these invariances correspond to the conditions in Proposition 32.

4. Spatial autoregressive models

We present two basic spatial Autoregressive models, the SAR and CAR, able to represent spatial dependency.

4.1. **Simultaneous AutoRegressive (SAR) models.**

Definition 39. "Autoregressive" representation for lattice random field $(Z_s; s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$ is called

$$\sum_{h \in \mathcal{H} \cup \{0\}} a_h \left(Z_{s+h} - \mu_{s+h}\right) = \varepsilon_s, \ s \in \mathcal{S} \tag{4.1}$$

where $\{a_h; h \in \mathcal{H}\}$, $\mathcal{H} = \left\{h \in \mathbb{Z}^d - \{0\} : s + h \in \mathcal{S}\right\}$, are coefficients and $\{\varepsilon_s; s \in \mathcal{S}\}$ is a discrete white noise random field with variance $\lambda_s = \mathrm{Var}(\varepsilon_s)$. We will denote it as SAR($\mathcal{H}$).

4.1.1. *Assuming stationarity.*

Note 40. We will focus our study on stationary CAR random field $(Z_s; s \in \mathcal{S})$, i.e. $\mu_s = \mu$ and $\lambda_s = \sigma^2_\varepsilon$ for $s \in \mathcal{S} \subseteq \mathbb{Z}^d$.

Note 41. $\{\varepsilon_s; s \in \mathcal{S}\}$ has c.f. $c_\varepsilon(h) = \sigma_\varepsilon^2 \delta_{\{0\}}(h)$ and hence spectral density $f_\varepsilon(\omega) = \sigma_\varepsilon^2/(2\pi)^d$. The spectral density f for $(Z_s; s \in \mathcal{S})$ is

$$f(\omega) = \frac{1}{|\tilde{a}(\omega)|^2} f_\varepsilon(\omega) = \frac{1}{|\tilde{a}(\omega)|^2} \left(\frac{1}{2\pi}\right)^d \sum_h e^{-i\omega^\top h} c_\varepsilon(h)\, \mathrm{d}h \tag{4.2}$$
$$= \frac{1}{|\tilde{a}(\omega)|^2} \left(\frac{1}{2\pi}\right)^d \sum_{h\in\mathbb{Z}^d} e^{-i\omega^\top h} \sigma_\varepsilon^2 \delta_{\{0\}}(h)\, \mathrm{d}h = \frac{1}{|\tilde{a}(\omega)|^2} \frac{\sigma_\varepsilon^2}{(2\pi)^d}$$

where $\tilde{a}(\omega) = \sum_h a_h e^{ih^\top \omega}$ since (26).

Note 42. For random field $(Z_s; s \in \mathcal{S})$ to be stationary, $f(\omega)$ in (5.1) must be integrable function and bounded function on $\omega \in (-\pi, \pi]^d$. Hence, we can set restrictions on coefficients

$$a_0 > 0, \text{ and } \sum_h |a_h| < a_0 \tag{4.3}$$

satisfying regularity conditions in (21).

Note 43. To make model (6.3) identifiable and use it for inference, we can introduce further restrictions $a_h = a_{-h}$.

Definition 44. Lattice random field $(Z_s; s \in \mathcal{S})$ is called Simultaneous AutoRegressive (SAR) if it is given in an "Autoregression" representation (6.3)

$$\sum_{h\in\mathcal{N}} a_h Z_{s+h} = \varepsilon_s,\ s \in \mathcal{S}$$

whose coefficients $\{a_h; h \in \mathcal{H}\}$ satisfy the symmetry condition

$$a_h = a_{-h},\ \forall h$$

and $f_Z(\omega)$ in (5.1) is an integrable function over $\omega \in (-\pi, \pi]^d$.

4.1.2. *Assuming Gaussian distribution.*

Note 45. Following we provide the matrix form of the definition used in software implementations.

Definition 46. Consider discrete set of sites $\mathcal{S} = \{s_i; i = 1, ..., n\}$ and a lattice random field $(Z_s; s \in \mathcal{S})$. Vectorize $Z = (Z_1, ..., Z_n)^\top$ with $Z_i = Z(s_i)$ and set

$$Z = \mu + A(Z - \mu) + E \Longleftrightarrow E = (I - A)(Z - \mu)$$

Assume that matrix A is such that $[A]_{i,i} = 0$ and $(I - A)^{-1}$ exists. Assume n-dimensional Gaussian random vector $E \sim \mathrm{N}_n(0, \Lambda)$ with $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$. We say that Z follows a "Gaussian" Simultaneous Autoregressive model.

Note 47. The joint distribution of Z following the SAR model in Definition 46 is

$$Z \sim \mathrm{N}\left(\mu, (I-A)^{-1}\Lambda\left(I-A^{\top}\right)^{-1}\right) \tag{4.4}$$

Proof. Z is a linear combination of Gaussian random vectors, hence if follows a Gaussian distribution. It's mean and variance are

$$\begin{aligned}\mathrm{E}(Z) &= \mathrm{E}\left((I-A)^{-1}E+\mu\right)=\mu,\\ \mathrm{Var}(Z) &= \mathrm{Var}\left((I-A)^{-1}E+\mu\right)=(I-A)^{-1}\mathrm{Var}(E)\left(I-A^{\top}\right)^{-1}=(I-A)^{-1}\Lambda\left(I-A^{\top}\right)^{-1}\end{aligned}$$

□

4.2. **Conditional autoregressive models (CAR).**

Definition 48. A lattice random field $(Z_s; s\in\mathcal{S})$, $\mathcal{S}\subseteq\mathbb{Z}^d$ is called Conditional AutoRegressive (CAR) model if

$$\begin{aligned}\mathrm{E}(Z_s|Z_{S-s}) &= \mu_s+\sum_{h\in\mathcal{H}} b_h\left(Z_{s+h}-\mu_{s+h}\right)\\ \mathrm{Var}(Z_s|Z_{S-s}) &= \kappa_s\end{aligned}$$

where $b_h=-b_h$ and $\mathcal{H}=\left\{h\in\mathbb{Z}^d-\{0\}: s+h\in\mathcal{S}\right\}$. We will denote it as CAR($\mathcal{H}$).

Note 49. Alternatively CAR lattice random field $(Z_s; s\in\mathcal{S})$, $\mathcal{S}\subseteq\mathbb{Z}^d$can be written as

$$Z_s=\mu_s+\sum_{h\in\mathcal{H}} b_h\left(Z_{s+h}-\mu_{s+h}\right)+\epsilon_s \tag{4.5}$$

where $(\epsilon_s : s\in\mathcal{S})$ is the residual random field with mean $\mathrm{E}(\epsilon_s)=0$ and variance $\mathrm{Var}(\epsilon_s)=\kappa_s$. Also $b_h=-b_h$ and $\mathcal{H}=\left\{h\in\mathbb{Z}^d-\{0\}: s+h\in\mathcal{S}\right\}$.

4.2.1. *Assuming stationary.*

Note 50. We will focus our study on stationary CAR random field $(Z_s; s\in\mathcal{S})$, and hence we set $\mu_s=\mu$ and $\kappa_s=\kappa$ for $s\in\mathcal{S}\subseteq\mathbb{Z}^d$.

Note 51. It is

$$\begin{aligned}\mathrm{E}(Z_s\epsilon_s) &= \mathrm{E}\left(\left(\epsilon_s-\mu-\sum_{h\in\mathcal{H}} b_h\left(Z_{s+h}-\mu\right)\right)\epsilon_s\right)\\ &= \mathrm{E}\left(\epsilon_s^2\right)-\mu\mathrm{E}(\epsilon_s)-\sum_{h\in\mathcal{H}} b_h\mathrm{E}\left(\left(Z_{s+h}-\mu\right)\epsilon_s\right)\\ &= \mathrm{E}\left(\epsilon_s^2\right)+0+0=\kappa\end{aligned}$$

Note 52. The covariance function $c(\cdot)$ of stationary CAR random field $(Z_s; s\in\mathcal{S})$ is

$$c(h)=\sum_{h'\in\mathcal{H}} b_{h'}c\left(h-h'\right)+\kappa\delta_{\{0\}}(h)$$

as, for $h \neq 0$ it is

$$\begin{aligned} c(h) &= \mathrm{E}\left(\left(Z_s-\mu\right)\left(Z_{s+h}-\mu\right)\right) = \mathrm{E}\left(\left(\sum_{h'\in\mathcal{H}} b_{h'}\left(Z_{s+h'}-\mu\right)+\epsilon_s\right)\left(Z_{s+h}-\mu\right)\right) \\ &= \mathrm{E}\left(\left(\sum_{h'\in\mathcal{H}} b_{h'}\left(Z_{s+h'}-\mu\right)+\epsilon_s\right)\left(Z_{s+h}-\mu\right)\right) \\ &= \sum_{h'\in\mathcal{H}} b_{h'}\mathrm{E}\left(Z_{s+h'}Z_{s+h}\right) - \mu^2\sum_{h'\in\mathcal{H}} b_{h'} + \mathrm{E}\left(\epsilon_s Z_{s+h}\right) \\ &= \sum_{h'\in\mathcal{H}} b_{h'}b_{h'}c\left(h-h'\right) - 0 + 0 \end{aligned}$$

and for $h = 0$

$$\begin{aligned} c(0) &= \mathrm{E}\left(\left(Z_s-\mu\right)\left(Z_s-\mu\right)\right) = \\ &= \mathrm{E}\left(\left(\sum_{h\in\mathcal{H}} b_h\left(Z_{s+h}-\mu\right)+\epsilon_s\right)\left(\sum_{h'\in\mathcal{H}} b_{h'}\left(Z_{s+h'}-\mu\right)+\epsilon_s\right)\right) \\ &= \sum_{h\in\mathcal{H}}\sum_{h'\in\mathcal{H}} b_{h'}b_{h'}c\left(h-h'\right)+\kappa \end{aligned}$$

Note 53. The spectral density of the covariance function $c(h)$ of stationary CAR random field $(Z_s; s \in \mathcal{S})$ is computed by inverse Fourier transform as

$$\begin{aligned} f(\omega) &= \left(\frac{1}{2\pi}\right)^d \sum_{\mathrm{h}} e^{-i\omega^\top h} c(h) \\ &= \left(\frac{1}{2\pi}\right)^d \sum_{\mathrm{h}} e^{-i\omega^\top h}\left(\sum_{h'\in\mathcal{H}} b_{h'}c\left(h-h'\right)+\kappa\delta_{\{0\}}(h)\right) \\ &= \tilde{b}(\omega) f(\omega) + \kappa\frac{1}{(2\pi)^d} \implies \\ f(\omega) &= \frac{\kappa}{(2\pi)^d}\frac{1}{1-\tilde{b}(\omega)},\ \tilde{b}(\omega) = \sum_{h\in\mathcal{H}} b_h e^{ih^\top\omega} = \sum_{h\in\mathcal{H}} b_h\cos\left(\omega^\top h\right) \end{aligned} \tag{4.6}$$

Hence sufficient conditions for CAR random field $(Z_s; s \in \mathcal{S})$ in (5.3) to be stationary is the spectral density (5.4) to be bounded. This is true if $\tilde{b}(\omega) < 1$ which is implied by

$$\sum_{h\in\mathcal{H}} |b_h| < \infty$$

satisfying regularity conditions in (21).

4.2.2. *Assuming Gaussian distribution.*

Note 54. Following we provide the matrix form of the definition used in software implementations.

Definition 55. "Gaussian" Conditional autoregressive model, CAR, assumes that the local characteristics $\{\mathrm{pr}_i(z_i|z_{\mathcal{S}-i})\}$ are Gaussian distributions

$$Z_i|z_{\mathcal{S}-i} \sim \mathrm{N}\left(\underbrace{\mu_i + \sum_{j\neq i} b_{i,j}(Z_j - \mu_j)}_{=\mathrm{E}(Z_i|Z_{\mathcal{S}-i})}, \underbrace{\kappa_i}_{=\mathrm{Var}(Z_i|Z_{\mathcal{S}-i})}\right), \quad \forall i \in \mathcal{S} \tag{4.7}$$

Proposition 56. *Let $K = diag(\{\kappa_i\})$ with $\kappa_i > 0$, matrix B with $B_{i,i} := [B]_{i,i} = 0$, and real vector μ with suitable dimensions. If Z follows a Gaussian CAR (Definition 55), $I - B$ is non-singular, and $(I-B)^{-1}K > 0$, then the joint distribution of Z is*

$$Z \sim N\left(\mu, (I-B)^{-1}K\right). \tag{4.8}$$

Proof. Without lose of generality, consider zero mean $\mu = 0$ (or equivalently set $Z := Z - \mu$). The full conditionals $Z_i|z_{\mathcal{S}-i}$ in (4.7) are compatible with the joint distribution $\mathrm{pr}_Z(z)$. By using Besag's factorization theorem (Theorem 37) with reference state/configuration $z^* = 0$ we get

$$\begin{aligned}
\mathrm{pr}_Z(z) &= \prod_{i=1}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*=0,...,z_n^*=0\right)}{\mathrm{pr}_i\left(z_i^*=0|z_1,...,z_{i-1},z_{i+1}^*=0,...,z_n^*=0\right)}\mathrm{pr}_Z(z^*=0) \\
&= \prod_{i=1}^{n} \frac{\mathrm{N}\left(z_i|\sum_{j<i} b_{i,j}z_j + 0, \kappa_i\right)}{\mathrm{N}\left(0|\sum_{j<i} b_{i,j}z_j + 0, \kappa_i\right)}\mathrm{pr}_Z(z^*=0) \\
&\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i - \sum_{j<i} b_{i,j}z_j\right)^2 + \frac{1}{2\kappa_i}\left(0 - \sum_{j<i} b_{i,j}z_j\right)^2\right) \\
&= \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i^2 - 2z_i\sum_{j<i} b_{i,j}z_j\right)\right)\mathrm{pr}_Z(z^*=0) \\
&= \exp\left(-\sum_i \frac{z_i^2}{2\kappa_i} + \frac{1}{2}2\sum_i\sum_{j<i}\frac{b_{i,j}}{\kappa_i}z_iz_j\right)\mathrm{pr}_Z(z^*=0) \\
&= \exp\left(-\frac{1}{2}z^\top K^{-1}z + \frac{1}{2}z^\top K^{-1}Bz\right)\mathrm{pr}_Z(z^*=0) = \exp\left(-\frac{1}{2}z^\top\left[K^{-1}(I-B)\right]z\right)\mathrm{pr}_Z(z^*=0) \\
&= \mathrm{N}\left(z|0,(I-B)^{-1}K\right)
\end{aligned} \tag{4.9}$$

Recovering the mean from (4.9), it is

$$\mathrm{pr}_Z(z) = \mathrm{N}\left(z-\mu|0,(I-B)^{-1}K\right) = \mathrm{N}\left(z|\mu,(I-B)^{-1}K\right)$$

□

Note 57. When CAR is used for modeling, B is often specified to be sparse either due to some natural problem specific property, or for our computational convenience as is may allow the use of sparse solvers. To achieve this, one way is to specify $B = \phi W$ where $\phi > 0$ and W is an adjacency/proximity matrix; that is $[B]_{i,j} = \phi 1(i \sim j)\, 1(i \neq j)$ will be non-zero only for adjacent pairs i and j.

Note 58. The system in (4.8) can be rewritten as

$$Z = \mu + B(Z-\mu) + E \Longleftrightarrow E = (I-B)(Z-\mu) \tag{4.10}$$

by setting $E = (I-B)(Z-\mu)$. The distribution of Z in (4.8) induces a distribution on E as $E \sim \mathrm{N}\left(0, K(I-B)^{\top}\right)$ because

$$\mathrm{E}(E) = \mathrm{E}((I-B)(Z-\mu)) = (I-B)\,\mathrm{E}(Z-\mu) = 0$$
$$\mathrm{Var}(E) = \mathrm{Var}((I-B)Z) = (I-B)\,\mathrm{Var}(Z)\,(I-B)^{\top} = (I-B)(I-B)^{-1}K(I-B)^{\top}$$

4.3. **A comparison between CAR and SAR.**

Note 59. In most of the cases a SAR model can be written as a CAR model implying that CAR family of models can be more general than that of SAR models.

Note 60. We compare the use and flexibility of the two models, in particular the stationary CAR($\mathcal{H}_{\mathrm{CAR}}$) and the stationary SAR($\mathcal{H}_{\mathrm{SAR}}$) on $\mathbb{Z}^d$.

Example 61. Consider the stationary SAR and CAR models on $\mathbb{Z}^d$.

(1) Every stationary SAR model on $\mathbb{Z}^d$ is a stationary CAR model on $\mathbb{Z}^d$.
(2) When $d \geq 2$, the family of CAR models is larger than that of the SAR models.

Proof. Consider the stationary models SAR($\mathcal{H}_{\mathrm{SAR}}$) and CAR($\mathcal{H}_{\mathrm{CAR}}$) :

$$\mathrm{CAR}(\mathcal{H}_{\mathrm{CAR}}): Z_s = \mu + \sum_{h\in\mathcal{H}_{\mathrm{CAR}}} b_h(Z_{s+h}-\mu) + \epsilon_s,\ s \in \mathcal{S} \tag{4.11}$$

$$\mathrm{Var}(\epsilon_s) = \kappa,\ \mathrm{E}(\epsilon_s) = 0 \tag{4.12}$$

$$\mathrm{SAR}(\mathcal{H}_{\mathrm{SAR}}): \sum_{h\in\mathcal{H}_{\mathrm{SAR}}} a_h(Z_{s+h}-\mu) = \varepsilon_s,\ s \in \mathcal{S} \tag{4.13}$$

$$\mathrm{Var}(\varepsilon_s) = \lambda,\ \mathrm{E}(\epsilon_s) = 0 \tag{4.14}$$

(1) The spectral density of CAR($\mathcal{H}_{\text{CAR}}$) is

$$f_{\text{CAR}(\mathcal{H}_{\text{CAR}})}(\omega) = \frac{\kappa}{(2\pi)^d} \frac{1}{1 - \sum_{h \in \mathcal{H}_{\text{CAR}}} b_h e^{ih^\top \omega}} = \frac{\lambda}{(2\pi)^d} \frac{1}{\frac{\lambda}{\kappa} - \sum_{h \in \mathcal{H}_{\text{CAR}}} \frac{\lambda}{\kappa} b_h e^{ih^\top \omega}}$$

I can choose to set $\mathcal{H}_{\text{CAR}} = \{i - j : i \in \mathcal{H}_{\text{SAR}} \cup \{0\} \text{ and } j \in \mathcal{H}_{\text{SAR}} \cup \{0\}\}$,

$$b_h = \begin{cases} -\frac{\kappa}{\lambda} \sum_{v \in \mathcal{H}_{\text{SAR}}} a_v a_{v+h} & , h \neq 0 \\ 1 & , h = 0 \end{cases}, \text{ for } h \in \mathcal{H}_{\text{CAR}}$$

then

$$\begin{aligned} f_{\text{CAR}(\mathcal{H}_{\text{CAR}})}(\omega) &= \frac{\lambda}{(2\pi)^d} \frac{1}{\frac{\lambda}{\kappa} - \sum_{h \in \mathcal{H}_{\text{CAR}}} \sum_{v \in \mathcal{H}_{\text{SAR}}} a_v a_{v+h} e^{i(h \pm v)^\top \omega}} \\ &= \frac{\lambda}{(2\pi)^d} \frac{1}{\left| \sum_{h \in \mathcal{H}_{\text{SAR}}} a_h e^{ih^\top \omega} \right|^2} \end{aligned}$$

which is the spectral density of SAR($\mathcal{H}_{\text{SAR}}$).

(2) We consider

$$Z_s = c \sum_{h \in \mathcal{H}_{\text{CAR}}} Z_{s+h} + \epsilon_s$$

where $\mathcal{H}_{\text{CAR}} = \{h \in \mathbb{Z}^2 : |h|_1 = 1\}$ and $c \neq 0$. Then

$$f_{\text{CAR}(\mathcal{H}_{\text{CAR}})}(\omega) = \frac{\kappa}{(2\pi)^d} \frac{1}{1 - \sum_{h \in \mathcal{H}_{\text{CAR}}} b_h e^{ih^\top \omega}} = \frac{\lambda}{(2\pi)^d} \frac{1}{c\left(1 - 2\cos(\omega_1 h_1) - 2\cos(\omega_2 h_2)\right)}$$

If some SAR($\mathcal{H}_{\text{SAR}}$) had this spectral density, then $\mathcal{H}_{\text{SAR}} \subset \mathcal{H}_{\text{CAR}}$. Noting that it must be either $a_{(1,0)} \neq 0$ or $a_{(0,-1)} \neq 0$ and either either $a_{(0,1)} \neq 0$ or $a_{(-1,0)} \neq 0$. Assume $a_{(1,0)} \neq 0$ and $a_{(0,1)} \neq 0$. In this case the spectral density should contain a term $\cos(\omega_1 h_1 - \omega_2 h_2)$ which is not the case. So stationary CAR model has no stationary SAR representation.

□

5. Related Random Fields with particular properties

Note 62. We introduce more general modeling structures for basic spatial models which are computationally convenient yet quite descriptive for spatial statistical modeling. Convenient because they aim to break a high-dimensional problem into smaller ones using conditional independence, and reasonable because they allow representation of spatial dependence as well. We introduce the Gibbs Random Fields and the Markov Random Fields. The aforesaid Ising, CAR, and SAR models are just special cases of modeling structures.

5.1. **Gibbs Random Fields.**

Notation 63. Recall notation $z_{\mathcal{A}} = (z_i : i \in \mathcal{A})$ and $\mathcal{Z}^{\mathcal{A}} = \left\{z_{\mathcal{A}} : z \in \mathcal{Z}^{\mathcal{S}}\right\}$ for $\mathcal{A} \subseteq \mathcal{S}$.

Definition 64. Let $\mathcal{S} \neq \emptyset$ be a finite collection of sites. Let $\mathcal{Z} \subset \mathbb{R}$. Interaction potential is a family $\mathcal{V} = \{V_{\mathcal{A}} : \mathcal{A} \subseteq \mathcal{S}\}$ of potential functions $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \to \mathbb{R}$ such that $V_{\emptyset}(\cdot) := 0$ and for every set $\mathcal{A} \subseteq \mathcal{S}$ the sum

$$U_{\mathcal{A}}^{\mathcal{V}}(z) = \sum_{\{\forall \mathcal{B} \subseteq \mathcal{S} : \mathcal{A} \cap \mathcal{B} \neq \emptyset\}} V_{\mathcal{B}}(z_{\mathcal{B}}) \tag{5.1}$$

exists.

Definition 65. In Definition 64, the function $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \to \mathbb{R}$ is called potential on $\mathcal{A}$.

Definition 66. In Definition 64, the function $U_{\mathcal{A}}^{\mathcal{V}}(z)$ in (5.1) is called energy function of interaction potential $\mathcal{V}$ on $\mathcal{A}$ is called.

Definition 67. The interaction potential $\mathcal{V}$ is said to be admissible if for all $\mathcal{B} \subseteq \mathcal{S}$ and $z_{\mathcal{S}\setminus\mathcal{B}} \in \mathcal{Z}^{\mathcal{S}\setminus\mathcal{B}}$

$$C_{\mathcal{A}}^{\mathcal{V}}\left(z_{\mathcal{S}\setminus\mathcal{A}}\right) = \int \exp\left(U_{\mathcal{A}}^{\mathcal{V}}\left(\left(z_{\mathcal{A}}, z_{\mathcal{S}\setminus\mathcal{A}}\right)\right)\right) \mathrm{d}z_{\mathcal{A}} < \infty$$

Note 68. This allow as to define a distribution corresponding to the energy.

Definition 69. Let Z be $\mathcal{Z}$ valued Random Field on a finite collection of sites $\mathcal{S}$ with $\mathcal{S} \neq \emptyset$, and let $\mathcal{V} = \{V_{\mathcal{A}} : \mathcal{A} \subseteq \mathcal{S}\}$ be an interaction potential of functions $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \to \mathbb{R}$. Assume that $\mathcal{V}$ is admissible. Then Z is a Gibbs Random Field with interaction potentials $\mathcal{V} = \{V_{\mathcal{A}} : \mathcal{A} \subseteq \mathcal{S}\}$ if

$$\mathrm{pr}_Z\left(z_{\mathcal{A}} | z_{\mathcal{S}\setminus\mathcal{A}}\right) = \frac{1}{C_{\mathcal{A}}^{\mathcal{V}}\left(z_{\mathcal{S}\setminus\mathcal{A}}\right)} \exp\left(\underbrace{\sum_{\{\mathcal{B} \subseteq \mathcal{S} \,:\, \mathcal{A} \cap \mathcal{B} \neq \emptyset\}} V_{\mathcal{B}}(z_{\mathcal{B}})}_{= U_{\mathcal{A}}^{\mathcal{V}}(z)}\right), \quad z \in \mathcal{Z}^{\mathcal{S}} \tag{5.2}$$

Definition 70. The normalizing integral $C_{\mathcal{A}}^{\mathcal{V}}$ in (5.2) is called partition function.

Notation 71. For the marginal $\mathrm{pr}_Z(z_{\mathcal{S}})$ we will denote

$$\mathrm{pr}_Z(z_{\mathcal{S}}) = \frac{1}{C_{\mathcal{S}}^{\mathcal{V}}} \exp\left(U_{\mathcal{S}}^{\mathcal{V}}(z)\right) = \frac{1}{C_{\mathcal{S}}^{\mathcal{V}}} \exp\left(\sum_{\mathcal{B} \subseteq \mathcal{S}} V_{\mathcal{B}}(z_{\mathcal{B}})\right), \; z \in \mathcal{Z}^{\mathcal{S}}$$

where $C_{\mathcal{S}}^{\mathcal{V}} < \infty$ is the constant. In this case (and when it is clear), to easy the notation, we can omit $\cdot_{\mathcal{S}}^{\mathcal{V}}$ and just write

$$\mathrm{pr}_Z(z_{\mathcal{S}}) = \frac{1}{C} \exp\left(\sum_{\mathcal{B} \subseteq \mathcal{S}} V_{\mathcal{B}}(z_{\mathcal{B}})\right), \; z \in \mathcal{Z}^{\mathcal{S}}$$

Example 72. (Ising model) In Example 17, the Ising model (1.1) has potentials

$$\begin{aligned} V_{\emptyset}(z) &= 0 \\ V_{\{i\}}(z), &= \alpha z_i \, \forall i \in \mathcal{S} \\ V_{\{i,j\}}(z) &= \begin{cases} \beta z_i z_j & \text{if } i \sim j \\ 0 & \text{if } i \not\sim j \end{cases} \\ V_{\mathcal{A}}(z) &= 0, \text{ if } \operatorname{card}(\mathcal{A}) > 2 \end{aligned}$$

it has energy function

$$U(z) := U_{\mathcal{S}}^{\mathcal{V}}(z_{\mathcal{S}}) = \alpha \sum_{i \in \mathcal{S}} z_i + \beta \sum_{\{i \in \mathcal{S}, j \in \mathcal{S}: i \sim j\}} z_i z_j$$

and it has energy function conditional on $\mathcal{S} \backslash \mathcal{B}$

$$U_{\mathcal{B}}^{\mathcal{V}}\left(z_{\mathcal{B}} | z_{\mathcal{S} \backslash \mathcal{B}}\right) = \alpha \sum_{i \in \mathcal{B}} z_i + \beta \sum_{\{i \in \mathcal{B}, j \in \mathcal{S}: i \sim j\}} z_i z_j$$

Note 73. In what follows we discuss identifiability matters related to the potential.

Definition 74. The interaction potential $\mathcal{V}$ is said to be normalized with respect to a normalizing reference point $\zeta \in \mathcal{Z}$ if there is $i \in \mathcal{S}$ which for any $z \in \mathcal{Z}^{\mathcal{S}}$ with $z_i = \zeta$ implies that $V_{\mathcal{B}}(z) = 0$ for every $\mathcal{B} \neq \emptyset$.

Note 75. In (5.2), the mapping $\mathcal{V} \to \mathrm{pr}_Z$ is in general non-identifiable because (5.2) can be constructed from a different interaction potential $\tilde{\mathcal{V}} = \{V_{\mathcal{B}} + c : \mathcal{B} \subseteq \mathcal{S}\}$ for any constant c. I.e. $U_{\mathcal{S}}^{\mathcal{V}}(z) = U_{\mathcal{S}}^{\tilde{\mathcal{V}}}(z)$.

Note 76. One way to make $\mathcal{V}$ identifiable is to impose restriction

$$\forall \mathcal{A} \neq \emptyset, \; V_{\mathcal{A}}(z) = 0, \text{ if for some } \; i \in \mathcal{A}, \; z_i = \zeta \tag{5.3}$$

Notation 77. For convenience, consider notation related to $z^{[\mathcal{B},\zeta]}$ such as

$$\left[z^{[\mathcal{B},\zeta]}\right]_i = \begin{cases} \zeta, & \text{if } i \notin \mathcal{B} \\ z_i, & \text{if } i \in \mathcal{B} \end{cases}$$

and $z_{\mathcal{A}}^{[\mathcal{B},\zeta]} = \left(z_s^{[\mathcal{B},\zeta]}; s \in \mathcal{A}\right)$, and $z_s^{[\mathcal{B},\zeta]} = z_{\{s\}}^{[\mathcal{B},\zeta]}$ for some fixed ζ.

Example 78. For instance if $z \in \mathcal{Z}^{\mathcal{S}}$ where $\mathcal{S} = \{1, ..., n\}$ then

$$z^{[\emptyset,\zeta]} = \left(\overbrace{\zeta, ..., \zeta}^{n \text{ times}} \right)^{\top}; \qquad z^{[\{i\},\zeta]} = \left(\zeta, ..., \zeta, \overbrace{z_i}^{i\text{th location}}, \zeta, ..., \zeta \right)^{\top};$$

$$z^{[\{i,j\},\zeta]} = \left(\zeta, ...\zeta, \overbrace{z_i}^{i\text{th location}}, \zeta, ..., \zeta, \overbrace{z_j}^{j\text{th location}}, ..., \zeta \right)^{\top}; \quad z^{[\mathcal{S},\zeta]} = (z_1, ..., z_n)^{\top};$$

Note 79. The following theorem uniquely associates potentials satisfying (5.3) with (5.2) with regards a normalizing point.

Theorem 80. *Let Z be an $\mathcal{Z}$-valued random field on a finite collection $\mathcal{S} \neq \emptyset$ of sites such that $pr_Z(z) > 0$ for all $z \in \mathcal{Z}^{\mathcal{S}}$. Then Z is a Gibbs Random Field with respect to the canonical potential*

$$\begin{aligned} V_{\mathcal{A}}(z_{\mathcal{A}}) &= \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{Card(\mathcal{A} \setminus \mathcal{B})} U_{\mathcal{B}}^{\mathcal{V}}\left(z^{[\mathcal{B},\zeta]}\right), \ z \in \mathcal{Z}^{\mathcal{S}} \\ &= \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{Card(\mathcal{A} \setminus \mathcal{B})} \log\left(pr_Z\left(z^{[\mathcal{B},\zeta]}\right)\right), \ z \in \mathcal{Z}^{\mathcal{S}} \end{aligned} \tag{5.4}$$

where $\zeta \in \mathcal{Z}$ is a fixed value and notation $z^{[\mathcal{B},\zeta]}$ denotes the vector based on $z \in \mathcal{Z}^{\mathcal{S}}$ but modified such that its i-th element is $\left[z^{[\mathcal{B},\zeta]}\right]_i = z_i$ if $i \in \mathcal{B}$ and $\left[z^{[\mathcal{B},\zeta]}\right]_i = \zeta$ if $i \notin \mathcal{B}$. This is the unique normalized potential w.r.t $\zeta \in \mathcal{Z}$.

Proof. The proof is is based on Möbius inversion formula, and hence out of scope. □

Corollary 81. *From Theorem 80, for all $i \in \mathcal{A}$ it is*

$$V_{\mathcal{A}}(z_{\mathcal{A}}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{Card(\mathcal{A} \setminus \mathcal{B})} \log\left(pr_i\left(z_i^{[\mathcal{B},\zeta]} | z_{\mathcal{S}\setminus\{i\}}^{[\mathcal{B},\zeta]}\right)\right), \ z \in \mathcal{Z}^{\mathcal{S}} \tag{5.5}$$

Note 82. The following example explains the use of Theorem 80 in terms of the Definition 64.

Example 83. Consider $\mathcal{S} = \{1, 2\}$. Let $z = (z_1, z_2)^{\top}$. Consider a fixed $\zeta \in \mathcal{Z}$. Then $\mathcal{V} = \{V_{\mathcal{A}} : \mathcal{A} \subseteq \mathcal{S}\} = \left\{V_{\{1\}}, V_{\{2\}}, V_{\{1,2\}}\right\}$. The decomposition of the energy $U\left(z = (z_1, z_2)^{\top}\right) := U_{\mathcal{S}}^{\mathcal{V}}(z)$ is written as

$$U(z_1, z_2) - U(\zeta, \zeta) = V_{\{1\}}(z_1) + V_{\{2\}}(z_2) + V_{\{1,2\}}(z_1, z_2)$$

by using (5.1) with

$$V_{\{1\}}(z_1) = U(z_1, \zeta) - U(\zeta, \zeta)$$
$$V_{\{2\}}(z_2) = U(\zeta, z_2) - U(\zeta, \zeta)$$
$$V_{\{1,2\}}(z_1, z_2) = U(z_1, z_2) - U(z_1, \zeta) - U(\zeta, z_2) + U(\zeta, \zeta)$$

by (5.4).

Example 84. (Ising model) We revisit Example 17 where

$$\mathrm{pr}_Z(z) \propto \exp\left(\alpha \sum_{i \in \mathcal{S}} z_i + \beta \sum_{\{i,j\}: i \sim j} z_i z_j\right), \ \forall z \in \mathcal{Z}^{\mathcal{S}}$$

Consider Notation 77, for instance,

$$z^{[\emptyset,\zeta]} = \left(\overbrace{\zeta, ..., \zeta}^{n \text{ times}}\right)^\top; \qquad z^{[\{i\},\zeta]} = \left(\zeta, ..., \zeta, \overbrace{z_i}^{i\text{th location}}, \zeta, ..., \zeta\right)^\top;$$

$$z^{[\{i,j\},\zeta]} = \left(\zeta, ...\zeta, \overbrace{z_i}^{i\text{th location}}, \zeta, ..., \zeta, \overbrace{z_j}^{j\text{th location}}, ..., \zeta\right)^\top; \quad z^{[\mathcal{S},\zeta]} = (z_1, ..., z_n)^\top;$$

It is $V_\emptyset = 0$ by definition. By using Theorem 80 and considering a reference point $\zeta = 0$, we get

$$V_{\{i\}}(z) = (-1)^{1-1} U\left(z^{[\{i\},\zeta]}\right) + (-1)^{1-0} U\left(z^{[\emptyset,\zeta]}\right) = a z_i, \tag{5.6}$$

for any $i \in \mathcal{S}$ and

$$\begin{aligned} V_{\{i,j\}}(z) &= \left[(-1)^{2-2} U\left(z^{[\{i,j\},\zeta]}\right)\right] + \left[(-1)^{2-1} U\left(z^{[\{i\},\zeta]}\right)\right] \\ &\quad + \left[(-1)^{2-1} U\left(z^{[\{j\},\zeta]}\right)\right] + \left[(-1)^{2-0} U\left(z^{[\emptyset,\zeta]}\right)\right] \\ &= [\alpha z_i + \alpha z_j + \beta z_i z_j] + [-\alpha z_i] + [-\alpha z_j] + [0] = \beta z_i z_j \end{aligned} \tag{5.7}$$

for any $i, j \in \mathcal{S}$, with $i \sim j$. Obviously, it is $V_{\{i,j\}}(z) = 0$ for any $i, j \in \mathcal{S}$, with $i \not\sim j$; and it is $V_{\mathcal{A}}(z) = 0$ for $\mathrm{card}(\mathcal{A}) > 2$.

6. Markov Random Fields

Note 85. Regarding spatial modeling, $\sim$ can describe adjacent sites which is in accordance to the spatial statistics "dogma" that *near things are more related than distant things.* Also it may be computationally convenient for big data problems (large number of sites) as it introduces sparsity and allows specialized numerical algorithms to be implemented.

Note 86. Markov Random Fields constrain the problem such that the conditional distribution of the label at some site i given those at all other sites $j \in \mathcal{S} - \{i\}$ depends only on the labels at neighbors of site i.

Example 87. Recall the Ising model in Example 84 whose sites are equipped with a symmetric relation "$\sim$". It's potentials $V_{\mathcal{A}}$ are non-zero only when $\mathcal{A}$ is a pair of sites $\{i, j\}$ satisfying the relation $\sim$ (5.7) or when $\mathcal{A}$ a singleton (5.6). Consequently, its local characteristics $\mathrm{pr}_i\left(z_i | z_{\mathcal{S}\setminus\{i\}}\right)$ depend only on the values of the sites $j \in \mathcal{S}\setminus\{i\}$ that satisfy $\sim$.

Definition 88. We define as the boundary of $\mathcal{A}$, $\mathcal{A} \subseteq \mathcal{S}$, for a given relation $\sim$ the set

$$\partial\mathcal{A} = \{s \in \mathcal{S}\setminus\mathcal{A} : \exists t \in \mathcal{A} \,\text{s.t.}\, s \sim t\}$$

Definition 89. Let $\partial\mathcal{A}$ be the boundary of $\mathcal{A} \subseteq \mathcal{S}$ for a symmetric relation $\sim$ the finite set $\mathcal{S} \neq \emptyset$. $Z = (Z_s; s \in \mathcal{S})$ is a Markov random field on $\mathcal{S}$ taking values in $\mathcal{Z}$ with respect to the symmetric relation $\sim$ if for each $\mathcal{A} \subset \mathcal{S}$ and $Z_{\mathcal{A}\setminus\mathcal{S}} \in \mathcal{Z}_{\mathcal{A}\setminus\mathcal{S}}$ the distribution of Z on $\mathcal{A}$ conditional on $Z_{\mathcal{A}\setminus\mathcal{S}}$ only depends on $Z_{\partial\mathcal{A}}$ (i.e. the configuration of Z on the neighborhood boundary of $\mathcal{A}$) i.e.

$$\mathrm{pr}_Z\left(z_{\mathcal{A}} | z_{\mathcal{S}\setminus\mathcal{A}}\right) = \mathrm{pr}_Z\left(z_{\mathcal{A}} | z_{\partial\mathcal{A}}\right) \tag{6.1}$$

when $\mathrm{pr}_Z\left(z_{\mathcal{S}\setminus\mathcal{A}}\right) > 0$.

Note 90. Definition 89 implies that (6.1) becomes

$$\mathrm{pr}_Z\left(z_i | z_{-i}\right) = \mathrm{pr}_Z\left(z_i | z_{\partial\{i\}}\right), \quad \forall i \in \mathcal{S} \tag{6.2}$$

when $\mathrm{pr}_Z\left(z_{\mathcal{S}\setminus\{i\}}\right) > 0$

Definition 91. A non-empty subset $\mathcal{C}$, $\mathcal{C} \subset \mathcal{S}$, is a clique in $\mathcal{S}$ with respect to $\sim$ if for all $s, t \in \mathcal{C}$ with $s \neq t$ it is $s \sim t$ or if $\mathcal{C}$ is a singleton set.

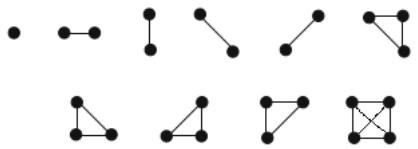


FIGURE 6.1. Examples of cliques

Notation 92. The set containing all the cliques in a lattice of sites in $\mathcal{S}$ equipped with a relation $\sim$ will be usually denoted as bold $\boldsymbol{\mathcal{C}}$.

Note 93. The following theorem shows that the distribution of any Markov random field such that $\mathrm{pr}_Z(z) > 0$ can be expressed in terms of interactions between neighbors.

Theorem 94. *(Hammersley–Clifford) Let $Z = (Z_s; s \in \mathcal{S})$ be an $\mathcal{Z}$-valued random field on a finite collection $\mathcal{S} \neq \emptyset$ of sites such that $pr_Z\left(z_{\mathcal{A}}|z_{\mathcal{C}\setminus\mathcal{A}}\right) > 0$ for all $\mathcal{A} \subset \mathcal{S}$ and $z \in \mathcal{Z}^{\mathcal{S}}$. Let $\sim$ be a symmetric relation on $\mathcal{S}$. Then Z is a Markov Random Field with respect to $\sim$ if and only if*

$$pr_Z(z) \propto \prod_{\mathcal{C}\in\boldsymbol{\mathcal{C}}} \varphi_{\mathcal{C}}(z_{\mathcal{C}}) \tag{6.3}$$

for some interaction functions $\varphi_{\mathcal{C}} : \mathcal{Z}^{\mathcal{C}} \to \mathbb{R}^+$ defined on cliques $\mathcal{C} \in \boldsymbol{\mathcal{C}}$.

Proof. □

For convenience, let $\left[z^{\mathcal{B},\delta}\right]_i = \begin{cases} \delta, & \text{if } i \notin \mathcal{B} \\ z_i, & \text{if } i \in \mathcal{B} \end{cases}$, *and* $z_{\mathcal{A}}^{\mathcal{B},\delta} = \left(z_s^{\mathcal{B},\delta}; s \in \mathcal{A}\right)$, *and* $z_s^{\mathcal{B},\delta} = z_{\{s\}}^{\mathcal{B},\delta}$.

for $\Longrightarrow$: *By Theorem 80, Z is Gibbs with a canonical potential (5.4)*

$$V_{\mathcal{A}}(z_{\mathcal{A}}) = \sum_{\mathcal{A}\subseteq\mathcal{B}} (-1)^{Card(\mathcal{A}\setminus\mathcal{B})} \log\left(pr_Z\left(z^{[\mathcal{B},\zeta]}\right)\right),$$

for $z \in \mathcal{Z}^{\mathcal{S}}$. We need to show that for all $\mathcal{A}$ which are not a cliques, $\mathcal{A} \notin \boldsymbol{\mathcal{C}}$.

Assume a set $\mathcal{A}$ with $\mathcal{A} \subseteq \mathcal{S}$ which is not a clique, $\mathcal{A} \notin \boldsymbol{\mathcal{C}}$, there are two distinct sites $s, t \in \mathcal{A}$ with $s \not\sim t$. Then,

$$\begin{aligned}
V_{\mathcal{A}}(z) =& \sum_{\mathcal{B}\subseteq\mathcal{A}} (-1)^{Card(\mathcal{A}\setminus\mathcal{B})} \log\left(pr_Z\left(z_s^{\mathcal{B},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B},\delta}\right)\right) \\
=& \sum_{\mathcal{B}\subseteq\mathcal{A}\setminus\{s,t\}} (-1)^{Card(\mathcal{A}\setminus\mathcal{B})} \log\left(pr_Z\left(z_s^{\mathcal{B},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B},\delta}\right)\right) \\
&+ \sum_{\mathcal{B}\subseteq\mathcal{A}\setminus\{s,t\}} (-1)^{Card(\mathcal{A}\setminus(\mathcal{B}\cup\{s\}))} \log\left(pr_Z\left(z_s^{\mathcal{B}\cup\{s\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s\},\delta}\right)\right) \\
&+ \sum_{\mathcal{B}\subseteq\mathcal{A}\setminus\{s,t\}} (-1)^{Card(\mathcal{A}\setminus(\mathcal{B}\cup\{t\}))} \log\left(pr_Z\left(z_s^{\mathcal{B}\cup\{t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{t\},\delta}\right)\right) \\
&+ \sum_{\mathcal{B}\subseteq\mathcal{A}\setminus\{s,t\}} (-1)^{Card(\mathcal{A}\setminus(\mathcal{B}\cup\{s,t\}))} \log\left(pr_Z\left(z_s^{\mathcal{B}\cup\{s,t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s,t\},\delta}\right)\right)
\end{aligned}$$

Rearranging I get simplifies

$$V_{\mathcal{A}}(z) = \sum_{\mathcal{B}\subseteq\mathcal{A}} (-1)^{Card(\mathcal{A}\setminus\mathcal{B})} \log\left(\frac{pr_Z\left(z_s^{\mathcal{B},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B},\delta}\right)}{pr_Z\left(z_s^{\mathcal{B}\cup\{t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{t\},\delta}\right)} \frac{pr_Z\left(z_s^{\mathcal{B}\cup\{s,t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s,t\},\delta}\right)}{pr_Z\left(z_s^{\mathcal{B}\cup\{s\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s\},\delta}\right)}\right)$$

Because $s \not\sim t$, it is $pr_Z\left(z_s^{\mathcal{B},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B},\delta}\right) = pr_Z\left(z_s^{\mathcal{B}\cup\{t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{t\},\delta}\right)$ and $pr_Z\left(z_s^{\mathcal{B}\cup\{s,t\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s,t\},\delta}\right) = pr_Z\left(z_s^{\mathcal{B}\cup\{s\},\delta}|z_{\mathcal{S}\setminus s}^{\mathcal{B}\cup\{s\},\delta}\right)$. This implies $V_{\mathcal{A}}(z) = 0$ for any subset $\mathcal{A}$ with $\mathcal{A} \subseteq \mathcal{S}$ which is not a clique. Hence (6.3) holds.

for $\Longleftarrow$: *By using (5.2), I can write*

$$pr_Z\left(z_{\mathcal{A}}|z_{\mathcal{S}\setminus\mathcal{A}}\right)=\frac{1}{C_{\mathcal{A}}\left(z_{\mathcal{S}\setminus\mathcal{A}}\right)}\exp\left(U_{\mathcal{A}}\left(z\right)\right)$$

where

$$U_{\mathcal{A}}\left(z\right)=\sum_{\{\mathcal{C}\subseteq\mathcal{S}\,:\,\mathcal{A}\cap\mathcal{C}\neq\emptyset\}}V_{\mathcal{C}}\left(z_{\mathcal{C}}\right)$$

depends only on $\{z_i : i\in\mathcal{A}\cup\partial\mathcal{A}\}$ *as* $pr_Z\left(\cdot\right)$ *is a Markov Random Field.*

Note 95. Because $\mathrm{pr}_Z\left(z\right)>0$, the Markov Random Field in (6.3) is a Gibbs Random Field as

$$\mathrm{pr}_Z\left(z\right)\propto\exp\left(\sum_{\mathcal{C}\in\boldsymbol{\mathcal{C}}}\log\left(\varphi_{\mathcal{C}}\left(z_{\mathcal{C}}\right)\right)\right)$$

with non-zero interaction potentials restricted to cliques $\mathcal{C}\in\boldsymbol{\mathcal{C}}$.

Note 96. Essentially Theorem 94 gives guidelines on using Markov RF and Gibbs RF that:

for $\Longrightarrow$: we need to show that there exists an interaction potential $\boldsymbol{\varphi}=\{\varphi_{\mathcal{C}}:\mathcal{C}\in\boldsymbol{\mathcal{C}}\}$ defined on the cliques $\boldsymbol{\mathcal{C}}$ such that $\mathrm{pr}_Z\left(\cdot\right)$ is a Gibbs Random Field with iteration potential $\boldsymbol{\varphi}$.

for $\Longleftarrow$: a Gibbs Random Field with potentials $\{\varphi_{\mathcal{C}}:\mathcal{C}\in\boldsymbol{\mathcal{C}}\}$ defined on the cliques $\boldsymbol{\mathcal{C}}$ is a Markov Random Field.

Example 97. (Ising model; Cont. Example 17). The joint PMF of the Ising model in Example 17 is

$$\begin{aligned}\mathrm{pr}\left(z\right)=&\frac{\exp\left(\alpha\sum_{i\in\mathcal{S}}z_i+\beta\sum_{\{i,j\}:i\sim j}z_iz_j\right)}{\sum_{z\in\mathcal{Z}^{\mathcal{S}}}\exp\left(\alpha\sum_{i\in\mathcal{S}}z_i+\beta\sum_{\{i,j\}:i\sim j}z_iz_j\right)}\\=&\frac{1}{\sum_{z\in\mathcal{Z}^{\mathcal{S}}}\exp\left(\alpha\sum_{i\in\mathcal{S}}z_i+\beta\sum_{\{i,j\}:i\sim j}z_iz_j\right)}\prod_{i\in\mathcal{S}}\exp\left(\alpha z_i\right)\prod_{i\in\mathcal{S}}\prod_{j:j\sim i}\exp\left(\beta z_iz_j\right)\end{aligned}$$

This Ising model is a Markov random field (and hence a Gibbs random field) as I can find that

$$\varphi_{\emptyset}=1/\sum_{z\in\mathcal{Z}^{\mathcal{S}}}\exp\left(\alpha\sum_{i\in\mathcal{S}}z_i+\beta\sum_{\{i,j\}:i\sim j}z_iz_j\right)$$

$$\varphi_{\{i\}}\left(z_{\{i\}}\right)=\exp\left(\alpha z_i\right),\qquad\forall i\in\mathcal{S}\tag{6.4}$$

$$\begin{aligned}\varphi_{\{i,j\}}\left(z_{\{i,j\}}\right)&=\exp\left(\beta z_iz_j\right),&&\forall i,j\in\mathcal{S}\quad\text{s.t.}\quad i\sim j\\\varphi_{\{i,j\}}\left(z_{\{i,j\}}\right)&=1,&&\forall i,j\in\mathcal{S}\quad\text{s.t.}\quad i\not\sim j\\\varphi_{\mathcal{A}}\left(z_{\mathcal{A}}\right)&=1,&&\forall\mathcal{A}\subset\mathcal{S}\quad\text{s.t.}\quad\mathrm{card}\left(\mathcal{A}\right)>2\end{aligned}\tag{6.5}$$

where $\{i\}$ and $\{i,j\}$ satisfying $i \sim j$ are cliques. Alternatively, as $\emptyset$ is not a clique if that $\varphi_\emptyset$ is just the constant term which can be absorbed by (6.4) and (6.5) and correspond to cliques.

Part 2. Model building for aerial data & related inference

7. Automodels

Note 98. We introduce a general class of models, the automodels and their special case Besag's automodels, which are associated to the exponential family of distributions and able to represent spatial dependence.

Definition 99. A random variable X taking values in $\mathcal{X}$ follows an exponential family labeled by parameter $\theta \in \Theta$ if the associated PMF/PDF $\mathrm{pr}_X(x|\theta)$ can be expressed in the form

$$\mathrm{pr}_X(x|\theta) = \exp\left(A(\theta)^\top B(x) + C(x) + D(\theta)\right),\ \forall x \in \mathcal{X}$$

where $A(\cdot)$, $B(\cdot)$, $C(\cdot)$, and $D(\cdot)$ are known functions.

7.1. Multi-parameter automodels.

Theorem 100. *Consider Markov random field $Z = (Z_s; s \in \mathcal{S})$ that takes values in $\mathcal{Z}$ on a finite set of points $\mathcal{S}$ and has energy function $U(\cdot)$, Assume that the following assumptions are satisfied with some fixed normalization configuration $\zeta = (\zeta, ..., \zeta)^\top \in \mathcal{Z}^{\mathcal{S}}$:*

(1) In the energy function $U(\cdot)$ the dependence between the sites is pairwise only, i.e.

$$U(z) = \sum_{i \in \mathcal{S}} V_i(z_i) + \sum_{\{\{i,j\} \in \mathcal{S}^2 : i \sim j\}} V_{i,j}(z_i, z_j),\ z \in \mathcal{Z}^{\mathcal{S}}$$

with $V_i(\zeta) = V_{i,j}(z_i, \zeta) = V_{i,j}(\zeta, z_j) = 0$ for all $i, j \in \mathcal{S}$.

(2) For all $i \in \mathcal{S}$, the conditional distributions (characteristics) are such that

$$\log\left(pr_i(z_i|z_{-i})\right) = \left(A_i(z_{-i})\right)^\top B_i(z_i) + C_i(z_i) + D_i(z_{-i}), \tag{7.1}$$

where $A_i(z_{-i}) \in \mathbb{R}^\ell$, $B_i(z_i) \in \mathbb{R}^\ell$, for $\ell \geq 1$ and $C_i(z_i) \in \mathbb{R}$, and $D_i(z_{-i}) \in \mathbb{R}$ with $C_i(\zeta) = 0$ and $B_i(\zeta) = 0$.

(3) For all $i \in \mathcal{S}$, span$\{B_i(z_i); z_i \in \mathcal{Z}\} = \mathbb{R}^\ell$, for $\ell \geq 1$.

Then,

(1) the functions $A_i(z_{-i}) \in \mathbb{R}^\ell$ take the form

$$A_i(z_{-i}) = \alpha_i + \sum_{i \neq j} \beta_{i,j} B_j(z_j),\ i \in \mathcal{S}$$

where $\{\alpha_i; i \in \mathcal{S}\}$ *is a family of* ℓ*-dimensional vectors, and* $\{\beta_{i,j}; i,j \in \mathcal{S}, i \neq j\}$ *is a family of* $\ell \times \ell$ *symmetric matrices, and*

(2) the potentials are given by

$$V_i(z_i) = (\alpha_i)^\top B_i(z_i) + C_i(z_i) \tag{7.2}$$

$$V_{i,j}(z_i, z_j) = (B_i(z_i))^\top \beta_{i,j} B_j(z_j) \tag{7.3}$$

Proof. Omitted, but can be found in

(1) Hardouin, C., & Yao, J. F. (2008). Multi-parameter automodels and their applications. Biometrika, 95(2), 335-349.

(2) Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 192-225.

□

Theorem 101. *Consider Markov random field* $Z = (Z_i; i \in \mathcal{S})$ *that takes values in* $\mathcal{Z}$ *on a finite set of points* $\mathcal{S}$ *and has energy function* $U(\cdot)$*, Assume that the following assumptions are satisfied with some fixed normalization configuration* $\zeta = (\zeta, ..., \zeta)^\top \in \mathcal{Z}^{\mathcal{S}}$ *:*

(1) energy function $U(\cdot)$ *involves only pairwise dependence between the sites, i.e.*

$$U(z) = \sum_{i \in \mathcal{S}} V_i(z_i) + \sum_{\{\{i,j\} \in \mathcal{S}^2 : i \sim j\}} V_{i,j}(z_i, z_j), \ z \in \mathcal{Z}^{\mathcal{S}}$$

with potentials

$$V_i(z_i) = (\alpha_i)^\top B_i(z_i) + C_i(z_i) \tag{7.4}$$

$$V_{i,j}(z_i, z_j) = (B_i(z_i))^\top \beta_{i,j} B_j(z_j) \tag{7.5}$$

and $V_i(\zeta) = V_{i,j}(z_i, \zeta) = V_{i,j}(\zeta, z_j) = 0$ *for all* $i, j \in \mathcal{S}$*.*

(2) energy function $U(\cdot)$ *is admissible; i.e.*

$$\int \exp(U(z))\, dz < \infty$$

Then,

(1) the family of conditional distributions $pr_i(z_i | z_{-i})$ *belongs to a multiparameter exponential family distributions such as*

$$\log(pr_i(z_i | z_{-i})) = (A_i(z_{-i}))^\top B_i(z_i) + C_i(z_i) + D_i(z_{-i}), \tag{7.6}$$

whose natural parameters $A_i(z_{-i}) \in \mathbb{R}^\ell$ *take the form*

$$A_i(z_{-i}) = \alpha_i + \sum_{i \neq j} \beta_{i,j} B_j(z_j), \ i \in \mathcal{S}$$

where $\{\alpha_i; i \in \mathcal{S}\}$ *is a family of* ℓ*-dimensional vectors, and* $\{\beta_{i,j}; i,j \in \mathcal{S}, i \neq j\}$ *is a family of* $\ell \times \ell$ *symmetric matrices.*

Proof. Omitted, but can be found in

(1) Hardouin, C., & Yao, J. F. (2008). Multi-parameter automodels and their applications. Biometrika, 95(2), 335-349.

(2) Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 192-225.

□

Definition 102. Automodel is called the model satisfying the assumptions of Theorem 100.

Definition 103. Univariate automodel is the automodel with $\ell = 1$ in Theorem 100.

Definition 104. Multi-parameter is the automodel with $\ell > 1$ in Theorem 100.

Remark 105. In the univariate automodel, $\ell = 1$, assumption 3 in Theorem 100 is not needed; it is automatically satisfied as B_i's are not identically zero. Yet, for $\ell = 1$, (7.2) and (7.3) become

$$V_i(z_i) = \alpha_i B_i(z_i) + C_i(z_i) \tag{7.7}$$

$$V_{i,j}(z_i, z_j) = \beta_{i,j} B_i(z_i) B_j(z_j) \tag{7.8}$$

7.2. **Besag auto-models.**

Definition. Random field $Z = (Z_s; s \in \mathcal{S})$ follows a Besag's auto-model if Z is real-valued and its joint distribution $\mathrm{pr}_Z(z)$ is given by

$$\mathrm{pr}_Z(z) = \frac{1}{C} \exp\left(\sum_{i \in \mathcal{S}} V_i(z_i) + \sum_{\{i,j\} \in \mathcal{S}^2 : i \sim j} \beta_{i,j} z_i z_j \right),\ z \in \mathcal{Z}^{\mathcal{S}} \tag{7.9}$$

with $\beta_{i,j} = \beta_{j,i}$ for all $i, j \in \mathcal{S}$.

Note 106. The following allows us to define a Markov Random Field model from a set of conditional distributions (characteristics) whose compatibility is automatically satisfied.

Proposition 107. *If each of the*

$$pr_i(z_i | z_{-i}), \quad for \quad i \in \mathcal{S}$$

is a family of real-valued $z_i \in \mathbb{R}$ *conditional distributions which are members of the exponential family of distributions (7.1) with* $B_i(z_i) = z_i$ *for* $i \in \mathcal{S}$*, then they are compatible a Besag's auto-model with distribution (7.9) if* $\beta_{i,j} = \beta_{j,i}$ *for all* $i, j \in \mathcal{S}$*.*

Proof. For

$$\text{pr}_i\left(z_i|z_{-i}\right)=\exp\left(A_i\left(z_{-i}\right)z_i+C_i\left(z_i\right)+D_i\left(z_{-i}\right)\right)$$

it is

$$\begin{aligned}V_i\left(z_i\right)=&\alpha_i B_i\left(z_i\right)+C_i\left(z_i\right)=\alpha_i z_i+C_i\left(z_i\right)\\V_{i,j}\left(z_i,z_j\right)=&\beta_{i,j}B_i\left(z_i\right)B_j\left(z_j\right)=\beta_{i,j}z_i z_j\end{aligned}$$

so

$$\text{pr}_Z\left(z\right)\propto\exp\left(\sum_i\left[\alpha_i z_i+C_i\left(z_i\right)\right]+\sum_{i\sim j}\beta_{i,j}z_i z_j\right),\ z\in\mathcal{Z}^{\mathcal{S}}$$

□

Example 108. (Logistic automodel / Ising model) Consider that $Z\left(s\right)$ represents presence or absence of a characteristic at location $s\in\mathcal{S}$. Mathematically, assume random field Z taking values on a set of indices $\mathcal{S}$ in $\mathcal{Z}=\{0,1\}$ on $\mathcal{S}=\{1,...,n\}$, $n\in\mathbb{N}-\{0\}$.

Consider that for a given z_{-i} it is

$$z_i|z_{-i}\sim\text{Logit}\left(\theta_i\left(z_{-i}\right)\right),\quad i\in\mathcal{S}.$$

Hint:: The PMF of distribution $x|\theta\sim\text{Logit}\left(\theta\right)$ can be written as $\text{pr}\left(x|\theta\right)=\frac{\exp(x\theta)}{1+\exp(\theta)}1\left(x\in\{0,1\}\right)$. Then the characteristics are

$$\text{pr}_i\left(z_i|z_{-i}\right)=\frac{\exp\left(z_i\theta_i\left(z_{-i}\right)\right)}{1+\exp\left(\theta_i\left(z_{-i}\right)\right)}1\left(z_i\in\{0,1\}\right)\tag{7.10}$$

Now, let's parameterize $\{\theta_i\left(\cdot\right)\}$ as

$$\theta_i\left(z_{-i}\right)=\alpha_i+\sum_{j:j\sim i}\beta_{i,j}z_j\tag{7.11}$$

for $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ with $\beta_{i,j}=\beta_{j,i}$. Then (7.10) becomes

(7.12)

$$\log\left(\text{pr}_i\left(z_i|z_{-i}\right)\right)=\underbrace{z_i}_{B_i(z_i)}\underbrace{\left(\alpha_i+\sum_{j\sim i}\beta_{i,j}\overbrace{z_j}^{B_i(z_j)}\right)}_{A_i(z_{-i})}+\underbrace{0}_{C_i(z_i)}+\underbrace{\left(-\log\left(1+\exp\left(\alpha_i+\sum_{j:j\sim i}\beta_{i,j}z_j\right)\right)\right)}_{D_i(z_{-i})}$$

Notice that all the conditionals $z_i|z_{-i}$ follow an Exponential family with

$$\begin{aligned} A_i\left(z_{-i}\right) =& \alpha_i + \sum_{j:j\sim i} \beta_{i,j} B_i\left(z_j\right) \\ B_i\left(z_i\right) =& z_i \\ C_i\left(z_i\right) =& 0 \\ D_i\left(z_{-i}\right) =& -\log\left(1+\exp\left(\alpha_i + \sum_{j:j\sim i}\beta_{i,j} z_j\right)\right) \end{aligned}$$

Also, I can get $C_i(\zeta) = 0$ and $B_i(\zeta) = 0$ by considering a reference point $\zeta = 0$. From Theorem 107, (7.10) with (7.11), the conditionals $z_i|z_{-i}$ are compatible as a Besag automodel with marginal distribution

$$\text{pr}_Z(z) \propto \exp\left(\overbrace{\sum_i \alpha_i \underbrace{z_i}_{B_i(z_i)} + \sum_i \underbrace{\sum_{j:j\sim i}}_{\sum_{\{i,j\}:j\sim i}} \beta_{i,j} z_i z_j}^{U(z)=}\right) \tag{7.13}$$

I observe that:

- Here the Ising model has spatially dependent coefficients $\{\alpha_i\}$ and $\{\beta_{i,j}\}$, unlike the Ising model in Example 17 where we considered $\{\alpha_i = \alpha\}$ and $\{\beta_{i,j} = \beta\}$.
- When $\beta_{i,j} = 0$, for all j such as $j \sim i$, it is $\text{pr}_i\left(z_i|z_{-i}\right) = \frac{\exp(z_i\alpha_i)}{1+\exp(\alpha_i)}$.
- Characteristic's present at site i is encouraged in neighboring site j when $\beta_{i,j} > 0$, and discouraged when $\beta_{i,j} < 0$.

The resulting spatial model is called Logistic automodel or Ising model (the latter name is from physics).

Example 109. (Poisson automodel) Consider that $Z(s)$ represents counts at location $s \in \mathcal{S}$. Mathematically we can consider Z taking values in $\mathcal{Z} = \mathbb{N}$ on a set of sites $\mathcal{S} = \{1, ..., n\}$, where $n \in \mathbb{N} - \{0\}$.

Consider that for a given z_{-i} it is

$$z_i|z_{-i} \sim \text{Poisson}\left(\lambda_i\left(z_{-i}\right)\right)$$

Hint:: The PMF of Poisson distribution $x|\lambda \sim \text{Poisson}(\lambda)$ can be written as

$$\text{pr}\left(x|\lambda\right) = \frac{1}{x!}\lambda^x \exp\left(-\lambda\right) 1\left(x \in \mathbb{N}\right)$$

with mean $\text{E}\left(x|\lambda\right) = \lambda$.

Then the full conditionals (characteristics) are

$$\text{pr}_i\left(z_i|z_{-i}\right)=\frac{1}{z_i!}\left(\lambda_i\left(z_{-i}\right)\right)^{z_i}\exp\left(-\lambda\right)1\left(z_i\in\mathbb{N}\right) \tag{7.14}$$

Now, let's parameterize $\{\lambda_i(\cdot)\}$ as

$$\log\left(\lambda_i\left(z_{-i}\right)\right)=\alpha_i+\sum_{j:j\sim i}\beta_{i,j}z_j \tag{7.15}$$

for $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ with $\beta_{i,j}=\beta_{j,i}$. So (7.14) becomes

$$\log\left(\text{pr}_i\left(z_i|z_{-i}\right)\right)=\underbrace{z_i}_{B_i(z_i)}\underbrace{\left(\alpha_i+\sum_{j\sim i}\beta_{i,j}\overbrace{z_j}^{B_i(z_j)}\right)}_{A_i(z_{-i})}+\underbrace{\log\left(z_i!\right)}_{C_i(z_i)}+\underbrace{0}_{D_i(z_{-i})}$$

with

$$\begin{aligned}A_i\left(z_i\right)=&\alpha_i+\sum_{j\sim i}\beta_{i,j}B_i\left(z_j\right)\\B_i\left(z_{-i}\right)=&z_i\\C_i\left(z_i\right)=&\log\left(z_i!\right)\\D_i\left(z_{-i}\right)=&0\end{aligned}$$

I can notice that all the conditionals $z_i|z_{-i}$ follow exponential of exponential. Also, I can get $C_i(\zeta)=0$ and $B_i(\zeta)=0$ by considering a reference point $\zeta=0$. From Theorem 107, (7.14) with (7.15), the the conditionals $z_i|z_{-i}$ are compatible as a Besag auto-model with marginal distribution

$$\text{pr}_Z\left(z\right)\propto\exp\left(\overbrace{\sum_i\left(\underbrace{\alpha_i\overbrace{z_i}^{B_i(z_j)}+\overbrace{\log\left(z_i!\right)}^{C_i(z_i)}}_{V_i(z_i)}\right)+\sum_i\sum_{j:j\sim i}\beta_{i,j}z_iz_j}^{U(z)=}\right)$$

or otherwise the energy function is

$$U\left(z\right)=\sum_i\left(\alpha_iz_i+\log\left(z_i!\right)\right)+\sum_{j\sim i}\beta_{i,j}z_iz_j$$

Furthermore, to ensure that $U(z)$ is admissible, we need to consider additional conditions. I observe that

$$\sum_{z\in\mathbb{N}^S}\exp\left(U\left(z\right)\right)=\sum_{z\in\mathbb{N}^S}\prod_i\left(\exp\left(\alpha_iz_i+\log\left(z_i!\right)\right)+\sum_{j\sim i}\beta_{i,j}z_iz_j\right)$$

- If we use additional condition $\beta_{i,j} \leq 0$ then

$$\sum_{z \in \mathbb{N}^{\mathcal{S}}} \exp\left(U\left(z\right)\right) \leq \sum_{z \in \mathbb{N}^{\mathcal{S}}} \prod_{i} \left(\exp\left(\alpha_i z_i + \log\left(z_i!\right)\right)\right) = \sum_{z \in \mathbb{N}^{\mathcal{S}}} \prod_{i} \frac{1}{z_i!} \exp\left(\alpha_i z_i\right) < \infty$$

  which converges. Modeling-wise, $\beta_{i,j} < 0$ introduces competition among the neighbors similar to the Ising model. So by introducing a competition such as $\beta_{i,j} \leq 0$ in the model I prevent the count z_i at i to explode.
- If $\beta_{i,j} > 0$, I discourage competition among neighboring sites. Admissibility can be satisfied if we truncate the state space as $z_i < M$ for some fixed upper bound M. For instance, the characteristics $z_i | z_{-i}$ can follow a Poisson distribution truncated at M.

$$\mathrm{pr}_i\left(z_i | z_{-i}\right) = \frac{1}{z_i!} \left(\lambda_i\left(z_{-i}\right)\right)^{z_i} \exp\left(-\lambda\right) 1\left(z_i \in \{0, 1, ..., M\}\right)$$

  So I can prevent z_i at i to explode by forcefully bounding it $z_i < M$ with a big enough value $M > 0$.

The resulting spatial model is called Poisson automodel.

Note 110. A CAR model is an automodel. Recall that CAR model is defined such as its local characteristics (full conditional distributions) are Gaussian distributions; however Gaussian distribution is an exponential distribution family. Hence the joint distribution of CAR model in Proposition 56 could have been derived from Theorem 100 as well.

7.3. **Parameterization matters in automodels.**

Remark 111. The unknown parameter vector $\theta = \left(\left(\alpha_i; i \in \mathcal{S}\right), \left(\beta_{i,j}; i, j \in \mathcal{S}\right)\right)$ in automodels (e.g., Besag's automodel (7.9)) can be further parameterised to have a particular structure without the need to consider any additional constrains in Theorems 100 & 101.

Remark 112. The dimensionality of automodel parameters $\theta = \left(\left(\alpha_i; i \in \mathcal{S}\right), \left(\beta_{i,j}; i, j \in \mathcal{S}\right)\right)$ may be too large leading to an over-parameterized model or prohibitively large computational cost when the size of the set of sites $\mathcal{S}$ is large (a usual case). To mitigate this issue, a way is to set a structure on $\left\{\alpha_i, \beta_{i,j}\right\}_{i,j \in \mathcal{S},}$ reducing its dimentionality.

E.g.: by setting

$$\alpha_i = a w_i, \quad \text{and} \quad \beta_{i,j} = b_i c_j; \text{ for } i, j \in \mathcal{S},$$

with some known weights $\{w_i;\ i \in \mathcal{S},\}$ and unknown $\{a, b_i, c_j;\ i, j \in \mathcal{S}\}$. Then learning $\mathrm{Card}\left(\mathcal{S}\right)\left(1 + \mathrm{Card}\left(\mathcal{S}\right)\right)$ unknown parameter $\{\alpha_i, \beta_{i,j};\ i, j \in \mathcal{S},\}$ reduces to learning just $1 + 2\mathrm{Card}\left(\mathcal{S}\right)$ unknown parameters $\{a, b_i, c_j;\ i, j \in \mathcal{S}\}$. Note, that $\beta_{i,j} = b_i c_j$ restricts the interaction between i, j.

Remark 113. When covariates $x_i = (x_{i,1}, ..., x_{i,p})^\top$ for $i \in \mathcal{S}$ are available (such as other characteristics or time), one could "link" them to the model via the parameters $\{\alpha_i, \beta_{i,j}\}_{i,j\in\mathcal{S}}$, For instance

E.g.: by setting

$$\alpha_i = a_i + \sum_{k=1}^{p} d_k x_{i,k}, \quad \text{and} \quad \beta_{i,j} = \beta_{i,j};\ \text{for } i,j \in \mathcal{S}, \tag{7.16}$$

where $\{a_i;\ i \in \mathcal{S}\}$, $\{d_k; k = 1, ..., k\}$ and $\{\beta_{i,j};\ i,j \in \mathcal{S}\}$ are unknown parameters. d_k represents the influence of k-th covariate $x_{i,k}$, for all $i \in \mathcal{S}$. $\beta_{i,j}$ represents the influence of the $z_{\partial i}$ at the neighboring sites of Z_i. Examination of the sign of $\beta_{i,j}$, and d_k or whether $\beta_{i,j} \neq 0$, $d_k \neq 0$ facilitates the discovery of pasterns and conditional dependencies.

E.g.: it t denotes time, we can make the automodel "dynamic" (aka spatio-temporal) by setting $x_i = (t_i, t_i^2)^\top$ for $i \in \mathcal{S}$ and

$$\alpha_i = a_i + d_1 t_i + d_2 (t_i)^2, \quad \text{and} \quad \beta_{i,j} = \beta_{i,j};\ \text{for } i,j \in \mathcal{S}.$$

Example 114. In Example 109, given observable covariates $x_i = (x_{i,1}, ..., x_{i,p})^\top$ for $i \in \mathcal{S}$, one may set (7.15) as

$$\log\left(\lambda_i\left(z_{-i}\right)\right) = \left[a_i + \sum_{k=1}^{p} d_k x_{i,k}\right] + \left[\sum_{j:j\sim i} \beta_{i,j} z_j\right] \tag{7.17}$$

Then d_k represent the influence of k-th covariate $x_{i,k}$, for all $i \in \mathcal{S}$, and $\beta_{i,j}$ represents the influence of the $z_{\partial i}$ at the neighboring sites of Z_i. For admissibility, a condition such as $\beta_{i,j} \leq 0$ should be specified (see Example 109). Further restrictions on the unknown parameters, or dimension reduction techniques, should be used because the number of unknowns is greater than the number of observations in (7.17).

Example 115. In Example 109, if the dataset is $\{(t_i, s_i, Z_i)\,; i \in \mathcal{S}\}$ where Z_i is the measurement (e.g. counts of a characteristic), at time t_i, at location $s_i \in \mathbb{R}^2$ of the i-th observation, a researcher may consider a parametrization

$$\log\left(\lambda_i\left(z_{-i}, t_i\right)\right) = \left[a_i + d_1 t_i\right] + \left[\sum_{j:j\sim i} \beta_{i,j} z_j\right] \tag{7.18}$$

and be interested in learning the unknown parameters $\{a_i\}$, d_1, and $\{\beta_{i,j}\}$. Obviously, the resulted model is space-time.

8. Frequentist modeling and likelihood based inference

Note 116. Consider a dataset $\{(s_i, Z_i = Z(s_i)); i = 1, ..., n\}$ where Z_i is the observation at site s_i for $i = 1, ..., n$. Assume that the the sampling distribution of $(Z_i)_{i=1}^n$ is specified by the researcher to be

$$Z \sim \text{pr}_Z(Z|\theta) \tag{8.1}$$

labeled by unknown parameter vecrtor θ. Parametric and predictive inference can be performed based based on the associated likelihood or its approximation PsuedoLikelihood.

Note 117. For easy of presentation we assume that the observables Z are a realization of an automodel and hence their sampling distribution (8.1) is that of an automodel with potentials 7.2 and 7.3, and unknown parameter $\theta = (\{\alpha_i\}, \{\beta_{i,j}\})^\top$.

8.1. **MLE: Maximum likelihood estimation.**

Note 118. We describe the maximum likelihood estimation in the automodel framework.

Remark 119. In the MLE framework, given a dataset $\{(s_i, Z_i = Z(s_i)); i = 1, ..., n\}$, estimation of the unknown parameters $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ of an automodel can be performed by maximizing the likelihood, as

$$\left(\{\hat{\alpha}_i\}, \left\{\hat{\beta}_{i,j}\right\}\right) = \underset{\{\alpha_i\},\{\beta_{i,j}\}}{\arg\max} \left(\text{pr}_Z(Z|\{\alpha_i\}, \{\beta_{i,j}\})\right) \tag{8.2}$$

$$\text{subject to } \beta_{i,j} = \beta_{j,i},\ \forall i, j \in \mathcal{S}$$

$$\text{...and any other problem specific restrictions}$$

where $\text{pr}_Z(Z|\{\alpha_i\}, \{\beta_{i,j}\})$ is the joint distribution (7.9) given the unknown parameters $\{\alpha_i\}$ and $\{\beta_{i,j}\}$.

Example 120. (Logistic automodel / Ising model) Assume that observables Z follow the Logistic automodel (7.13) in Example 108. Computing MLE $\{\hat{\alpha}_i\}, \left\{\hat{\beta}_{i,j}\right\}$ of $\{\alpha_i\}, \{\beta_{i,j}\}$ requires

$$\begin{aligned}\left(\{\hat{\alpha}_i\}, \left\{\hat{\beta}_{i,j}\right\}\right) &= \underset{\{\alpha_i\},\{\beta_{i,j}\}}{\arg\max} \left(\log\left(\text{pr}_Z(Z|\{\alpha_i\}, \{\beta_{i,j}\})\right)\right) \\ &= \underset{\{\alpha_i\},\{\beta_{i,j}\}}{\arg\max} \left(\sum_i \alpha_i z_i + \sum_{\{i,j\}:j\sim i} \beta_{i,j} z_i z_j - \log\left(C(\{\alpha_i\}, \{\beta_{i,j}\})\right)\right) \end{aligned} \tag{8.3}$$

where

$$C(\{\alpha_i\}, \{\beta_{i,j}\}) = \sum_{\forall z \in \mathcal{Z}^{\mathcal{S}}} \exp\left(\sum_i \alpha_i z_i + \sum_{\{i,j\}:j\sim i} \beta_{i,j} z_i z_j\right) \tag{8.4}$$

is the normalizing constant. Optimization in (8.3) can be done numerically by using a recursive optimization algorithm such as Newton-Raphson.

Note 121. The optimization problem (8.2) can be too computationally expensive. For instance, in Example 120, a recursive optimization algorithm, like Newton-Raphson, requires several iterations. At each iteration the evaluation of the (parameter dependent) constant (8.4) has to be evaluated. A computation of that constant can be too expensive when the set of sites $i \in \mathcal{S}$ is large because the sum $\sum_{\forall z \in \mathcal{Z}^{\mathcal{S}}}$ in (8.4) implies scanning all the possible configurations of $z \in \mathcal{Z}^{\mathcal{S}}$. A way to mitigate this is to use instead an "approximation" of the likelihood, such as the Pseudo-likelihood.

8.2. **MPLE: Maximum pseudo likelihood estimation.**

Note 122. We describe the maximum pseudo likelihood estimation in the automodel framework.

Definition. The pseudo likelihood pseudo$L(Z;\theta)$ of observables $Z = (Z_1, ..., Z_n)^\top$ given parameters θ is an approximation of the (exact) likelihood $L(Z;\theta)$ of observables $Z = (Z_1, ..., Z_n)^\top$ given parameters θ which is equal to

$$\text{pseudo}L(Z;\theta) = \prod_i \text{pr}(Z_i|Z_{-i},\theta)$$

where $\text{pr}(Z_i|Z_{-i},\theta)$ are the conditionals of the joint pdf/pmf of the sampling distribution $\text{pr}(Z|\theta)$ of Z given parameter θ.

Definition. (Maximum PseudoLikelihood Estimator) The Maximum Pseudo-Likelihood Estimator (MPLE) $\tilde{\theta}$ of θ is the maximizer of the pseudo likelihood function pseudo$L(Z;\theta)$ where the parameter θ is the argument and the observables $Z = (Z_1, ..., Z_n)^\top$ are fixed values.

$$\tilde{\theta} = \arg\max_\theta \left(\text{pseudo}L(Z;\theta)\right)$$

Remark 123. Then (8.2) becomes: In the MPLE framework, given a dataset $\{(s_i, Z_i = Z(s_i)); i = 1, ..., n\}$, e
of the unknown parameters $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ of an automodel can be performed by maximizing the pseudo-likelihood, as

$$\left(\{\hat{\alpha}_i\}, \left\{\hat{\beta}_{i,j}\right\}\right) = \arg\max_{\{\alpha_i\},\{\beta_{i,j}\}} \left(\prod_{i\in\mathcal{S}} \text{pr}_Z(Z_i|Z_{-i},\theta)\right) \tag{8.5}$$

$$= \arg\max_{\{\alpha_i\},\{\beta_{i,j}\}} \left(\sum_{i\in\mathcal{S}} \log\left(\text{pr}_Z(Z_i|Z_{-i},\theta)\right)\right) \tag{8.6}$$

$$\text{subject to } \beta_{i,j} = \beta_{j,i},\ \forall i,j \in \mathcal{S}$$

...and any other problem specific restrictions

Example 124. (Logistic automodel / Ising model) (Cont. Example 120) Assume that observables Z follow the Logistic automodel (7.13) in Example 108. From (7.12), the conditionals (local characteristics) are computed to be such as

$$\log\left(\mathrm{pr}_i\left(z_i|z_{-i}\right)\right) = z_i\left(\alpha_i + \sum_{j\sim i}\beta_{i,j}z_j\right) - \log\left(1+\exp\left(\alpha_i+\beta_{i,j}\sum_{j:j\sim i}\beta_{i,j}z_j\right)\right)$$

and hence

$$\begin{aligned}\left(\left\{\hat{\alpha}_i\right\},\left\{\hat{\beta}_{i,j}\right\}\right) &= \underset{\{\alpha_i\},\{\beta_{i,j}\}}{\arg\max}\left(\sum_{i\in\mathcal{S}}\log\left(\mathrm{pr}_i\left(z_i|z_{-i}\right)\right)\right)\\ &= \underset{\{\alpha_i\},\{\beta_{i,j}\}}{\arg\max}\left(\sum_{i\in\mathcal{S}}z_i\left(\alpha_i+\sum_{j\sim i}\beta_{i,j}z_j\right) - \sum_{i\in\mathcal{S}}\log\left(1+\exp\left(\alpha_i+\beta_{i,j}\sum_{j:j\sim i}\beta_{i,j}z_j\right)\right)\right)\end{aligned}$$

which does not depend on the normalizing constant (8.4) and hence its computation is less computationally demanding.

9. Hierarchical modeling (Bayesian modeling)

9.1. **A general framework for the hierarchical modeling.**

Note 125. Uncertainty in spatial statistics can be decomposed as a Bayesian hierarchical spatial model

$$\begin{cases} Z|Y,\vartheta_1,\vartheta_2 & \text{data model}\\ Y|\vartheta_1,\vartheta_2 & \text{spatial process model}\\ \vartheta_1|\vartheta_2 & \text{hyper-parameter prior model}\end{cases} \tag{9.1}$$

where uncertainty is described by

$$\mathrm{pr}\left(Z,Y,\vartheta_1|\vartheta_2\right) = \mathrm{pr}\left(Z|Y,\vartheta_1|\vartheta_2\right)\mathrm{pr}\left(Y|\vartheta_1,\vartheta_2\right)\mathrm{pr}\left(\vartheta_1|\vartheta_2\right).$$

Let $\vartheta = \left(\vartheta_1,\vartheta_2\right)^\top$ be unknown hyper-parameters. Let ϑ_1 and ϑ_2 be unknown random and fixed hyper-parameters.

Data model: expresses the measurement uncertainty as it is quantified via the distribution $\mathrm{pr}\left(Z|Y,\vartheta\right)$ possibly labeled by some parameter ϑ. It is often specified/modeled so that it can measure the goodness of fit between Z and Y.

Spatial process model: expresses the scientific uncertainty (e.g., that coming from (Y_s)) as it is quantified via the specified distribution $\mathrm{pr}\left(Y|\vartheta\right)$ possibly labeled by some parameter ϑ. It is often specified/modeled with purpose (among others) to encourage spatial coherence and represente spatial dependence.

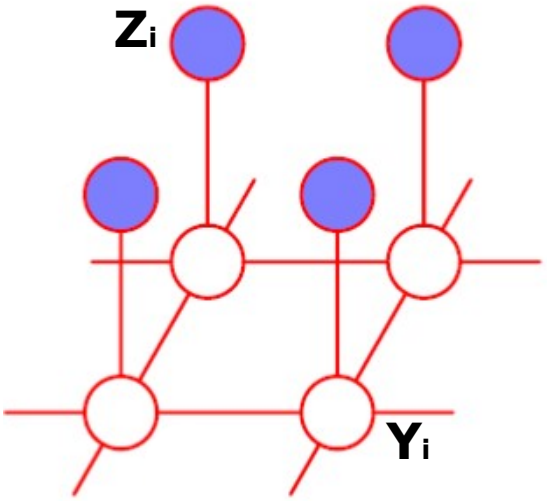


FIGURE 9.1. Hierarchical spatial model structure. $\{Y_i\}$ is the spatial process model which is hidden. $\{Z_i\}$ is the data model. The cartoon depicts a hierarchical spatial model with the special conditional independence structure $Z_i|\{Y_i\},\vartheta \sim \prod_i \text{pr}(Z_i|Y_i,\vartheta)$ and $Y|\vartheta \sim \text{pr}(Y|\vartheta)$

Hyper-parameter prior model: expresses uncertainty about specific unknown model hyper-parameters

See for example Figure 9.1

Note 126. Fixed ϑ_2 can be learned pointwise by computing the ML-II point estimator

$$\hat{\vartheta}_2 = \arg\min_{\vartheta_2} \left(-2\log\left(\text{pr}\left(Z|\vartheta_2\right)\right)\right) \tag{9.2}$$

as the maximizer of the marginal likelihood

$$\text{pr}\left(Z|\vartheta_2\right) = \int \text{pr}\left(Z,Y,\vartheta_1|\vartheta_2\right) \text{d}Y \text{d}\vartheta_1$$

or by computing the pseudo ML-II point estimator

$$\tilde{\vartheta}_2 = \arg\min_{\vartheta_2} \left(-2\log\left(\prod_i \text{pr}\left(Z_i|Z_{-i},\vartheta_2\right)\right)\right) \tag{9.3}$$

as the maximizer of the pseudo marginal likelihood

$$\text{pseudo}L\left(Z|\vartheta_2\right) = \prod_i \text{pr}\left(Z_i|Z_{-i},\vartheta_2\right)$$

$\tilde{\vartheta}_2$ in (9.3) is a computationally cheaper approximation of the MLE $\hat{\vartheta}_2$ in 9.2.

Note 127. Random ϑ_1 can be learned by computing the posterior pdf/pmf of ϑ_1 given Y and $\vartheta_2 = \hat{\vartheta}_2$

$$\text{pr}\left(\vartheta_1|Z, \vartheta_2 = \hat{\vartheta}_2\right) = \frac{\text{pr}\left(Z|\vartheta_1, \vartheta_2 = \hat{\vartheta}_2\right)\text{pr}\left(\vartheta_1|\vartheta_2 = \hat{\vartheta}_2\right)}{\text{pr}\left(Z|\vartheta_2 = \hat{\vartheta}_2\right)}$$

where the value $\hat{\vartheta}_2$ (or $\tilde{\vartheta}_2$) is plugged in.

Note 128. General interest lies in computing the posterior distributions of the spatial process model $(Y_i; i \in \mathcal{S})$, (or latent process, or noiseless process) given the data Z

$$\text{pr}\left(Y|Z, \vartheta_2 = \hat{\vartheta}_2\right) = \int \text{pr}\left(Y|Z, \vartheta_2 = \hat{\vartheta}_2\right)\text{pr}\left(\vartheta_1|Z, \vartheta_2 = \hat{\vartheta}_2\right)\text{d}\vartheta_1$$

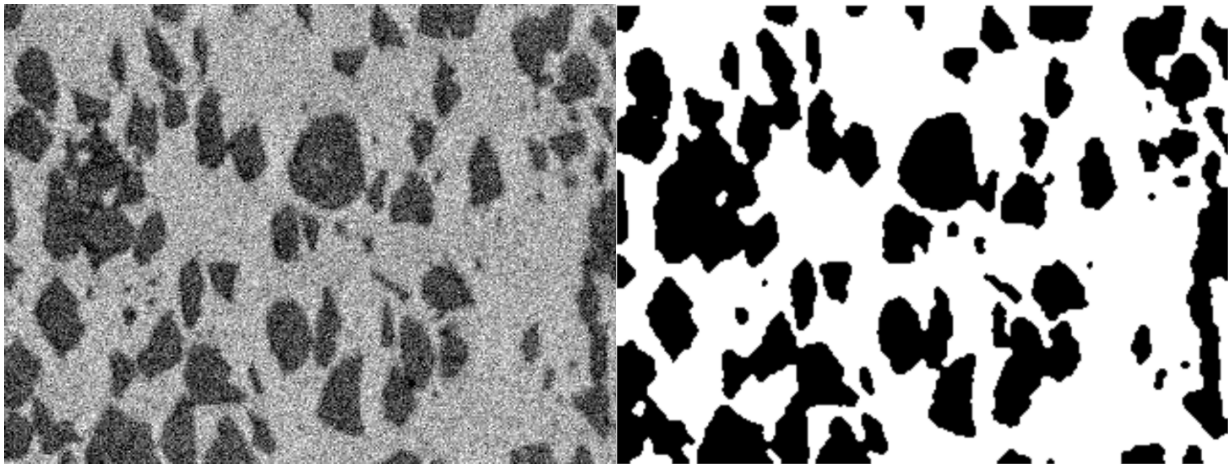
Note 129. Below we give two examples in aerial data.

9.2. **Examples.**

9.2.1. *A simplified spatial model for binary data (e.g. Image denoising).*

Example 130. (Image denoising) A central aim in image processing is to reconstruct an object (e.g. image) $Y = (Y_i; i \in \mathcal{S})$ based on a measurement (observation) $Z = (Z; i \in \mathcal{S})$ which is contaminated by errors $\varepsilon = (\varepsilon_i; i \in \mathcal{S})$. The framework of hierarchical modeling for aerial spatial data is suitable to address such cases.

Consider the image restoration dataset in Example 25 in Handout 1: Types of spatial data. (Figure 9.2a) We have a black and white noisy image with size 240×320 pixels.



(A) Noised image (B) Reconstructed image

FIGURE 9.2. Ferrite-Pearlite steel image (Image restoration)

Mathematically, denote $(Z_i)_{i\in\mathcal{S}}$ as the error contaminated (observed) image. The observables are coded as $Z_i = 1$ for black and $Z_i = 0$ for white at site $i \in \mathcal{S} = \{1, ..., 240 \times 320\}$.

Let $n = \text{Card}(\mathcal{S})$. Hence $(Z_i)_{i\in\mathcal{S}}$ is a realization from the data model. The aim is to recover/learn the unknown real (error free) image $(Y_i)_{i\in\mathcal{S}}$ given the measurement/observation $(Z_i)_{i\in\mathcal{S}}$.

The data model can be specified (for instance) by "assuming" that the observation Z_i has been contaminated by iid noise with some "probability" p for all pixels $i \in \mathcal{S}$; i.e. $p = \text{pr}(\{Z_i \neq Y_i\}|p) = 1 - \text{pr}(\{Z_i = Y_i\})$ for all $i \in \mathcal{S}$. Hence

$$\text{pr}(Z_i|Y_i, p) = p^{1-1(\{Z_i=Y_i\})}(1-p)^{1(\{Z_i=Y_i\})},\ i \in \mathcal{S}$$

Consequently, the data model is

$$\begin{aligned}\text{pr}(Z|Y, p) &= \prod_{i=1}^{n} p^{1-1(\{Z_i=Y_i\})}(1-p)^{1(\{Z_i=Y_i\})} = p^{n_{(Z,Y)}}(1-p)^{n-n_{(Z,Y)}} \\ &= \exp\left(n_{(Z,Y)}\log\left(\frac{p}{1-p}\right) + (1-p)^n\right)\end{aligned}$$

where $n_{(Z,Y)} = \sum_{i\in\mathcal{S}} 1(\{Z_i = Y_i\})$.

The spatial process $(Y_i)_{i\in\mathcal{S}}$ is unknown (unobserved), and, according the Bayesian paradigm, we need to specify a prior process on $(Y_i)_{i\in\mathcal{S}}$ account for the uncertainty. To introduce spatial dependence, the researcher may judge to specify (for example) an Logistic automodel (Ising model) process prior such as

$$\text{pr}(Y|\alpha, \beta) \propto \exp\left(\alpha\sum_{i\in\mathcal{S}} Y_i + \beta\sum_{\{i,j\}:i\sim j} Y_iY_j\right),\ \{0,1\}^{\mathcal{S}}$$

with symmetric relation $i \sim j$ considering only the adjacent pixels.

The researcher may be uncertain about the "real" value of p and hence he/she may want to specify a conjugate Beta prior[2] $p \sim \text{Be}(g, h)$ with known g and h to account for the uncertainty. The researcher may set certain fixed values on α and β; hence consider that g, h α, and β are known values.

The Hierarchical Bayesian model is summarized as

$$\begin{cases} Z|Y, p \sim \text{pr}(Z|Y, p) & \text{data model} \\ Y \sim \text{pr}(Y|\alpha, \beta) & \text{spatial process model} \\ p \sim \text{Be}(g, h) & \text{hyper-parameter prior model} \end{cases} \tag{9.4}$$

[2] $\text{Be}(p|g, h) = p^{g-1}(1-p)^{h-1}1_{(0,1)}(p)/\text{B}(g, h)$

To learn $Y|Z$, one can compute the Bayesian MAP estimator of Y, i.e.

$$\begin{aligned}\hat{Y} &= \arg\max_{Y} \left(\log\left(\text{pr}\left(Z|Y\right)\text{pr}\left(Z\right)/\text{pr}\left(Z\right)\right)\right) \\ &= \arg\min_{Y} \left(-\log\left(\text{pr}\left(Z|Y\right)\right) - \log\left(\text{pr}\left(Y\right)\right)\right)\end{aligned}$$

where

$$\begin{aligned}\text{pr}\left(Z|Y\right) &= \int p^{n(Z,Y)}\left(1-p\right)^{n-n(Z,Y)}\text{Be}\left(p|g,h\right)\text{d}p \\ &= \int p^{n(Z,Y)}\left(1-p\right)^{n-n(Z,Y)}\frac{p^{g-1}\left(1-p\right)^{h-1}}{\text{B}\left(g,h\right)}\text{d}p \\ &= \frac{1}{\text{B}\left(g,h\right)}\int p^{n(Z,Y)+g-1}\left(1-p\right)^{n-n(Z,Y)+h-1}\text{d}p \\ &= \frac{1}{\text{B}\left(g,h\right)}\text{B}\left(n\left(Z,Y\right)+g, n-n\left(Z,Y\right)+h\right)\end{aligned}$$

via an optimization numerical algorithm, or perhaps the posterior expectation, i.e.

$$\hat{Y} = \text{E}\left(Y|Z\right) = \int Z\text{pr}\left(Y|Z\right)\text{d}Y$$

via MCMC, INLA, etc... Here the marginal posterior can be computed analytically as

$$\begin{aligned}\text{pr}\left(Y|Z\right) &= \int \text{pr}\left(Y,p|Z\right)\text{d}p = \int \frac{\text{pr}\left(Z|Y,p\right)\text{pr}\left(Y\right)\text{pr}\left(p\right)}{\int \text{pr}\left(Z|Y,p\right)\text{pr}\left(Y\right)\text{pr}\left(p\right)\text{d}p\text{d}Z}\text{d}p \\ &\propto \underbrace{\int \text{pr}\left(Z|Y,p\right)\text{pr}\left(p\right)\text{d}p}_{=\text{pr}(Z|Y)}\text{pr}\left(Y\right) = \text{pr}\left(Z|Y\right)\text{pr}\left(Y\right) \\ &\propto \underbrace{\frac{1}{\text{B}\left(g,h\right)}\text{B}\left(n_{(Z,Y)}+g, n-n\left(Z,Y\right)+h\right)}_{=\text{pr}(Z|Y)} \\ &\quad \times \underbrace{\frac{\exp\left(\alpha\sum_i Y_i + \beta\sum_{j\sim i} Y_iY_j\right)}{\sum_{Y\in\{0,1\}^n}\exp\left(\alpha\sum_i Y_i + \beta\sum_{j\sim i} Y_iY_j\right)}}_{=\text{pr}(Y)} \\ &\propto \text{B}\left(n\left(Z,Y\right)+g, n-n\left(Z,Y\right)+h\right)\exp\left(\alpha\sum_i Y_i + \beta\sum_{j\sim i} Y_iY_j\right) \qquad (9.5)\end{aligned}$$

Note that the only reason that we ignored the constant from the Ising process prior in (9.5) was because, in this particular example, the researcher considered α and β as known constants. Of course, that constant should not have be ignored if α and β had been considered as unknown, and hence we had to learn them.

Figure 9.2b shows the restored image as the Bayesian MAP estimator $\hat{Y} = \arg\max_{Y} (\log (\text{pr} (Y|Z)))$ of $Y|Z$ by using an R optimization function against (9.5).

9.2.2. *A simplified spatial model for count data (e.g. Counts analysis).*

Example 131. Consider the statistical problem scenario where there is available a dataset $\{(X_i, s_i, Z_i) ; i = 1, ..., n\}$, where $Z_i \in \mathbb{N}$ is the count of the occurrence of an event in a particular time interval, at a location s_i, and associated with a vector of covariates (other measurements) $X_i = (X_{i,1}, ..., X_{i,k})^\top$, for $i \in \mathcal{S}$, with $\mathcal{S} = \{1, ..., n\}$, and $n \in \mathbb{N} - \{0\}$ fixed. So, denote $(Z_i)_{i \in \mathcal{S}}$ as the observed vector. Assume that $Z_i \in \mathcal{Z}^{\mathcal{S}}$, with $\mathcal{Z} = \mathbb{N}$ and $\mathcal{S} = \{1, ..., n\}$.

- Such a scenario is suitable for the Columbus OH data set which concerns spatially correlated count data arising from small area sampling of some underlying process. This is the R dataset `columbus{spdep}`. Briefly, the Columbus data frame has 49 rows and 22 columns. Unit of analysis is 49 neighborhoods in Columbus, OH, 1980 data. The date frame has among others variables

  NEIG: neighborhood id value (1-49); conforms to id value used in Spatial Econometrics book.

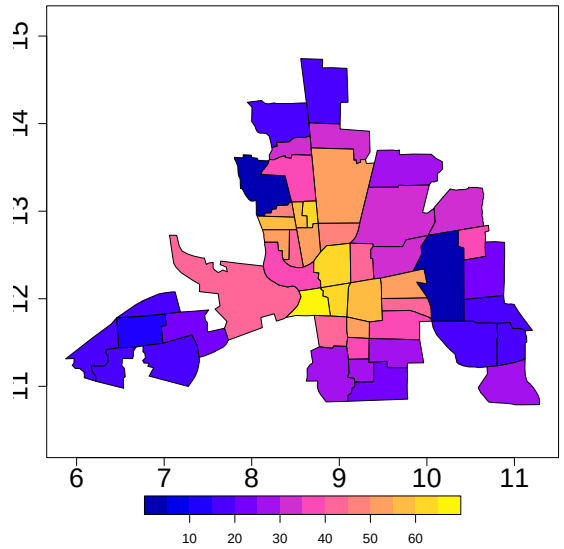
  CRIME: residential burglaries and vehicle thefts per thousand households in the neighborhood

  HOVAL: housing value (in 1,000 USD)

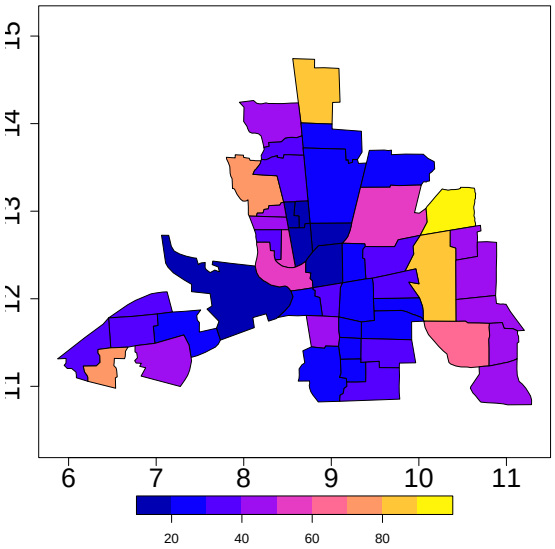
  INC: household income (in 1,000 USD)

$$\left\{ \left(\underbrace{\text{HOVAL}_i, \text{INC}_i}_{X_i}, \underbrace{\text{NEIG}}_{s_i}, \underbrace{\text{CRIME}_i}_{Z_i} \right)^\top ; i = 1, ..., \underbrace{49}_{n} \right\}$$

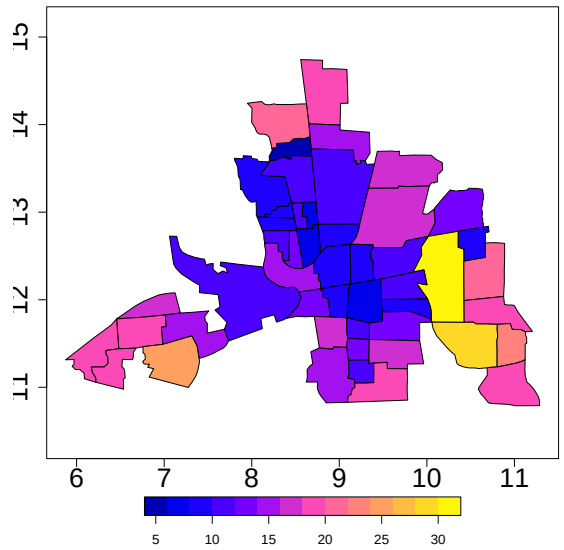
- Figure 1a shows the Property crime (number per thousand households) in 49 districts/neighborhood in Columbus in 1980, as well as the average value of the house in USD. Figure 1b presents the corresponding average house value. For privacy reasons, these are typically aggregated over areas that are large enough to ensure that the counts cannot be traced back to individuals.
- Interest may lie to find whether high rates of crime are clustered in a particular areas, and, perhaps what is the association of it with the value of the houses in the area.



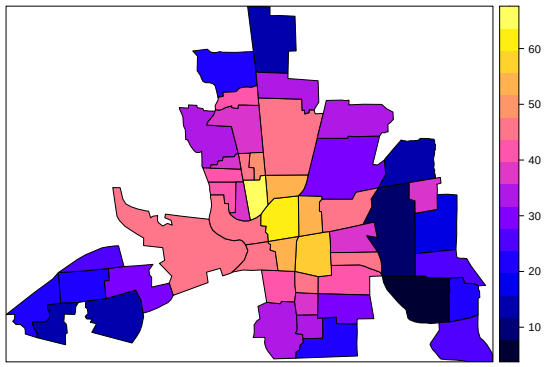
(A) CRIME: residential burglaries and vehicle thefts per thousand households in the neighborhood



(B) HOVAL: housing value (in 1,000 USD)



(C) INC: household income (in 1,000 USD)



(D) Fitted CRIME ~ HOVAL+INC in a Poisson model with rate modeled as CAR with mean CRIME ~ HOVAL+INC

FIGURE 9.3. Columbus Columbus OH spatial analysis dataset

For the data model, it is natural to assume that for site s_i the observable count Z_i is sampled from a Poisson distribution with some given rate $\lambda_i := \mathrm{E}\left(Z_i|Y_i\right) = \log\left(Y_i\right)$, different for different sites s_i and depending on an unknown/unobserved and underpinning spatial

process $(Y_i)_{i\in\mathcal{S}}$. The researcher may specify the data model as

$$Z_i|Y_i \overset{\text{ind}}{\sim} \text{Poisson}\left(\lambda_i\left(Y_i\right)\right), \text{ for } i=1,...,n \tag{9.6}$$
$$\text{where} \quad \log\left(\lambda_i\right) = Y_i$$

This imposes the (rather strong) assumption that Z_i and Z_j are conditionally independent given the spatial process $(Y_i)_{i\in\mathcal{S}}$.

- For the Columbus dataset, data model (9.6) is reasonable because the observation Z_i represents count namely number of event in a specific time and space and with fixed rate λ_i for each individual neighborhood $i \in \mathcal{S}$.

The spatial process Y is unknown. To specify the uncertainty on Y, the researcher may judge to assign a CAR model prior on $(Y_i)_{i\in\mathcal{S}}$, for instance

$$Y_i|Y_{-i} \sim \text{N}\left(\mu_i + \beta_{i,j}\left(Y_j - \mu_j\right), \kappa_i\right), \text{ for } i=1,...,n \tag{9.7}$$
$$\mu_i = X_i^\top \alpha.$$

To reduce parametric dimensionality, we impose a more restrictive structure such that $\kappa_i = 1/\tau$ for all $i = 1, ..., n$, and

$$\beta_{i,j} = \begin{cases} \phi & \text{if } i \sim j \\ 0 & \text{if } i \not\sim j \text{ or } i = j \end{cases}$$

- In the Columbus example we Spatial process (9.7) is suitable with regressors $X_i = (1, \text{HOVAL}_i, \text{INC}_i)^\top$; that is CRIME ~ HOVAL + INC.

$$\mu_i = \alpha_0 + \alpha_1 \text{HOVAL}_i + \alpha_2 \text{INC}_i, \quad i = 1, ..., n$$

  This is because we are interested in investigating "whether high rates of crime (CRIME) are clustered in a particular areas ($i \in \mathcal{S}$), eg areas with expensive houses (HOVAL), and if yes, perhaps what is the association of it with the value of the houses in the area (INC)". Hence we can use our model in order to see the association of CRIME with SPACE (i.e. $i \in \mathcal{S}$), HOVAL (i.e. house value), and INC (i.e. income).

- Note that unlike the usual linear model, here I have managed to introduce spatial dependence in the model as well by

$$\begin{aligned}\mathrm{E}\left(Y_i|Y_{-i}\right) =& \underbrace{\alpha_0+\alpha_1\mathrm{HOVAL}_i+\alpha_2\mathrm{INC}_i}_{\text{cocariate dependence only}}+\underbrace{\sum_{i\sim j}\phi\left(Y_j-\mu_j\right)}_{\text{spatial dependence}}\\ =& \underbrace{\alpha_0+\alpha_1\mathrm{HOVAL}_i+\alpha_2\mathrm{INC}_i}_{\text{cocariate dependence only}}+\underbrace{\sum_{i\sim j}\phi Y_j}_{\text{spatial dependence only}}\\ &-\underbrace{\sum_{i\sim j}\phi\left[\alpha_0+\alpha_1\mathrm{HOVAL}_j+\alpha_2\mathrm{INC}_j\right]}_{\text{interaction of space and covariates}}\end{aligned}$$

This essentially produces a joint model (see Section 4.2 or just use Theorem 100)

$$Y\sim\mathrm{N}\left(X\alpha,\left(I-\phi N\right)^{-1}\tau^{-1}\right)$$

where N is an $n\times n$ matrix with $[N]_{i,j}=1\left(\{i\sim j\},i\neq j\right)$, where $\sim$ is defined to denote adjacent sites (or otherwise spatial locations sharing same boarders).

For the unknown hyper-parameters α, ϕ and κ, the researcher may consider hyper-priors $\alpha\sim\mathrm{N}\left(0,\Sigma_\alpha\right)$, $\phi\sim\mathrm{U}\left(0,\phi_{\max}\right)$, and $\tau\sim\chi^2\left(\nu\right)$; the prior distributions here are chosen for demonstration. The rest hyper-parameters $\Sigma_\alpha>0$, $\phi_{\max}\in\{\phi>0:I-\phi N\text{ is non singular}\}$, $\nu>0$ are considered as unknown fixed constants set by the researcher based on his/her subjective believes.

The Bayesian spatial hierarchical model becomes

$$\begin{cases} Z_i|Y_i\overset{\text{ind}}{\sim}\mathrm{Poisson}\left(\exp\left(Y_i\right)\right),\forall i & \text{data model}\\ Y|\alpha,\phi,\tau\sim\mathrm{N}\left(X\alpha,\left(I-\phi N\right)^{-1}\tau^{-1}\right) & \text{spatial process model}\\ \alpha|\tau\sim\mathrm{N}\left(\mu_\alpha,\tau^{-1}\Sigma_\alpha\right) & \text{hyper-prior model}\\ \tau\sim\chi^2\left(\nu\right) & \text{hyper-prior model}\\ \phi\sim\mathrm{U}\left(0,\phi_{\max}\right) & \text{hyper-prior model}\end{cases}\tag{9.8}$$

The joint probability model becomes

$$\begin{aligned}\mathrm{pr}\left(Z,Y,\alpha,\beta,\tau\right)&=\prod_{i\in\mathcal{S}}\mathrm{pr}\left(Z_i|Y_i\right)\mathrm{pr}\left(Y|\alpha,\phi,\tau\right)\mathrm{pr}\left(\alpha|\tau\right)\mathrm{pr}\left(\tau\right)\mathrm{pr}\left(\phi\right)\\ &=\prod_{i\in\mathcal{S}}\mathrm{Poisson}\left(Z_i|\exp\left(Y_i\right)\right)\mathrm{N}\left(Y|X\alpha,\left(I-\phi N\right)^{-1}\tau^{-1}\right)\\ &\quad\times\mathrm{N}\left(\alpha|\mu_\alpha,\tau^{-1}\Sigma_\alpha\right)\mathrm{ChiSq}\left(\tau|\nu\right)\mathrm{U}\left(\phi|0,\phi_{\max}\right)\end{aligned}$$

Interest lies in learning $Y|Z$ which can be addressed for instance by the Bayesian MAP estimator

$$\hat{\lambda} = \arg\max_{Y} \left(\text{pr} \left(\lambda | Z \right) \right)$$

or the posterior expectation estimator

$$\hat{\lambda} = \text{E}_{\text{pr}} \left(\lambda | Z \right) = \text{E}_{\text{pr}} \left(\exp \left(Y \right) | Z \right)$$

$\text{pr} \left(\lambda | Z \right)$ can be computed via random variable transformation from $\text{pr} \left(Y | Z \right)$ which is given by the Bayesian theorem as

$$\text{pr} \left(Y | Z \right) = \int \text{pr} \left(Y, \alpha, \tau, \phi | Z \right) \text{d}\alpha \text{d}\tau \text{d}\phi$$

$$\text{pr} \left(Y, \alpha, \tau, \phi | Z \right) \propto \text{pr} \left(Z, Y, \alpha, \tau, \phi \right) = \text{pr} \left(Z | Y \right) \text{pr} \left(Y | \alpha, \tau, \phi \right) \text{pr} \left(\alpha | \tau \right) \text{pr} \left(\phi \right)$$

The above integration is analytically intractable, and hence its numerical computation can be performed by methods such as MCMC, INLA, etc...

Some marginal pdf/pmf

$$\begin{aligned} \text{pr} \left(Y, \alpha, \tau, \phi | Z \right) &\propto \text{pr} \left(Z, Y, \alpha, \tau, \phi \right) \\ &= \prod_{i \in \mathcal{S}} \text{Poisson} \left(Z_i | \exp \left(Y_i \right) \right) \text{N} \left(Y | X\alpha, \left(I - \phi N \right)^{-1} \tau^{-1} \right) \text{N} \left(\alpha | \mu_\alpha, \Sigma_\alpha \tau^{-1} \right) \\ &\quad \times \text{U} \left(\phi | 0, \phi_{\max} \right) \text{ChiSq} \left(\tau | \nu \right) \end{aligned}$$

and

$$\begin{aligned} \text{pr} \left(Y, \tau, \phi | Z \right) &\propto \int \text{pr} \left(Z, Y, \alpha, \tau, \phi \right) \text{d}\alpha \\ &= \int \prod_{i \in \mathcal{S}} \text{pr} \left(Z_i | Y_i \right) \text{pr} \left(Y | \alpha, \tau, \phi \right) \text{pr} \left(\alpha | \tau \right) \text{pr} \left(\tau \right) \text{pr} \left(\phi \right) \text{d}\alpha \\ &= \prod_{i \in \mathcal{S}} \text{pr} \left(Z_i | Y_i \right) \underbrace{\int \text{pr} \left(Y | \alpha, \tau, \phi \right) \text{pr} \left(\alpha | \tau \right) \text{d}\alpha}_{= \text{pr}(Y | \phi, \tau)} \text{pr} \left(\tau \right) \text{pr} \left(\phi \right) \\ &= \prod_{i \in \mathcal{S}} \text{Poisson} \left(Z_i | \exp \left(Y_i \right) \right) \text{N} \left(Y | X\mu_\alpha, \Sigma \left(\phi \right) \tau^{-1} \right) \text{ChiSq} \left(\tau | \nu \right) \text{U} \left(\phi | 0, \phi_{\max} \right) \end{aligned}$$

where

$$\begin{aligned} \text{pr} \left(Y | \phi \right) &= \int \text{pr} \left(Y | \alpha, \tau, \phi \right) \text{pr} \left(\alpha | \tau \right) \text{d}\alpha \\ &= \text{N} \left(Y | X\mu_\alpha, \left(\left(I - \phi \right)^{-1} + X \Sigma_\alpha X^\top \right) \tau^{-1} \right) \\ &= \text{N} \left(Y | X\mu_\alpha, \left(\left(I - \phi N \right) + \left(I - \phi N \right) X \left(\Sigma_\mu^{-1} + X^\top \left(I - \phi N \right) X \right)^{-1} X^\top \left(I - \phi N \right) \right) \tau^{-1} \right) \\ &= \text{N} \left(Y | X\mu_\alpha, \Sigma \left(\phi \right) \tau^{-1} \right) \end{aligned}$$

where

$$\Sigma(\phi) = \left((I - \phi N) + (I - \phi N) X \left(\Sigma_{\mu}^{-1} + X^{\top} (I - \phi N) X \right)^{-1} X^{\top} (I - \phi N) \right)$$

and

$$\begin{aligned}
\operatorname{pr}(Y, \phi | Z) &\propto \int \operatorname{pr}(Z, Y, \tau, \phi) \, \mathrm{d}\tau \\
&= \int \prod_{i \in \mathcal{S}} \operatorname{pr}(Z_i | Y_i) \operatorname{pr}(Y | \tau, \phi) \operatorname{pr}(\tau) \operatorname{pr}(\phi) \, \mathrm{d}\alpha \\
&= \prod_{i \in \mathcal{S}} \operatorname{pr}(Z_i | Y_i) \underbrace{\int \operatorname{pr}(Y | \tau, \phi) \operatorname{pr}(\tau) \, \mathrm{d}\tau}_{= \operatorname{pr}(Y | \phi)} \operatorname{pr}(\phi) \\
&= \prod_{i \in \mathcal{S}} \operatorname{Poisson}(Z_i | \exp(Y_i)) \operatorname{T}\left(Y | X \mu_{\alpha}, \frac{1}{\nu} \Sigma(\phi), \nu \right) \operatorname{U}(\phi | 0, \phi_{\max})
\end{aligned}$$

because

Hint:: The following definition is given:
A d dimensional random vector y follows a Student t distribution with degrees of freedom ν, mean parameter μ, and scale parameter Σ iff it can be represented as $y = \mu + \sqrt{\nu/\xi} x$ results as $\xi \sim \chi^2_{\nu}$, and $x \sim \mathrm{N}(0, \Sigma)$. It is denoted as $y \sim \mathrm{T}(\mu, \Sigma, \nu)$
If $\Sigma > 0$, then x has pdf

$$\operatorname{T}(y | \mu, \Sigma, \nu) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2) \nu^{\frac{d}{2}} \pi^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} (y - \mu)^{\top} \Sigma^{-1} (y - \mu) \right)^{-\frac{\nu + d}{2}}$$

$$\begin{aligned}
\operatorname{pr}(Y | \phi) &= \int \operatorname{pr}(Y | \phi, \tau) \operatorname{pr}(\tau) \, \mathrm{d}\tau \\
&= \int \mathrm{N}\left(Y | X \mu_{\alpha}, \Sigma(\phi) \tau^{-1} \right) \operatorname{ChiSq}(\tau | \nu) \, \mathrm{d}\tau \\
&= \int \mathrm{N}\left(Y | X \mu_{\alpha}, \Sigma(\phi) \tau^{-1} \frac{1}{\nu} \nu \right) \operatorname{ChiSq}(\tau | \nu) \, \mathrm{d}\tau \\
&= \int \mathrm{N}\left(Y | X \mu_{\alpha}, \frac{\nu}{\tau} \left(\frac{1}{\nu} \Sigma(\phi) \right) \right) \operatorname{ChiSq}(\tau | \nu) \, \mathrm{d}\tau \\
&= \int \mathrm{N}\left(Y | X \mu_{\alpha}, \frac{\nu}{\tau} \left(\frac{1}{\nu} \Sigma(\phi) \right) \right) \operatorname{ChiSq}(\tau | \nu) \, \mathrm{d}\tau \\
&= \mathrm{T}\left(y | X \mu_{\alpha}, \frac{1}{\nu} \Sigma(\phi), \nu \right)
\end{aligned}$$

Notice that

$$\operatorname{pr}(Y | Z) = \int \operatorname{pr}(Y, \phi | Z) \, \mathrm{d}\phi$$

involves an analytically intractable one-dimensional integral which can be easily approximated by using a standard integration algorithm (e.g. parallelogram rule).

- In Columbus example, the resulted Bayesian hierarchical model is as in (9.8). Estimation was facilitated via MCMC methods. The estimates are computed as the posterior expected values given the data Z as

$$\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_{\text{const}} \\ \hat{\alpha}_{\text{HOVAL}} \\ \hat{\alpha}_{\text{INC}} \end{pmatrix} = \begin{pmatrix} 54.3139189 \\ -0.2821969 \\ -0.9882862 \end{pmatrix}$$

  $\hat{\phi} = 0.1589004$, and $\hat{\kappa} = 87.65$. The fitted counts $\hat{Y}$ are presented in Figure 9.3d.

By comparing Figure 9.3a and Figure 9.3d, we see that there are certain locations where the fitted counts $\hat{Y}$ and Z are substantially different. Perhaps, we could improve our parameterization in (9.7) by considering a less restrictive $\beta_{i,j}$ or by including more covariates in the mean μ_i.