Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Handout 4: 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] Gaetan, C., & Guyon, X. (2010; Ch 3). Spatial statistics and modeling (Vol. 90). New York: Springer.

Specialized reading.

[3] Kent, J. T., & Mardia, K. V. (2022). Spatial analysis (Vol. 72). John Wiley & Sons. (on Spatial analysis)

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

Note 1. Recall from Section 2.2 of "Handout 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 adjacency relation or neighborhood structure.

Note 2. This means we need to introduce new basic building models able to acceptably 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 adjusting it 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.

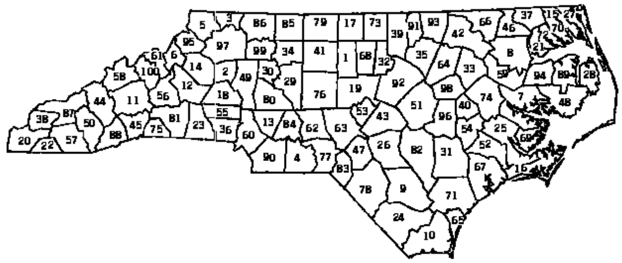

FIGURE 1.1. Lattice of spatial sites for North Carolina SIDS data. Each county is a site. Each site is coded according to its alphabetical order. The collection of sites is the lattice of sites.

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

$$\mathrm{pr}_{Z}\left(z\right)=\mathrm{pr}\left(Z=z\right)=\mathrm{pr}\left(\left\{Z_{s}=z_{s};\ s\in\mathcal{S}\right\}\right),\ \forall z_{s}\in\mathcal{Z}^{\mathcal{S}}$$

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

$$Z=\left(Z_{s};s\in\mathcal{S}\right)^{\top}=\left(Z_{i}=Z\left(s_{i}\right);i=1,...n\right)^{\top}.$$

Notation 10. The notation $i\sim j$ between two sites $i,j\in\mathcal{S}$ means that "sites i and j are adjacent".

Example 11. Recall the North Carolina SIDS data Ex 24 in Handout 1. Fig 1.1 presents the sites and the lattice of sites. Each county is a site. Each site is coded according to its alphabetical order. The collection of sites is the lattice of sites coded according to alphabetical order of the county name. One may define the "adjacency between sites $i\sim j$" as the counties that share common boarders. Then for site $i=43$, $i\sim j$ involves any $j\in\{63,53,19,92,51,82,26,47\}$ in Fig 1.1.

Example 12. (Logistic/Ising model) Consider Z_i denotes a characteristic presence coded as 1 or absence coded as 0 on a region labeled by $i\in\mathcal{S}$. Then $\mathcal{Z}=\{0,1\}$. The Ising model is defined by the (joint) PMF

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

E.g., it can model a black & white image 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, the characteristic is observed with probability $\mathrm{pr}_{Z_i}\left(z_i=1\right)=\frac{\exp(\alpha)}{1+\exp(\alpha)}$

when $\beta = 0$. The characteristic's present is encouraged in neighboring sites when $\beta > 0$, and discouraged when $\beta < 0$.

Notation 13. We use notation, for $\mathcal{A} \subset \mathcal{S}$

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

Definition 14. Local characteristics of a random field Z on $\mathcal{S}$ with values in $\mathcal{Z}$ are the conditionals

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

Example 15. The 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. Compatibility of conditional distributions

Note 16. Essentially, we attempt to answer the following question. Under what conditions a parameterized family $\{\pi_i\left(z_i|z_{\mathcal{S}-i}\right); i \in \mathcal{S}\}$ of distributions on $\mathcal{S}$ conditioned on $z_{\mathcal{S}-i}$ can represent conditional distributions of a joint distribution $\mathrm{pr}_Z(\cdot)$?

Note 17. To answer the above we need to be able to specify partially or wholly the joint and conditional distribution of pr_Z. However, an arbitrary chosen set of conditional distributions $\{\pi_i(\cdot|\cdot)\}$ is not generally compatible, and hence we need to impose conditions.

Proposition 18. [1]*(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 $\int u(x)\,dG(x|y) < \infty$*

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

Note 19. 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 20. The conditionals $x|y \sim \mathrm{N}\left(a+by, \sigma^2+\tau^2y^2\right)$ and $y|x \sim \mathrm{N}\left(c+dx, \tilde{\sigma}^2+\tilde{\tau}^2x^2\right)$ are compatible if $\tau^2 = \tilde{\tau}^2 = 0$ and $d/\tilde{\sigma}^2 = b/\sigma^2$.

Solution. See Ex 24 in the Exercise sheet.

Note 21. Proposition 18 can be extended to more dimensions.

Note 22. The following theorem shows that local characteristics can determine the entire distribution in certain cases.

Theorem 23. *(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*

$$\frac{pr_Z(z)}{pr_Z(z^*)} = \prod_{i=1}^{n} \frac{pr_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{pr_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}, \quad \forall z, z^* \in \mathcal{Z}^{\mathcal{S}} \tag{2.1}$$

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

$$\mathrm{pr}_Z(z) = \frac{\mathrm{pr}_n(z_n|z_1, ..., z_{n-1})}{\mathrm{pr}_n(z_n^*|z_1, ..., z_{n-1})} \mathrm{pr}_Z(z_1, ..., z_{n-1}, z_n^*) \tag{2.2}$$

Proposition P_1 is true

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

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 (2.1) is correct according to the induction principle. □

Note 24. The theorem 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 $\mathrm{pr}_Z(\cdot)$, as this reconstruction has to be invariant wrt the coordinate permutation $\{1,...,n\}$ and the reference state z^*– these invariances correspond to the conditions in Proposition 18.

3. Gaussian Autoregressive models

Modeling snapshot

Definition 25. Adjacency matrix N is called a matrix N with $[N]_{i,j} = 1(i \sim j)$ for some symmetric neighborhood relation $\sim$ on $\mathcal{S}$. It aims at spatially connecting unites i and j.

Definition 26. Proximity matrix W is called a matrix W which aims at spatially connecting unites i and j in some fashion for some symmetric neighbourhood relation $\sim$ on $\mathcal{S}$. Usually $[W]_{i,i} = 0$

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

Definition 27. Assume a random field $Z = (Z_s; s \in \mathcal{S})$ on a set of indexes $\mathcal{S}$ with values in $\mathcal{Z}$. We say that Z follows a conditional autoregressive model (CAR) if the distribution of each element Z_s of the random field Z is specified conditionally on the values at the neighboring sites of s.

3.1.1. *Gaussian CAR.*

Definition 28. Gaussian CAR assumes that the local characteristics $\{\mathrm{pr}_i(z_i|z_{\mathcal{S}-i})\}$ are Gaussian distributions

$$Z_i|z_{\mathcal{S}-i} \sim \mathrm{N}\left(\mu_i + \sum_{j\neq i} b_{i,j}(Z_j - \mu_j), \kappa_i\right) \tag{3.1}$$

with mean $\mathrm{E}(Z_i|Z_{\mathcal{S}-i}) = \mu_i + \sum_{j\neq i} b_{i,j}(Z_j - \mu_j)$ and variance $\mathrm{Var}(Z_i|Z_{\mathcal{S}-i}) = \kappa_i$ for $i \in \mathcal{S}$.

Proposition 29. *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 dimentions. If Z follows a Gaussian CAR (Def 28), $I - B$ is non-singular, and $(I-B)^{-1} K > 0$, then the joint distribution of Z is*

$$Z \sim N\left(\mu, (I-B)^{-1} K\right). \tag{3.2}$$

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 (3.1) are compatible with the joint distribution $\text{pr}_Z(z)$. By using Besag's factorization theorem with reference state $z^* = 0$ we get

$$\begin{aligned}
\text{pr}_Z(z) &= \prod_{i=1}^{n} \frac{\text{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*=0,...,z_n^*=0\right)}{\text{pr}_i\left(z_i^*=0|z_1,...,z_{i-1},z_{i+1}^*=0,...,z_n^*=0\right)}\text{pr}_Z(z^*=0) \\
&= \prod_{i=1}^{n} \frac{\text{N}\left(z_i|\sum_{j<i} b_{i,j}z_j+0,\kappa_i\right)}{\text{N}\left(0|\sum_{j<i} b_{i,j}z_j+0,\kappa_i\right)}\text{pr}_Z(z^*=0) \\
&\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i-\sum_{j<i} b_{i,j}z_j\right)^2+\frac{1}{2\kappa_i}\left(0-\sum_{j<i} b_{i,j}z_j\right)^2\right) \\
&= \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i^2-2z_i\sum_{j<i} b_{i,j}z_j\right)\right)\text{pr}_Z(z^*=0) \\
&= \exp\left(-\sum_i \frac{z_i^2}{2\kappa_i}+\frac{1}{2}2\sum_i\sum_{j<i}\frac{b_{i,j}}{\kappa_i}z_iz_j\right)\text{pr}_Z(z^*=0) \\
&= \exp\left(-\frac{1}{2}z^\top K^{-1}z+\frac{1}{2}z^\top K^{-1}Bz\right)\text{pr}_Z(z^*=0) = \exp\left(-\frac{1}{2}z^\top\left[K^{-1}(I-B)\right]z\right)\text{pr}_Z(z^*=0) \\
&= \text{N}\left(z|0,(I-B)^{-1}K\right)
\end{aligned} \tag{3.3}$$

Recovering the mean from (3.3), it is

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

□

Note 30. 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 $B = \phi N$ where $\phi > 0$ and N is an adjacency 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 31. The system in (3.2) can be rewritten as

$$Z = \mu + B(Z-\mu) + E \Longleftrightarrow E = (I-B)(Z-\mu) \tag{3.4}$$

by setting $E = (I-B)(Z-\mu)$. The distribution of Z in (3.2) induces a distribution on E as $E \sim \mathrm{N}\left(0, K(I-B)^{\top}\right)$ because

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

3.2. **Simultaneus Autoregressive (SAR) models.**

3.2.1. *Gaussian SAR.*

Note 32. CAR sets the AR relation, and specifies the distribution on Z which induces the distribution on E; see 3.4. SAR does does the reverse; sets the same AR relation but it specifies the distribution on E which induces the distribution on Z –this is more might be more intuitive (?).

Definition 33. Consider discrete set of sites $\mathcal{S} = \{s_i; i = 1, ..., n\}$. Consider a random field $Z = (Z_s; s \in \mathcal{S})^{\top} = (Z_i = Z(s_i); i = 1, ...n)^{\top}$ on the discrete set of indexes $\mathcal{S}$ with values in $\mathcal{Z}$. Define

$$Z = \mu + \tilde{B}(Z-\mu) + E \Longleftrightarrow E = \left(I-\tilde{B}\right)(Z-\mu)$$

Assume that matrix $\tilde{B}$ is such that $\left(I-\tilde{B}\right)^{-1}$ exists, and $\left[\tilde{B}\right]_{i,i} = 0$. Assume that $E = (E_i; i = 1, ...n)$ is an n-dimensional Gaussian random vector $E \sim \mathrm{N}_n(0, \Lambda)$ with $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$ whose elements are indexed by $\mathcal{S}$. Then we say that Z follows a Gaussian Simultaneous Autoregressive (SAR) model.

Proposition 34. *The joint distribution of Z following the SAR model in Def 33 is*

$$Z \sim N\left(\mu, \left(I-\tilde{B}\right)^{-1}\Lambda\left(I-\tilde{B}^{\top}\right)^{-1}\right) \tag{3.5}$$

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

$$\mathrm{E}(Z) = \mathrm{E}\left(\left(I-\tilde{B}\right)^{-1}E + \mu\right) = \mu,$$
$$\mathrm{Var}(Z) = \mathrm{Var}\left(\left(I-\tilde{B}\right)^{-1}E + \mu\right) = \left(I-\tilde{B}\right)^{-1}\mathrm{Var}(E)\left(I-\tilde{B}^{\top}\right)^{-1} = \left(I-\tilde{B}\right)^{-1}\Lambda\left(I-\tilde{B}^{\top}\right)^{-1}$$

□

3.3. **CAR vs SAR.**

Remark 35. From (3.2) and (3.5), CAR and SAR are equivalent iff

$$\underbrace{(I-B)^{-1}K}_{\text{CAR}} = \underbrace{\left(I-\tilde{B}\right)^{-1}\Lambda\left(I-\tilde{B}^{\top}\right)^{-1}}_{\text{SAR}}$$

Proposition 36. *Any SAR can be written as a CAR model.*

Proof. Let Λ be $n\times n$ positive diagonal matrix. Let $\tilde{B}$ be $n\times n$ positive matrix where $I-\tilde{B}$ is non-singular and $\tilde{B}_{i,i} := \left[\tilde{B}\right]_{i,i} = 0$. Then $\left(I-\tilde{B}\right)^{-1}\Lambda\left(I-\tilde{B}^{\top}\right)^{-1}$ is well defined and I need to solve wrt B and $K=\mathrm{diag}(\kappa_1,...,\kappa_n)$

$$\begin{aligned}(I-B)^{-1}K &= \left(I-\tilde{B}\right)^{-1}\Lambda\left(I-\tilde{B}^{\top}\right)^{-1} \Leftrightarrow\\ K^{-1}(I-B) &= \left(I-\tilde{B}^{\top}\right)\Lambda^{-1}\left(I-\tilde{B}\right) \Leftrightarrow\\ K^{-1}-K^{-1}B &= \Lambda^{-1}-\tilde{B}^{\top}\Lambda^{-1}-\Lambda^{-1}\tilde{B}+\tilde{B}^{\top}\Lambda^{-1}\tilde{B}\end{aligned}$$

If I focus of the diagonal part and set $B_{i,i} := [B]_{i,i} = 0$

$$\left[K^{-1}\right]_{i,i} - \cancelto{=0}{\left[K^{-1}B\right]_{i,i}} = \left[\Lambda^{-1}\right]_{i,i} - \cancelto{=0}{\left[\tilde{B}^{\top}\Lambda^{-1}\right]_{i,i}} - \cancelto{=0}{\left[\Lambda^{-1}\tilde{B}\right]_{i,i}} + \left[\tilde{B}^{\top}\Lambda^{-1}\tilde{B}\right]_{i,i}$$

so

$$\kappa_i = \left(\frac{1}{\lambda_i}+\sum_{j=1}^{n}\frac{\tilde{B}_{j,i}^2}{\lambda_j}\right)^{-1} > 0,\ \forall i=1,...,n$$

and hence I can solve with respect to K and Bin a manner that they satisfy the assumptions of CAR. □

Remark 37. The converse of Prop 36 is not true.

Proposition 38. *Any positive-definite covariance matrix Σ can be expressed as the covariance matrix of a CAR model $\Sigma_{CAR} = (I-B)^{-1}K$, for a unique pair of matrices B and K where $(I-B)$ is non-singular and K is diagonal.*

Proof. Express

$$\Sigma^{-1} = D - R$$

for

$$[D]_{i,j} = \begin{cases}\left[\Sigma^{-1}\right]_{i,i} & i=j\\ 0 & i\neq j\end{cases},\ \text{and}\ [R]_{i,j} = \begin{cases}0 & i=j\\ -\left[\Sigma^{-1}\right]_{i,j} & i\neq j\end{cases}$$

then

$$\Sigma = (D-R)^{-1} = \left(D\left(I-D^{-1}R\right)\right)^{-1} = \left(I-D^{-1}R\right)^{-1}D^{-1}$$

Now define $B = D^{-1}R$ and $K = D^{-1}$, and you get $\Sigma = \Sigma_{\text{CAR}}$. Now regarding the uniqueness, assume there is another pair of $\mathring{B}$, and $\mathring{K}$ such that $\Sigma_{\text{CAR}} = \left(I - \mathring{B}\right)^{-1} \mathring{K}$. Then

$$\operatorname{diag}\left(\Sigma^{-1}\right) = \operatorname{diag}\left(\Sigma_{\text{CAR}}^{-1}\right) = \operatorname{diag}\left(\mathring{K}^{-1}\left(I - \mathring{B}\right)\right) = \operatorname{diag}\left(\mathring{K}^{-1}\right)$$

and similarly $\operatorname{diag}\left(\Sigma^{-1}\right) = \operatorname{diag}\left(K^{-1}\right)$. Hence it has to be $\mathring{K} = K$ because both are diagonal matrices. Then it is

$$\left(I - \mathring{B}\right)^{-1} \mathring{K} = (I - B)^{-1} K \overset{\mathring{K}=K}{\Longleftrightarrow} \mathring{B} = B.$$

So the representation is unique. □

Proposition 39. *Any positive-definite covariance matrix Σ can be expressed as the covariance matrix of a SAR model $\Sigma_{SAR} = \left(I - \tilde{B}\right)^{-1} \Lambda \left(I - \tilde{B}^{\top}\right)^{-1}$ for a (non-unique) pair of matrices $\tilde{B}$ and Λ where $\left(I - \tilde{B}\right)$ is non-singular, $\left[\tilde{B}\right]_{i,i} = 0$, and Λ is diagonal.*

Proof. Express

$$\Sigma^{-1} = LL^{\top}$$

where L is a lower triangular matrix with $[L]_{i,i} > 0$. Such matrix decomposition can be done by Cholesky decomposition, square-matrix decomposition, etc... and hence it is not always unique. Then

$$\Sigma = \left(LL^{\top}\right)^{-1} = L^{-\top}L^{-1}$$

Now express, $L = D - C$ for

$$[D]_{i,j} = \begin{cases} [L]_{i,i} & i = j \\ 0 & i \neq j \end{cases}, \text{ and } [C]_{i,j} = \begin{cases} 0 & i = j \\ -[L]_{i,j} & i \neq j \end{cases}$$

then

$$\begin{aligned}\Sigma &= (D - C)^{-\top} (D - C)^{-1} = \left(I - D^{-1}C\right)^{-\top} D^{-\top}D^{-1} \left(I - D^{-1}C\right)^{-1} \\ &= \left(I - C^{\top}D^{-\top}\right)^{-1} D^{-\top}D^{-1} \left(I - \left(C^{\top}D^{-\top}\right)^{\top}\right)^{-1}\end{aligned}$$

Set $\tilde{B} = C^{\top}D^{-\top}$ and $\Lambda = D^{-\top}D^{-1}$ and you get $\Sigma_{\text{SAR}} = \Sigma$ for non-unique pairs of $\tilde{B}$ and Λ. □

Proposition 40. *Any SAR model can be written as a unique CAR model.*

Proof. SAR and CAR are both Gaussian's with the same mean. SAR's variance matrix is positive definite, and hence it can be written in a unique manner as a CAR's variance matrix by Prop 38. □

Proposition 41. *Any CAR model can be written as a non-unique SAR model.*

Proof. SAR and CAR are both Gaussian's with the same mean. CAR's variance matrix is positive definite, and hence it can be written in a non-unique manner as a SAR's variance matrix by Prop 39. $\square$

Example 42. Show that

(1) Z_i and E_j are independent for $i \neq j$ in Gaussian CAR
(2) Z_i and E_j are not necessarily independent for $i \neq j$ in Gaussian SAR

Solution.

(1) For Gaussian CAR,

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

which is a diagonal; hence Z_i and E_j are independent for $i \neq j$.

(2) For Gaussian SAR,

$$\text{Cov}\,(Z, E) = \text{Cov}\left(\left(I - \tilde{B}\right)^{-1} E, E\right) = \left(I - \tilde{B}\right)^{-1} \text{Var}\,(E) = \left(I - \tilde{B}\right)^{-1} \Lambda$$

which is not a diagonal matrix in general; hence Z_i and E_j may be dependent for $i \neq j$.

4. Related Random Fields

Note 43. We introduce general modeling structures of basic building models which are computationally convenient yet reasonable for use in spatial statistics models. 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 Ising model, CAR, and SAR are just particular cases of such models.

4.1. Gibbs Random Fields.

Notation 44. 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 45. 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

$$U_{\mathcal{A}}^{\mathcal{V}}(z) = \sum_{\{\forall \mathcal{B} \subseteq \mathcal{S} : \mathcal{A} \cap \mathcal{B} \neq \emptyset\}} V_{\mathcal{B}}(z_{\mathcal{B}}) \tag{4.1}$$

exists.

Definition 46. The function $V_{\mathcal{A}} : \mathcal{Z}^{\mathcal{A}} \to \mathbb{R}$ in Def 45 is called potential on $\mathcal{A}$.

Definition 47. The function $U_{\mathcal{A}}^{\mathcal{V}}(z)$ in (4.1) in Def 45 is called energy function of interaction potential $\mathcal{V}$ on $\mathcal{A}$ is called.

Definition 48. 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 49. This allow as to define a distribution corresponding to the energy.

Definition 50. 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

$$\text{pr}_Z\left(z_{\mathcal{A}} | z_{\mathcal{S}\setminus\mathcal{A}}\right) = \frac{1}{C_{\mathcal{A}}^{\mathcal{V}}\left(z_{\mathcal{S}\setminus\mathcal{A}}\right)} \exp\left(\underbrace{\sum_{\{\mathcal{B}\subseteq\mathcal{S}\,:\,\mathcal{A}\cap\mathcal{B}\neq\emptyset\}} V_{\mathcal{B}}\left(z_{\mathcal{B}}\right)}_{=U_{\mathcal{A}}^{\mathcal{V}}(z)}\right), \quad z \in \mathcal{Z}^{\mathcal{S}} \tag{4.2}$$

Definition 51. The normalizing integral $C_{\mathcal{A}}^{\mathcal{V}}$ in (4.2) is called partition function.

Notation 52. Obviously for the marginal $\text{pr}_Z\left(z_{\mathcal{S}}\right)$ we will denote for $z \in \mathcal{Z}^{\mathcal{S}}$

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

where $C_{\mathcal{S}}^{\mathcal{V}} < \infty$ is the constant. For easy of the notation, in this case, we can omit $\cdot_{\mathcal{S}}^{\mathcal{V}}$ and just write

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

Example 53. (Ising model) In Ex 12, the Ising model has non-zero 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 energy function conditional on $\mathcal{S} \backslash \mathcal{B}$

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

Identifiability of the potential.

Definition 54. The interaction potential $\mathcal{V}$ is said to be normalized with respect to $\zeta \in \mathcal{Z}$ if there is $i \in \mathcal{S}$ which for any for any $z \in \mathcal{Z}^{\mathcal{S}}$with $z_i = \zeta$ implies that $V_{\mathcal{B}}(z) = 0$.

Note 55. The mapping $\mathcal{V} \to \mathrm{pr}_Z$ in (4.2) is non-identifiable as 4.2 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 56. One way to make $\mathcal{V}$ identifiable is to impose restriction

$$\forall \mathcal{A} \neq \emptyset,\ V_{\mathcal{A}}(z) = 0,\ \text{if for some}\ \ i \in \mathcal{A},\ z_i = \zeta \tag{4.3}$$

This follows from the following theorem which uniquely associates potentials satisfying (4.3) with (4.2).

Notation 57. 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 58. For instance if $z \in \mathcal{Z}^{\mathcal{S}}$ where $\mathcal{S} = \{1, ..., n\}$ then

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

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

Theorem 59. *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*

$$V_{\mathcal{A}}\left(z_{\mathcal{A}}\right)=\sum_{\mathcal{B}\subseteq\mathcal{A}}\left(-1\right)^{Card\left(\mathcal{A}\backslash\mathcal{B}\right)}U_{\mathcal{B}}^{\mathcal{V}}\left(z^{[\mathcal{B},\zeta]}\right),\ z\in\mathcal{Z}^{\mathcal{S}}$$
$$=\sum_{\mathcal{B}\subseteq\mathcal{A}}\left(-1\right)^{Card\left(\mathcal{A}\backslash\mathcal{B}\right)}\log\left(pr_{Z}\left(z^{[\mathcal{B},\zeta]}\right)\right),\ z\in\mathcal{Z}^{\mathcal{S}} \tag{4.4}$$

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[2], and hence out of scope. ☐

Corollary 60. *From Thm 59, for all $i\in\mathcal{A}$ it is*

$$V_{\mathcal{A}}\left(z_{\mathcal{A}}\right)\sum_{\mathcal{B}\subseteq\mathcal{A}}\left(-1\right)^{Card\left(\mathcal{A}\backslash\mathcal{B}\right)}\log\left(pr_{Z}\left(z_i^{[\mathcal{B},\zeta]}|z_{\mathcal{S}\backslash\{i\}}^{[\mathcal{B},\zeta]}\right)\right),\ z\in\mathcal{Z}^{\mathcal{S}} \tag{4.5}$$

Note 61. The following example explains the use of Thm 59 regarding the Def 45.

Example 62. 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 (from (4.1))

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

where (from (4.4)) it is

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

[2]Rota, G. C. (1964). On the foundations of combinatorial theory: I. Theory of Möbius functions. In Classic Papers in Combinatorics (pp. 332-360). Boston, MA: Birkhäuser Boston.

Example 63. (Ising model) Revisiting Ex 12, w.r.t Theorem 59. Consider $\zeta = 0$. Note that we use Notation 57, for instance,

$$z^{[\emptyset,\zeta]} = \Bigg(\overbrace{\zeta,...,\zeta}^{n\,\text{times}}\Bigg)^{\top};$$

$$z^{[\{i\},\zeta]} = \Bigg(\zeta,...,\zeta,\overbrace{z_i}^{\substack{i\text{th location}\\ \downarrow}},\zeta,...,\zeta\Bigg)^{\top};$$

$$z^{[\{i,j\},\zeta]} = \Bigg(\zeta,...\zeta,\overbrace{z_i}^{\substack{i\text{th location}\\ \downarrow}},\zeta,...,\zeta,\overbrace{z_j}^{\substack{j\text{th location}\\ \downarrow}},...,\zeta\Bigg)^{\top}$$

By using Thm 59, $V_{\emptyset} = 0$, and for any $i \in \mathcal{S}$, it is

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

for any $i, j \in \mathcal{S}$, with $i \sim j$ it is

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

Obviously, for any $i, j \in \mathcal{S}$, with $i \nsim j$ it is $V_{\{i,j\}}(z) = 0$, and for $\text{card}(\mathcal{A}) > 2$ it is $V_{\mathcal{A}}(z) = 0$.

4.2. **Markov Random Fields.**

Note 64. Recall the Ising model whose sites are 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 $\sim$ or when $\mathcal{A}$ a singleton. Consequently, its local characteristics $\text{pr}_i\left(z_i | z_{\mathcal{S}\setminus\{i\}}\right)$ depend only on the values of the sites $j \in \mathcal{S}\setminus\{i\}$ that satisfy $\sim$.

Note 65. Regarding spatial modeling, $\sim$ can describe adjacent sites which is in accordance to "dogma" that *near things are more related than distant things.* Also it is computationally convenient for big data problems (large number of sites) as it introduces sparsity and allows specialized numerical algorithms to be implemented.

Note 66. 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.

Definition 67. We define as the boundary of $\mathcal{A}$, $\mathcal{A} \subseteq \mathcal{S}$, for a given relation $\sim$ the set

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

Definition 68. 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 $Z_{\mathcal{A}\backslash\mathcal{S}} \in \mathcal{Z}_{\mathcal{A}\backslash\mathcal{S}}$the distribution of Z on $\mathcal{A}$ conditional on $Z_{\mathcal{A}\backslash\mathcal{S}}$ only depends on $Z_{\partial\mathcal{A}}$ (i.e. the configuration of Z on the neighborhood boundary of $\mathcal{A}$) i.e.

$$\mathrm{pr}_Z\left(z_{\mathcal{A}}|z_{\mathcal{S}\backslash\mathcal{A}}\right) = \mathrm{pr}_Z\left(z_{\mathcal{A}}|z_{\partial\mathcal{A}}\right) \tag{4.6}$$

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

Note 69. Def 68 implies that (4.6) becomes

$$\mathrm{pr}_Z\left(z_i|z_{-i}\right) = \mathrm{pr}_Z\left(z_i|z_{\partial\{i\}}\right), \quad \forall i \in \mathcal{S} \tag{4.7}$$

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

Definition 70. 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.

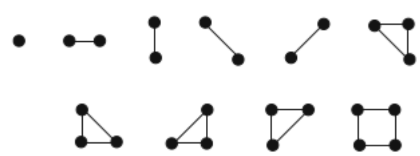

FIGURE 4.1. Examples of cliques

Note 71. The following theorem shows that the distribution of any Markov random field such that $\mathrm{pr}_Z(z)$ is positive can be expressed in terms of interactions between neighbors.

Theorem 72. *(Hammersley–Clifford) Let Z be an $\mathcal{Z}$-valued random field on a finite collection $\mathcal{S} \neq \emptyset$ of sites such that $pr_Z\left(z_{\mathcal{A}}|z_{\mathcal{C}\backslash\mathcal{A}}\right) > 0$ 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*

$$pr_Z(z) = \prod_{\mathcal{C} \in \boldsymbol{\mathcal{C}}} \varphi_{\mathcal{C}}\left(z_{\mathcal{C}}\right) \tag{4.8}$$

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 Thm 59, Z is Gibbs with a canonical potential (4.4)*

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

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,

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

Rearranging I get

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

Because $s\not\sim t$, it is $pr_Z\left(z_s^{\mathcal{B},\delta}|z_{\mathcal{S}\backslash s}^{\mathcal{B},\delta}\right)=pr_Z\left(z_s^{\mathcal{B}\cup\{t\},\delta}|z_{\mathcal{S}\backslash s}^{\mathcal{B}\cup\{t\},\delta}\right)$ and $pr_Z\left(z_s^{\mathcal{B}\cup\{s,t\},\delta}|z_{\mathcal{S}\backslash s}^{\mathcal{B}\cup\{s,t\},\delta}\right)=pr_Z\left(z_s^{\mathcal{B}\cup\{s\},\delta}|z_{\mathcal{S}\backslash s}^{\mathcal{B}\cup\{s\},\delta}\right)$. This implies $V_{\mathcal{A}}(z)=0$ for any subset $\mathcal{A}$ with $\mathcal{A}\subseteq\mathcal{S}$ which is not a clique. Hence (4.8) holds.

for $\Longleftarrow$: *By using (4.2), I can write*

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

where

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

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

Note 73. Because $\mathrm{pr}_Z(z)>0$, the Markov Random Field in (4.8) is a Gibbs Random Field as

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

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

Note 74. Essentially Thm 72, says 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 $\text{pr}_Z(\cdot)$ 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 75. (Ising model) Revisiting Ex 12... The joint PMF 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

$$\begin{aligned}
\varphi_{\emptyset} &= 1/\sum_{z\in\mathcal{Z}^{\mathcal{S}}} \exp\left(\alpha \sum_{i\in\mathcal{S}} z_i + \beta \sum_{\{i,j\}:i\sim j} z_i z_j\right) \\
\varphi_{\{i\}}\left(z_{\{i\}}\right) &= \exp(\alpha z_i), \qquad \forall i \in \mathcal{S} \\
\varphi_{\{i,j\}}\left(z_{\{i,j\}}\right) &= \exp(\beta z_i z_j), \qquad \forall i,j \in \mathcal{S} \quad \text{s.t.} \quad i \sim j \\
\varphi_{\{i,j\}}\left(z_{\{i,j\}}\right) &= 1, \qquad \forall i,j \in \mathcal{S} \quad \text{s.t.} \quad i \nsim j \\
\varphi_{\mathcal{A}}(z_{\mathcal{A}}) &= 1, \qquad \forall \mathcal{A} \subset \mathcal{S} \quad \text{s.t.} \quad \text{card}(\mathcal{A}) > 2
\end{aligned}$$

where $\{i\}$ and $\{i,j\}$ satisfying $i \sim j$ are cliques. Notice that $\varphi_{\emptyset}$ is just the constant term that can be absorbed to the other $\varphi's$ correspond to cliques.