

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 $(\Omega, \mathfrak{F}, \text{pr})$ 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

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

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

Definition 8. The discrete set of sites $\mathcal{S} = \{s_i; i = 1, \dots, n\}$ 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 $\mathcal{S} = \{s_i; i = 1, \dots, n\}$, we can consider a more convenient notation

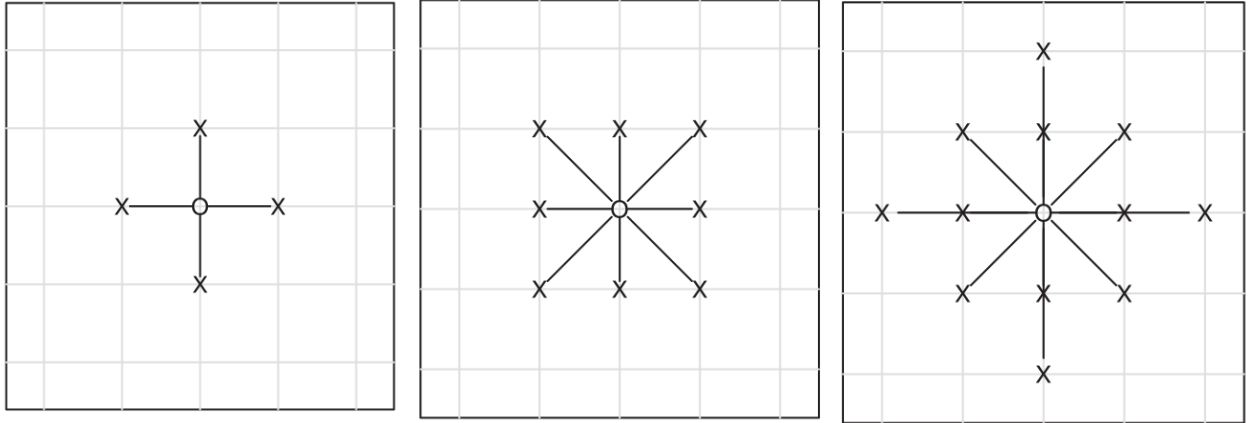
$$Z = (Z_s; s \in \mathcal{S})^\top = (Z_i = Z(s_i); i = 1, \dots, n)^\top.$$

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.

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

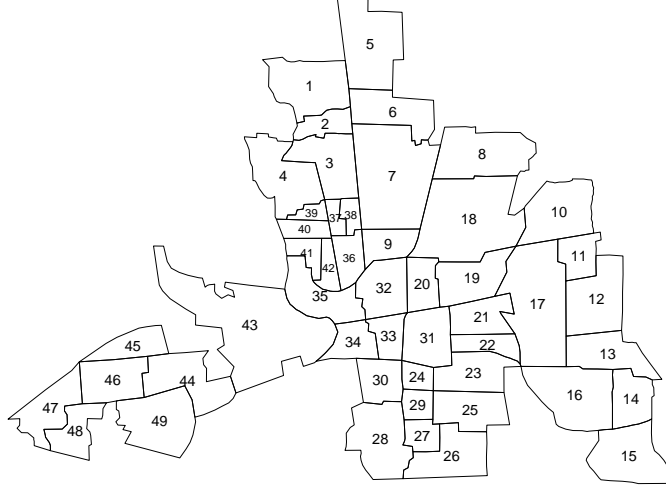


FIGURE 1.1. Lattice of spatial sites for Columbus dataset. Each neighborhood is a site. Each site is labeled. The collection of sites is the lattice of sites.

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 borders (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

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

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 $\text{pr}_{Z_i}(z_i = 1) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}$ 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 $\mathcal{A} \subset \mathcal{S}$

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

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

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

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

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

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

$$\tilde{\beta}(\omega) = \sum_{s \in \mathcal{S}} \beta_s e^{is^\top \omega}, \quad \omega \in (-\pi, \pi]^d.$$

- The inverse Fourier transform is

$$\beta_s = \int_{(-\pi, \pi]^d} e^{-is^\top \omega} \tilde{\beta}(\omega) d\omega$$

- 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 $c : \mathbb{Z}^d \rightarrow \mathbb{R}$ 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

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

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

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

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

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

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

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

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

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

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

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

with spectral measure

$$(2.1) \quad dF_V(\omega) = \left| \tilde{\beta}(\omega) \right|^2 dF_U(\omega), \quad \omega \in (-\pi, \pi]^d$$

where

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

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

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

$$(1) N_g = N_q = N$$

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

Proof. Omitted¹. □

¹See 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 $x|y \sim N(a + by, \sigma^2 + \tau^2 y^2)$ and $y|x \sim N(c + dx, \tilde{\sigma}^2 + \tilde{\tau}^2 x^2)$ 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 $\mathcal{S} = \{1, \dots, n\}$ with $n \in \mathbb{N}$, and such as $pr_Z(z) > 0$, $\forall z \in \mathcal{Z}^{\mathcal{S}}$. Then for all

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

Proof. I will show that

$$pr_Z(z) = \prod_{i=1}^n \frac{pr_i(z_i|z_1, \dots, z_{i-1}, z_{i+1}^*, \dots, z_n^*)}{pr_i(z_i^*|z_1, \dots, z_{i-1}, z_{i+1}^*, \dots, z_n^*)} pr_Z(z^*)$$

It is

$$pr_Z(z_1, \dots, z_n) = \frac{pr_n(z_n|z_1, \dots, z_{n-2}, z_{n-1})}{pr_n(z_n^*|z_1, \dots, z_{n-2}, z_{n-1})} pr_Z(z_1, \dots, z_{n-1}, z_n^*)$$

Let proposition P_j be

$$pr_Z(z) = \prod_{i=n-j}^n \frac{pr_i(z_i|z_1, \dots, z_{i-1}, z_{i+1}^*, \dots, z_n^*)}{pr_i(z_i^*|z_1, \dots, z_{i-1}, z_{i+1}^*, \dots, z_n^*)} pr_Z(z_1, \dots, z_{n-j-1}, z_{n-j}^*, \dots, z_n^*)$$

Proposition P_0 is true

$$(3.2) \quad pr_Z(z) = \frac{pr_n(z_n|z_1, \dots, z_{n-1})}{pr_n(z_n^*|z_1, \dots, z_{n-1})} pr_Z(z_1, \dots, z_{n-1}, z_n^*)$$

Proposition P_1 is true

$$pr_Z(z_1, \dots, z_{n-1}, z_n^*) = \frac{pr_{n-1}(z_{n-1}|z_1, \dots, z_{n-2}, z_n^*)}{pr_{n-1}(z_{n-1}^*|z_1, \dots, z_{n-2}, z_n^*)} pr_Z(z_1, \dots, z_{n-2}, z_{n-1}^*, z_n^*)$$

Assume that P_j is true. Then proposition P_{j+1} is true as well, because

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

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

Note 38. Theorem 37 shows that the joint $\text{pr}_Z(\cdot)$ can be constructed from its conditionals $\{\text{pr}_i(\cdot|\cdot)\}$ if distributions $\{\text{pr}_i(\cdot|\cdot)\}$ are compatible for $\text{pr}_Z(\cdot)$, under the requirement that this construction is invariant wrt the coordinate permutation $\{1, \dots, n\}$ 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

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

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

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 $\lambda_s = \sigma_\varepsilon^2$ 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

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

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

$$(4.3) \quad a_0 > 0, \text{ and } \sum_h |a_h| < a_0$$

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)

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

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

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

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

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

Assume that matrix A is such that $[A]_{i,i} = 0$ and $(I - A)^{-1}$ exists. Assume n -dimensional Gaussian random vector $E \sim N_n(0, \Lambda)$ with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. 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

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

Proof. Z is a linear combination of Gaussian random vectors, hence it follows a Gaussian distribution. Its mean and variance are

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

□

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

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

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

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

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

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

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

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

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

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

as, for $h \neq 0$ it is

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

and for $h = 0$

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

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

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

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 $\{\text{pr}_i(z_i|z_{\mathcal{S}-i})\}$ are Gaussian distributions

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

Proposition 56. Let $K = \text{diag}(\{\kappa_i\})$ 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

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

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 $\text{pr}_Z(z)$. By using Besag’s factorization theorem (Theorem 37) with reference state/configuration $z^* = 0$ we get

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

Recovering the mean from (4.9), it is

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

□

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 it 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

$$(4.10) \quad Z = \mu + B(Z - \mu) + E \iff E = (I - B)(Z - \mu)$$

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

$$E(E) = E((I - B)(Z - \mu)) = (I - B)E(Z - \mu) = 0$$

$$\text{Var}(E) = \text{Var}((I - B)Z) = (I - B)\text{Var}(Z)(I - B)^\top = (I - B)(I - B)^{-1}K(I - B)^\top$$

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 $\text{CAR}(\mathcal{H}_{\text{CAR}})$ and the stationary $\text{SAR}(\mathcal{H}_{\text{SAR}})$ 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 $\text{SAR}(\mathcal{H}_{\text{SAR}})$ and $\text{CAR}(\mathcal{H}_{\text{CAR}})$:

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

$$(4.12) \quad \text{Var}(\epsilon_s) = \kappa, \quad E(\epsilon_s) = 0$$

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

$$(4.14) \quad \text{Var}(\varepsilon_s) = \lambda, \quad E(\varepsilon_s) = 0$$

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

(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

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

If some $\text{SAR}(\mathcal{H}_{\text{SAR}})$ 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 $a_{(0,1)} \neq 0$ 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 $\mathcal{Z}^{\mathcal{A}} = \{z_{\mathcal{A}} : z \in \mathcal{Z}^{\mathcal{S}}\}$ 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 $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \rightarrow \mathbb{R}$ such that $V_{\emptyset}(\cdot) := 0$ and for every set $\mathcal{A} \subseteq \mathcal{S}$ the sum

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

exists.

Definition 65. In Definition 64, the function $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \rightarrow \mathbb{R}$ 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 $z_{\mathcal{S} \setminus \mathcal{B}} \in \mathcal{Z}^{\mathcal{S} \setminus \mathcal{B}}$

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

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 $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \rightarrow \mathbb{R}$. 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

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

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

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

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

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 $\cdot^{\mathcal{V}}$ and just write

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

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

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

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 $\mathcal{S} \setminus \mathcal{B}$

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

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 $V_{\mathcal{B}}(z) = 0$.

Note 75. In (5.2), the mapping $\mathcal{V} \rightarrow \text{pr}_{\mathcal{Z}}$ 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

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

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

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

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 $\mathcal{S} = \{1, \dots, n\}$ then

$$\begin{aligned} z^{[\emptyset, \zeta]} &= \left(\underbrace{\zeta, \dots, \zeta}_{n \text{ times}} \right)^{\top} ; & z^{\{i\}, \zeta} &= \left(\zeta, \dots, \zeta, \underbrace{z_i}_{i\text{th location}}, \zeta, \dots, \zeta \right)^{\top} ; \\ z^{\{i, j\}, \zeta} &= \left(\zeta, \dots, \zeta, \underbrace{z_i}_{i\text{th location}}, \zeta, \dots, \zeta, \underbrace{z_j}_{j\text{th location}}, \dots, \zeta \right)^{\top} ; & z^{[\mathcal{S}, \zeta]} &= (z_1, \dots, z_n)^{\top} ; \end{aligned}$$

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

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

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 $[z^{[\mathcal{B}, \zeta]}]_i = z_i$ if $i \in \mathcal{B}$ and $[z^{[\mathcal{B}, \zeta]}]_i = \zeta$ if $i \notin \mathcal{B}$. This is the unique normalized potential w.r.t $\zeta \in \mathcal{Z}$.

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

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

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

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 $\mathcal{V} = \{V_{\mathcal{A}} : \mathcal{A} \subseteq \mathcal{S}\} = \{V_{\{1\}}, V_{\{2\}}, V_{\{1,2\}}\}$. The decomposition of the energy $U(z = (z_1, z_2)^{\top}) := U_{\mathcal{S}}^{\mathcal{V}}(z)$ 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

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

by (5.4).

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

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

Consider Notation 77, for instance,

$$\begin{aligned} z^{[\emptyset, \zeta]} &= \left(\underbrace{\zeta, \dots, \zeta}_{n \text{ times}} \right)^{\top}; & z^{[\{i\}, \zeta]} &= \left(\zeta, \dots, \zeta, \underbrace{\zeta_i}_{\substack{\downarrow \\ \text{ith location}}}, \zeta, \dots, \zeta \right)^{\top}; \\ z^{[\{i,j\}, \zeta]} &= \left(\zeta, \dots, \zeta, \underbrace{\zeta_i}_{\substack{\downarrow \\ \text{ith location}}}, \zeta, \dots, \zeta, \underbrace{\zeta_j}_{\substack{\downarrow \\ \text{jth location}}}, \dots, \zeta \right)^{\top}; & z^{[\mathcal{S}, \zeta]} &= (z_1, \dots, z_n)^{\top}; \end{aligned}$$

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

$$(5.6) \quad V_{\{i\}}(z) = (-1)^{1-1} U(z^{[\{i\}, \zeta]}) + (-1)^{1-0} U(z^{[\emptyset, \zeta]}) = \alpha z_i,$$

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

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

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 $\text{card}(\mathcal{A}) > 2$.

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 ”. Its 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 $\text{pr}_i(z_i|z_{\mathcal{S}\setminus\{i\}})$ 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 is a 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.

$$(6.1) \quad \text{pr}_Z(z_{\mathcal{A}}|z_{\mathcal{S}\setminus\mathcal{A}}) = \text{pr}_Z(z_{\mathcal{A}}|z_{\partial\mathcal{A}})$$

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

Note 90. Definition 89 implies that (6.1) becomes

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

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

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.

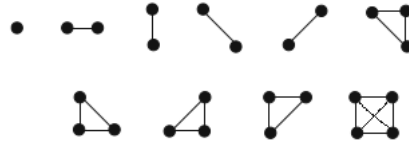


FIGURE 6.1. Examples of cliques

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 \mathcal{C} .

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

Theorem 94. (Hammersley–Clifford) Let Z be an \mathcal{Z} -valued random field on a finite collection $\mathcal{S} \neq \emptyset$ of sites such that $pr_Z(z_A | z_{\mathcal{C} \setminus A}) > 0$ for all $A \subset \mathcal{S}$ 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

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

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

Proof. □

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

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

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

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

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

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

Rearranging I get simplifies

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

Because $s \not\sim t$, it is $pr_Z(z_s^{\mathcal{B}, \delta} | z_{\mathcal{S} \setminus s}^{\mathcal{B}, \delta}) = pr_Z(z_s^{\mathcal{B} \cup \{t\}, \delta} | z_{\mathcal{S} \setminus s}^{\mathcal{B} \cup \{t\}, \delta})$ and $pr_Z(z_s^{\mathcal{B} \cup \{s, t\}, \delta} | z_{\mathcal{S} \setminus s}^{\mathcal{B} \cup \{s, t\}, \delta}) = pr_Z(z_s^{\mathcal{B} \cup \{s\}, \delta} | z_{\mathcal{S} \setminus s}^{\mathcal{B} \cup \{s\}, \delta})$. 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 \Leftarrow : By using (5.2), I can write

$$\text{pr}_Z(z_{\mathcal{A}}|z_{\mathcal{S}\setminus\mathcal{A}}) = \frac{1}{C_{\mathcal{A}}(z_{\mathcal{S}\setminus\mathcal{A}})} \exp(U_{\mathcal{A}}(z))$$

where

$$U_{\mathcal{A}}(z) = \sum_{\{C \subseteq \mathcal{S} : \mathcal{A} \cap C \neq \emptyset\}} V_C(z_C)$$

depends only on $\{z_i : i \in \mathcal{A} \cup \partial\mathcal{A}\}$ as $\text{pr}_Z(\cdot)$ is a Markov Random Field.

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

$$\text{pr}_Z(z) \propto \exp\left(\sum_{C \in \mathcal{C}} \log(\varphi_C(z_C))\right)$$

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

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

for \Rightarrow : we need to show that there exists an interaction potential $\varphi = \{\varphi_C : C \in \mathcal{C}\}$ defined on the cliques \mathcal{C} such that $\text{pr}_Z(\cdot)$ is a Gibbs Random Field with interaction potential φ .

for \Leftarrow : a Gibbs Random Field with potentials $\{\varphi_C : C \in \mathcal{C}\}$ defined on the cliques \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

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

I can find that

$$\begin{aligned} \varphi_{\emptyset} &= 1 / \sum_{z \in \mathcal{Z}^{\mathcal{S}}} \exp\left(\alpha \sum_{i \in \mathcal{S}} z_i + \beta \sum_{\{i,j\}: i \sim j} z_i z_j\right) \\ (6.4) \quad \varphi_{\{i\}}(z_{\{i\}}) &= \exp(\alpha z_i), \quad \forall i \in \mathcal{S} \\ (6.5) \quad \varphi_{\{i,j\}}(z_{\{i,j\}}) &= \exp(\beta z_i z_j), \quad \forall i, j \in \mathcal{S} \text{ s.t. } i \sim j \\ \varphi_{\{i,j\}}(z_{\{i,j\}}) &= 1, \quad \forall i, j \in \mathcal{S} \text{ s.t. } i \not\sim j \\ \varphi_{\mathcal{A}}(z_{\mathcal{A}}) &= 1, \quad \forall \mathcal{A} \subset \mathcal{S} \text{ s.t. } \text{card}(\mathcal{A}) > 2 \end{aligned}$$

where $\{i\}$ and $\{i, j\}$ satisfying $i \sim j$ are cliques. Alternatively, as \emptyset is not a clique if that φ_\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 $\text{pr}_X(x|\theta)$ can be expressed in the form

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

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 that takes values in \mathcal{Z} on a finite set of points \mathcal{S} and has marginal probability

$$(7.1) \quad \text{pr}_Z(z) = \frac{\exp(U(z))}{\int \exp(U(z)) dz}, \quad z \in \mathcal{Z}^{\mathcal{S}}.$$

Consider some fixed normalization configuration $\zeta = (\zeta, \dots, \zeta)^\top \in \mathcal{Z}^{\mathcal{S}}$. Assume that the following assumptions are satisfied:

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

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

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 exponential family distributions

$$(7.2) \quad \log(\text{pr}_i(z_i|z_{-i})) = (A_i(z_{-i}))^\top B_i(z_i) + C_i(z_i) + D_i(z_{-i}),$$

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

Then, necessarily,

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

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

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

$$(7.3) \quad V_i(z_i) = (\alpha_i)^\top B_i(z_i) + C_i(z_i)$$

$$(7.4) \quad V_{i,j}(z_i, z_j) = (B_i(z_i))^\top \beta_{i,j} B_j(z_j)$$

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 101. Automodel is called the model satisfying the assumptions of Theorem 100.

Definition 102. Univariate automodel is the automodel with $\ell = 1$ in Theorem 100.

Definition 103. Multi-parameter automodel with $\ell \geq 1$ in Theorem 100.

Remark 104. 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 $\ell = 1$, (7.3) and (7.4) become

$$(7.5) \quad V_i(z_i) = \alpha_i B_i(z_i) + C_i(z_i)$$

$$(7.6) \quad V_{i,j}(z_i, z_j) = \beta_{i,j} B_i(z_i) B_j(z_j)$$

7.2. Besag auto-models.

Definition. Z follows a Besag's auto-model if Z is real-valued and its joint distribution $\text{pr}_Z(z)$ is given by

$$(7.7) \quad \text{pr}_Z(z) = \frac{1}{C} \exp \left(\sum_{i \in \mathcal{S}} V_i(z_i) + \sum_{\{i,j\} \in \mathcal{S}^2: i \sim j} \beta_{i,j} z_i z_j \right), \quad z \in \mathcal{Z}_{\mathcal{S}}$$

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

Note 105. The following allows us to define a Markov Random Field model from a set of conditional distributions (characteristics) whose compatibility is automatically satisfied.

Proposition 106. *If each of the*

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

is a family of real-valued $z_i \in \mathbb{R}$ conditional distributions which are members of the exponential family of distributions (7.2) with $B_i(z_i) = z_i$ for $i \in \mathcal{S}$, then they are compatible a Besag's auto-model with distribution (7.7) if $\beta_{i,j} = \beta_{j,i}$ for all $i, j \in \mathcal{S}$.

Proof. For

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

it is

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

so

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

□

Example 107. (Logistic automodel / Ising model) Consider that $Z(s)$ 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 $\mathcal{S} = \{1, \dots, n\}$, $n \in \mathbb{N} - \{0\}$.

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

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

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

Then the characteristics are

$$(7.8) \quad \text{pr}_i(z_i|z_{-i}) = \frac{\exp(z_i \theta_i(z_{-i}))}{1 + \exp(\theta_i(z_{-i}))} 1(z_i \in \{0, 1\})$$

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

$$(7.9) \quad \theta_i(z_{-i}) = \alpha_i + \sum_{j:j \sim i} \beta_{i,j} z_j$$

for $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ with $\beta_{i,j} = \beta_{j,i}$. Then (7.8) becomes

$$(7.10) \quad \log(\text{pr}_i(z_i|z_{-i})) = \underbrace{\underbrace{z_i}_{B_i(z_i)} \left(\underbrace{\alpha_i + \sum_{j \sim i} \beta_{i,j} z_j}_{A_i(z_{-i})} \right)}_{A_i(z_{-i})} + \underbrace{0}_{C_i(z_i)} + \underbrace{\left(-\log \left(1 + \exp \left(\alpha_i + \sum_{j:j \sim i} \beta_{i,j} z_j \right) \right) \right)}_{D_i(z_{-i})}$$

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

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

Also, I can get $C_i(\zeta) = 0$ and $B_i(\zeta) = 0$ by considering a reference point $\zeta = 0$. From Theorem 106, (7.8) with (7.9), the conditionals $z_i|z_{-i}$ are compatible as a Besag auto-model with marginal distribution

$$(7.11) \quad \text{pr}_Z(z) \propto \exp \left(\overbrace{\sum_i \alpha_i z_i + \sum_i \sum_{j:j \sim i} \beta_{i,j} z_i z_j}^{U(z)=} \right)$$

$\underbrace{\sum_i \alpha_i z_i}_{V_i(z_i)} \quad \underbrace{\sum_i \sum_{j:j \sim i} \beta_{i,j} z_i z_j}_{\sum_{\{i,j\}:j \sim i}}$

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 $\text{pr}_i(z_i|z_{-i}) = \frac{\exp(z_i \alpha_i)}{1 + \exp(\alpha_i)}$.
- 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 108. (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 $\mathcal{S} = \{1, \dots, n\}$, where $n \in \mathbb{N} - \{0\}$.

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

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

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

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

with mean $E(x|\lambda) = \lambda$.

Then the full conditionals (characteristics) are

$$(7.12) \quad \text{pr}_i(z_i|z_{-i}) = \frac{1}{z_i!} (\lambda_i(z_{-i}))^{z_i} \exp(-\lambda_i) \mathbf{1}(z_i \in \mathbb{N})$$

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

$$(7.13) \quad \log(\lambda_i(z_{-i})) = \alpha_i + \sum_{j:j \sim i} \beta_{i,j} z_j$$

for $\{\alpha_i\}$ and $\{\beta_{i,j}\}$ with $\beta_{i,j} = \beta_{j,i}$. So (7.12) becomes

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

with

$$A_i(z_i) = \alpha_i + \sum_{j \sim i} \beta_{i,j} B_i(z_j)$$

$$B_i(z_{-i}) = z_i$$

$$C_i(z_i) = \log(z_i!)$$

$$D_i(z_{-i}) = 0$$

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 106, (7.12) with (7.13), the conditionals $z_i|z_{-i}$ are compatible as a Besag auto-model with marginal distribution

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

or otherwise the energy function is

$$U(z) = \sum_i (\alpha_i z_i + \log(z_i!)) + \sum_{j \sim i} \beta_{i,j} z_i z_j$$

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

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

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

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

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 .

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

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 109. 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.