

SPATIAL STATISTICS

空间统计学

GUANXU (GLEASON) WANG

University of Glasgow

Notes and Past Papers of STAT5012 Spatial Statistics, UofG

2025 年 5 月 16 日

目录

1	Introduction	1
2	Geostatistics (地统计学)	5
2.1	Geostatistical theory (理论)	5
2.1.1	Means, variances and covariances (均值, 方差和协方差)	5
2.1.2	Stationarity and isotropy (平稳性与各向同性)	7
2.1.3	Gaussian process (高斯过程)	9
2.1.4	Valid covariance function (协方差函数的有效性判断)	10
2.1.5	Variogram (变异函数)	11
2.2	Autocovariance and semi-variogram models (自协方差和半变异模型)	14
2.3	Geostatistical modelling (建模)	18
2.3.1	Overall modelling strategy (建模策略)	18
2.3.2	Exploratory data analysis (EDA)	20
2.3.3	Plotting the data (制图)	20
2.3.4	Modelling the mean function (均值函数建模)	21
2.3.5	Investigating the presence of spatial autocorrelation (调查空间自相关是否存在)	21
2.4	Parameter estimation (参数估计)	27
2.5	Model adequacy and comparison (模型适宜度与比较)	31
2.6	Spatial prediction (空间预测)	33
2.6.1	Kriging (克里金插值)	34

3	Areal unit processes (区域单元过程)	41
3.1	Exploratory Analysis for Areal Unit Data (EDA)	41
3.1.1	Mapping (制图)	41
3.1.2	Defining spatial closeness (空间邻近性)	42
3.1.3	Estimating the extent of spatial autocorrelation (估计空间自相关程度)	44
3.2	Gaussian Markov random field theory (高斯马尔可夫随机场理论)	49
3.2.1	Hammersley-Clifford theorem (GMRF 的充分条件)	51
3.2.2	Conditional autoregressive (CAR) models (条件自回归模型)	54
3.3	Modelling areal unit data (建模)	60
3.3.1	Modelling framework (建模框架)	60
3.3.2	Parameter estimation (参数估计)	62
3.3.3	An alternative estimation approach (参数估计的替代办法)	62
3.4	Extensions to the ICAR model (内在条件自回归模型的扩展)	63
4	Point Process (点过程)	67
4.1	Point process theory (理论)	69
4.2	Modelling and testing for complete spatial randomness (CSR) (完全随机的建模与测试)	71
4.2.1	A model for complete spatial randomness (CSR 的建模)	71
4.2.2	Testing for complete spatial randomness (CSR 的测试)	75
4.3	Modelling spatially varying intensities (空间变化强度的建模)	79
4.3.1	Model and data likelihood (模型与数据似然)	79
4.3.2	Non-parametric estimation of $\lambda(\mathbf{s})$ (强度函数的非参数估计)	81
4.3.3	Parametric models for $\lambda(\mathbf{s})$ (强度函数的参数模型)	82

4.4	Overview of models for clustered point processes (聚类点过程的模型)	83
4.4.1	Processes with offspring (后代)	83
4.4.2	Processes generated by thinning (稀疏/减薄/抽稀)	83
4.4.3	Log Gaussian Cox processes	84
5	Supplement	85
5.1	OLS	85

1 Introduction

定义 1.1 Spatial stochastic process (空间随机过程)

A **spatial stochastic process** is a family of random variables

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\},$$

indexed by spatial locations $\mathbf{s} \in D$.

- D 是过程的**空间域 (spatial domain)**, 即可以进行的观测的地理区域;
- $Z(\mathbf{s})$ 是一个随机变量, 表示在位置 \mathbf{s} 处测量的量.

存在三种不同的空间数据类型, 每种类型的统计模型都不同, 且在**空间域** D 的定义上差异很大.

定义 1.2 Geostatistical process (地统计学过程)

A **geostatistical process** is also known as a **point-referenced process (点参考过程)**, and is the stochastic process

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\},$$

其中 D 是二维空间 \mathbb{R}^2 的一个固定子集. 数据可能发生的位置 \mathbf{s} 在 D 上连续变化. 然而, 在实践中, 数据是在有限的位置上观测的, 相应的随机变量表示为 $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$.

地统计分析的目标

1. 找到一个能够充分解释数据中观察到的空间趋势和依赖性的统计模型.

2. 预测未测量位置的空间过程（以及不确定性度量），并创建整个研究区域的过程预测图.
3. 使用预测地图来对正在研究的过程进行推断。
4. 估计在空间上与响应位置相同的协变量测量的影响。

定义 1.3 Areal unit process (区域单元过程)

Areal unit process is also known as **lattice process (格点过程)**, 空间域 D 被划分为 n 个不重叠 (non-overlapping, 不相交, disjoint) 的面元, 记为

$$D = \{B_1, B_2, \dots, B_n\}.$$

这一组 areal units 可以形成一个规则的 grid (lattice), 也可以是不规则的形状单元, 但必须满足

- $\bigcup_{i=1}^n B_i = D$ (即区域被划分成的子域没有遗漏), 和
- $B_i \cap B_j = \emptyset$ for each $i \neq j$ (即不重叠).

根据这个研究区域划分, areal unit process 是一个随机过程

$$\mathbf{Z} = (Z(B_1), \dots, Z(B_n))$$

只定义在 n 个 areal units $\{B_1, B_2, \dots, B_n\}$ 上, 而不是无限多可能的位置上.

有时会使用另一种表述来表示 areal unit process, 即

$$\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$$

其中, 每个 $\mathbf{s}_i \in B_i \subset D$ 表示代表该区域 B_i 的位置, 一般用其中心点 (central point, 质心, centroid).

区域单元分析的目标

1. 找到一个能够充分解释数据中观察到的趋势和依赖关系的统计模型, 以理解其空间模式. 对于疾病数据, 这个研究领域通常被称为疾病制图 (disease mapping).
2. 使用统计模型来估计暴露对反应的影响, 同时考虑到残差在空间上是自相关的. 这个研究领域通常被称为生态回归 (ecological regression).

定义 1.4 Point process (点过程)

考虑一个空间域 D , 并令 $A \subset D$. 然后令 $Z(A)$ 表示在 A 中出现的点的数量. 因此集合

$$Z = \{Z(A) : A \subset D\},$$

是一个**空间点过程**.

空间点过程有两种类型:

- 对于**无标记 (unmarked) 点过程**, 仅记录数据的位置.
- 对于**标记 (marked) 点过程**, 记录数据的位置, 以及每个位置的相关测量值.

点过程分析的目标

1. 可视化和描述数据中的空间依赖结构.
2. 拟合模型以识别任何有许多点聚集在一起形成集群的位置.

3. 估计协变量对数据点位置的影响.

2 Geostatistics (地统计学)

定义 2.1 Geostatistical process (地统计学过程)

A **geostatistical process** is also known as a **point-referenced process** (点参考过程), and is the stochastic process

$$Z(\mathbf{s}) : \mathbf{s} \in D,$$

其中 D 是二维空间 \mathbb{R}^2 的一个固定子集. 数据可能发生的位置 \mathbf{s} 在 D 上连续变化. 然而, 在实践中, 数据是在有限的位置上观测的, 相应的随机变量表示为 $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$.

2.1 Geostatistical theory (理论)

2.1.1 Means, variances and covariances (均值, 方差和协方差)

定义 2.2 Mean function

The **mean function** of the stochastic process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is

$$\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})], \quad \forall \mathbf{s} \in D.$$

它是位置 \mathbf{s} 的理论均值 (期望), 是在随机过程 $Z(\mathbf{s})$ 可能生成的所有值的分布上取的.

$$\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})] = \int_{-\infty}^{\infty} z(\mathbf{s})f(z(\mathbf{s})) \, dz(\mathbf{s}) \quad \text{或者} \quad \mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})] = \sum_{z_i(\mathbf{s}) \in \mathcal{S}} z_i(\mathbf{s})f(z_i(\mathbf{s})).$$

定义 2.3 Autocovariance function

The **autocovariance function** of the stochastic process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is

$$\begin{aligned} C(\mathbf{s}, \mathbf{t}) &= \text{Cov}[Z(\mathbf{s}), Z(\mathbf{t})] \\ &= \mathbb{E}[(Z(\mathbf{s}) - \mu(\mathbf{s}))(Z(\mathbf{t}) - \mu(\mathbf{t}))] \\ &= \mathbb{E}[Z(\mathbf{s})Z(\mathbf{t})] - \mu(\mathbf{s})\mu(\mathbf{t}), \end{aligned}$$

for 2 locations (\mathbf{s}, \mathbf{t}) . 自协方差衡量了 $Z(\mathbf{s})$ 和 $Z(\mathbf{t})$ 之间的线性依赖性的强度.

The **variance function** of $Z(\mathbf{s})$ is the special case of the autocovariance with $\mathbf{s} = \mathbf{t}$, which gives

$$\begin{aligned} \text{Var}[Z(\mathbf{s})] &= C(\mathbf{s}, \mathbf{s}) = \text{Cov}[Z(\mathbf{s}), Z(\mathbf{s})] \\ &= \mathbb{E}[(Z(\mathbf{s}) - \mu(\mathbf{s}))^2] \\ &= \nu(\mathbf{s}). \end{aligned}$$

定理 2.1 Symmetric (对称性)

The autocovariance function is symmetric in its arguments, that is $C(\mathbf{s}, \mathbf{t}) = C(\mathbf{t}, \mathbf{s})$ for each $\mathbf{s}, \mathbf{t} \in D$.

定理 2.2 Nonnegative definite (非负定性)

The autocovariance function is $C(\mathbf{s}, \mathbf{t})$ is a nonnegative definite function, 即其满足

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0.$$

定义 2.4 Autocorrelation function

The autocorrelation function of $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is given by

$$\rho(\mathbf{s}, \mathbf{t}) = \text{Corr}[Z(\mathbf{s}), Z(\mathbf{t})] = \frac{C(\mathbf{s}, \mathbf{t})}{\sqrt{C(\mathbf{s}, \mathbf{s})C(\mathbf{t}, \mathbf{t})}}.$$

自相关函数衡量 $Z(\mathbf{s})$ 和 $Z(\mathbf{t})$ 之间的线性相关性的强度, 并且只是自协方差函数的缩放版本.

2.1.2 Stationarity and isotropy (平稳性与各向同性)

定义 2.5 Strictly Stationary (严平稳)

A geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is **strictly stationary** if

$$f(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) = f(Z(\mathbf{s}_1 + \mathbf{h}), \dots, Z(\mathbf{s}_n + \mathbf{h})).$$

for any displacement (位移, spatial lag) vector \mathbf{h} and any set of n locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$. 本质上这意味着一组随机变量的联合分布不受空间平移的影响.

定义 2.6 Weakly Stationary (宽平稳)

A geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is **weakly stationary** if

1. $\mathbb{E}[Z(\mathbf{s})] = \mu(\mathbf{s}) = \mu$ for some finite constant μ which does not depend on \mathbf{s} .
2. $\text{Cov}[Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})] = C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(\mathbf{h}) < \infty$, a finite constant that can depend on \mathbf{h} but not on \mathbf{s} .

定义 2.7 Isotropic (各向同性)

A weakly stationary geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is **isotropic** if the autocovariance function $C(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ can be further simplified to

$$C(\mathbf{h}) = C(h = \|\mathbf{h}\|),$$

where $h = \|\mathbf{h}\|$ denotes the length of the lag vector \mathbf{h} as measured by its Euclidean distance. That is in two dimensions for $\mathbf{h} = (h_1, h_2)$ we have that $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2}$.

如果自协方差/自相关函数是**宽平稳**且**各向同性**的, 它们就变成一个关于一维量 $h = (h_1, h_2)$ 的函数, 即所讨论的两个点之间的距离.

2.1.3 Gaussian process (高斯过程)

定义 2.8 Gaussian process (高斯过程)

A geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is a **Gaussian process**, if 对于任何 n 个位置的集合 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, 随机变量 $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$ 的联合分布是多元高斯 (正态) 分布, 即

$$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

for a mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$.

我们将重点关注在通过回归模型去除任何空间趋势后, 对**高斯宽平稳各向同性过程**进行建模, 这意味着表示在 n 个位置的随机过程 $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$ 服从多元高斯分布, 其矩为:

$$\mathbb{E}[Z(\mathbf{s})] = \mu,$$

$$\text{Var}[Z(\mathbf{s})] = \nu^2,$$

$$C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(h = \|\mathbf{h}\|) \Leftrightarrow \rho(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \rho(h = \|\mathbf{h}\|).$$

From this the joint probability density function is given by

$$f(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})\right)$$

where $\boldsymbol{\mu} = (\mu, \dots, \mu)_{n \times 1}^\top$ and the (j, k) element of the autocovariance matrix $\boldsymbol{\Sigma}$ is $\Sigma_{jk} = C(\|\mathbf{s}_j - \mathbf{s}_k\|)$.

2.1.4 Valid covariance function (协方差函数的有效性判断)

定理 2.3 Bochner's theorem (博克纳定理)

Define mathematically the two conditions that a general function $C(\mathbf{h})$ must satisfy to be a **valid covariance function** for a weakly stationary process.

1. The function $C(\mathbf{h})$ must be an even function as $C(-\mathbf{h}) = C(\mathbf{h})$ for all \mathbf{h} ;
2. The function $C(\mathbf{s}, \mathbf{s} + \mathbf{h})$ must be non-negative definite, which means that

$$C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \sum_{j=1}^n \sum_{k=1}^n a_j a_k C(\mathbf{s}_j, \mathbf{s}_k) \geq 0,$$

for any scalars (a_1, \dots, a_n) , and any locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$.

类似地, 自相关函数 $\rho(\mathbf{h})$ 的有效性亦为如此.

例 2.1 Bookwork 2.2.4

Consider a weakly stationary and isotropic mean zero white noise process with autocovariance function

$$C(h) = \begin{cases} \tau^2, & \text{if } h = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Solution. This is clearly a valid autocovariance function under Bochner's theorem, because:

1. It is even as $C(h) = C(-h) = 0$ for $h \neq 0$.

2. Consider n scalars (a_1, \dots, a_n) and n locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, and let $h_{jk} = \|\mathbf{s}_j - \mathbf{s}_k\|$. Then

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n a_j a_k C(\mathbf{s}_j, \mathbf{s}_k) &= \sum_{j=1}^n \sum_{k=1}^n a_j a_k C(h_{jk}) \\ &= \sum_{j=1}^n a_j^2 \tau^2 \geq 0. \end{aligned}$$

hence it is non-negative definite.

2.1.5 Variogram (变异函数)

定义 2.9 Semi-variogram (半变异函数)

A **semi-variogram** of a geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is a function denoted by $\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h})$, and measures the variance of the difference in the process at two spatial locations \mathbf{s} and $\mathbf{s} + \mathbf{h}$. It is defined as:

$$\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \frac{1}{2} \text{Var} [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})],$$

where traditionally $2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h})$ is called the **variogram** and $\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h})$ is called the **semi-variogram**.

使用协方差规则, 我们可以将半变异图与自协方差函数关联起来, 如下所示:

$$\begin{aligned} \gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \frac{1}{2} \text{Var} [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})] \\ &= \frac{1}{2} \text{Cov} [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})] \\ &= \frac{1}{2} (C(\mathbf{s}, \mathbf{s}) - 2C(\mathbf{s}, \mathbf{s} + \mathbf{h}) + C(\mathbf{s} + \mathbf{h}, \mathbf{s} + \mathbf{h})). \end{aligned}$$

当过程是宽平稳时, 即 $C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(\mathbf{h})$, 我们可以将其简化为

$$\begin{aligned}\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \gamma(\mathbf{h}) = \frac{1}{2}(2C(\mathbf{0}) - 2C(\mathbf{h})) \\ &= C(\mathbf{0}) - C(\mathbf{h}) \\ &= \nu^2 - C(\mathbf{h}),\end{aligned}$$

并且在各向同性条件下, 即 $C(\mathbf{h}) = C(h = \|\mathbf{h}\|)$, 我们得到

$$\gamma(\mathbf{h}) = \gamma(h) = C(0) - C(h) = \nu^2 - C(h),$$

其中, $h = \|\mathbf{h}\|$, 以及 $\nu^2 = \text{Var}[Z(\mathbf{s})]$.

例 2.2 from 2021.Q1(a)-iii

Derive the semi-variogram that corresponds to the covariance function

$$C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(h = \|\mathbf{h}\|) = \begin{cases} \tau^2, & \text{if } h = 0, \\ \frac{\theta}{h+1}, & \text{otherwise.} \end{cases}$$

Solution.

- For $h = 0$, we have that $\gamma(0) = C(0) - C(0) = 0$.
- For $h > 0$, we have that $\gamma(h) = C(0) - C(h) = \tau^2 - \frac{\theta}{h+1}$.

Thus we have that

$$\gamma(h) = \begin{cases} \gamma^2 - \frac{\theta}{h+1}, & h > 0, \\ 0, & h = 0. \end{cases}$$

例 2.3

Consider the following semi-variogram model for an isotropic geostatistical process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$,

$$\gamma_Z(h) = \begin{cases} 0, & h = 0, \\ h^2, & h > 0, \end{cases}$$

which is accompanied by the mean model $\mu_Z(\mathbf{s}) = 0$ for all $\mathbf{s} \in D$. Is this geostatistical process weakly stationary? Justify your answer.

Solution. 宽平稳性要求两个条件:

- **Mean constant:** mean model $\mu_Z(\mathbf{s}) = 0$ for all $\mathbf{s} \in D$.
- **Covariance finite constantly and only depend on h :** $C(h) = C(0) - \gamma_Z(h) = C(0) - h^2$. As h increases, $C(h)$ tends towards negative infinity, leading to a negative covariance even potentially unbounded.

因此不满足宽平稳条件.

2.2 Autocovariance and semi-variogram models (自协方差和半变异模型)

性质 2.1 Parameters

In summary, the following parameters are often used to describe variograms:

- **Nugget** (块金): is the limiting value of the **semi-variogram** as $h \rightarrow 0$, from the right.
 - 如果没有 Nugget, 即 $\text{Nugget} = 0$, 则曲线从 $(0, 0)$ 平滑开始.
 - 如果存在 Nugget, 即 $\text{Nugget} > 0$, 曲线会在 $h = 0$ 处有一个垂直的跳跃, 然后再平滑增长.
- **Sill** (基台): 半变异函数的 **稳定值**. 如果没有稳定值, 即不收敛, 则过程不平稳.
- **Range** (变程): 使半变异函数达到 Sill 的 **距离** h .
- **Partial sill** = Sill - Nugget.

	用 $\gamma(h)$ 计算	用 $C(h)$ 计算
Nugget	$\lim_{h \rightarrow 0+} \gamma(h)$	Sill - Partial sill
Sill	$\lim_{h \rightarrow \infty} \gamma(h)$	$C(0)$
Range	$\gamma(h) = \text{Sill}$ 时的距离	$C(h) = 0$ 时的距离
Partial sill	Sill - Nugget	$\lim_{h \rightarrow 0+} C(h)$

例 2.4

Suppose that $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is a stationary isotropic geostatistical process with zero-mean and an autocovariance function given by

$$C(h) = \begin{cases} \sigma^2 \exp(-|h/\phi|^r), & h > 0, \\ \sigma^2 + \tau^2, & h = 0, \end{cases}$$

where $0 < r < 2$. What are the range, sill, nugget, and partial sill for this autocovariance model?

Solution. 我们可以得到:

$$\gamma(h) = C(0) - C(h) = \begin{cases} \sigma^2 + \tau^2 - \sigma^2 \exp(-|h/\phi|^r), & h > 0 \\ 0, & h = 0. \end{cases}$$

1. **Nugget:**

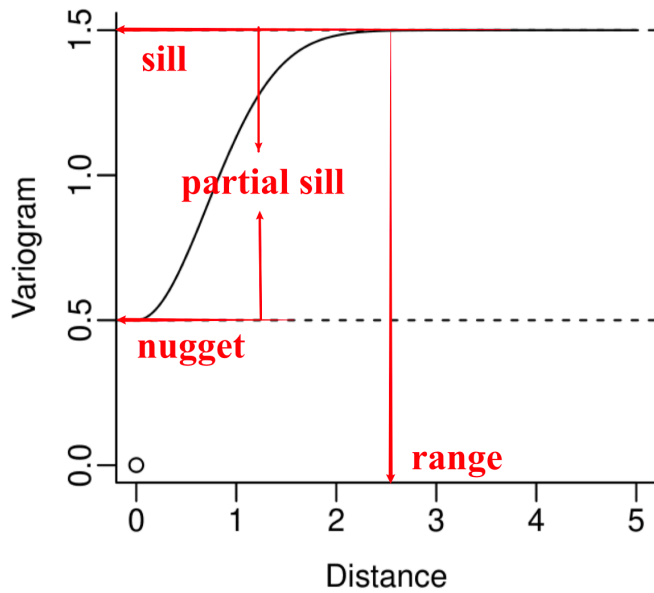
$$\lim_{h \rightarrow 0^+} \gamma(h) = \sigma^2 + \tau^2 - \sigma^2 = \tau^2.$$

2. **Sill:** 半变异函数的最终稳定值, 易知 $\gamma(h)$ 在 $h \rightarrow \infty$ 处收敛, 则

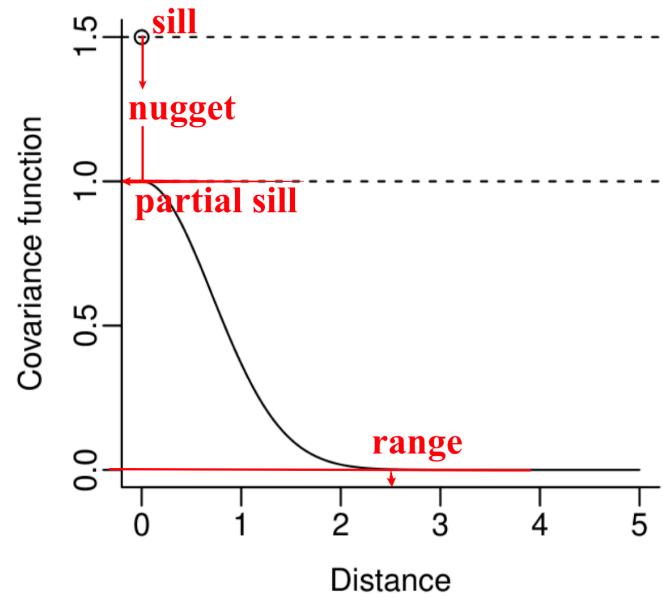
$$\lim_{h \rightarrow \infty} \gamma(h) = \sigma^2 + \tau^2 \quad \text{or} \quad C(0) = \sigma^2 + \tau^2.$$

3. **Range:** $\gamma(h)$ 在 $h \rightarrow \infty$ 处收敛到 Sill, 因此 the range is infinite.

4. **Partial sill** = Sill - Nugget = $\sigma^2 + \tau^2 - \tau^2 = \sigma^2$. Alternatively, $\lim_{h \rightarrow 0^+} C(h) = \sigma^2$.



(a) Variogram $\gamma(h)$



(b) Covariance function $C(h)$

定义 2.10 Exponential autocovariance function

The most commonly used parametric model.

$$C(h) = \begin{cases} \sigma^2 \exp(-h/\phi), & h > 0, \\ \tau^2 + \sigma^2, & h = 0. \end{cases}$$

定义 2.11 Gaussian semi-variogram model

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-h^2/\phi)), & h > 0, \\ 0, & h = 0. \end{cases}$$

定义 2.12 Spherical (球面) autocovariance function

$$C(h) = \begin{cases} 0, & h > \phi \\ \sigma^2 \left[1 - \frac{3}{2}(h/\phi) + \frac{1}{2}(h/\phi)^3 \right], & 0 < h \leq \phi, \\ \tau^2 + \sigma^2, & h = 0, \end{cases}$$

with semi-variogram

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2, & h > \phi \\ \tau^2 + \sigma^2 \left[\frac{3}{2}(h/\phi) - \frac{1}{2}(h/\phi)^3 \right], & 0 < h \leq \phi, \\ 0, & h = 0. \end{cases}$$

定义 2.13 Wave autocovariance function

$$C(h) = \begin{cases} \sigma^2 \left[\frac{\sin(h/\phi)}{(h/\phi)} \right], & h > 0, \\ \tau^2 + \sigma^2, & h = 0, \end{cases}$$

with semi-variogram

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{\sin(h/\phi)}{(h/\phi)} \right], & h > 0, \\ 0, & h = 0, \end{cases}$$

2.3 Geostatistical modelling (建模)

2.3.1 Overall modelling strategy (建模策略)

假设我们观察到以下地统计数据:

1. 响应数据 $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))$.
2. 协变量 (covariate) 风险因素 $\mathbf{x}(\mathbf{s}_i) = (x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i))$ 在每个位置 \mathbf{s}_i , 其中 $x_1(\mathbf{s}_i) = 1$ 对应于截距项, 所有 n 个位置的全部协变量都包含在设计矩阵 $\mathbf{X}_{n \times p}$ 中.

响应数据 \mathbf{z} 是地统计过程 $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ 的一个实现, 我们希望对 $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ 在任意位置 \mathbf{s} 的概率分布进行推断. 我们假设所建模的过程具有某种**空间平滑性 (spatial smoothness)**, 因为这允许我们在未测量位置进行过程预测. 一种常见的建模方法是加性分解:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}),$$

其中:

- $\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})] = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}$ 是空间变化的均值.
- $\epsilon(\mathbf{s})$ 是一个空间地统计误差过程.

假设 $\epsilon(\mathbf{s})$ 是高斯分布, 均值为零, 并且具有弱平稳和各向同性的自协方差函数 $C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(h)$.

考虑高斯模型:

$$\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta})),$$

其中:

- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ 是要估计的回归参数向量, 并形成线性回归模型.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ 是由自协方差函数 $C(h, \boldsymbol{\theta})$ 隐含的弱平稳各向同性协方差矩阵, 其中 $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$ 分别描述自相关结构的 partial sill, nugget 和 range.

例 2.5

考虑给定的指数自协方差函数

$$C(h) = \begin{cases} \sigma^2 \exp(-h/\phi), & h > 0, \\ \tau^2 + \sigma^2, & h = 0. \end{cases}$$

那么 \mathbf{Z} 对应的协方差矩阵 $\Sigma(\boldsymbol{\theta})$ 为

$$\Sigma(\boldsymbol{\theta}) = \sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I},$$

其中 $\mathbf{D} = (d_{ij})$ 是一个 $n \times n$ 的距离矩阵, 其值 $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, 并且 \mathbf{I} 是单位矩阵.

2.3.2 Exploratory data analysis (EDA)

大多数 EDA 技术强调数据的某些特征, 而隐藏其他特征, 因此在一个数据集上使用多种不同的 EDA 技术是一个好主意. EDA 有时被认为**琐碎 (trivial) 且不重要**, 但如果操作不当, 它可能会导致后续建模问题. EDA **很重要**.

2.3.3 Plotting the data (制图)

EDA 的第一阶段是绘制数据的空间地图, 有许多软件包可以做到这一点. 其中一个不错的叫做 `leaflet`, 因为它将数据叠加在 OpenStreetMap 上.

注意, 要使用 `leaflet`, 数据坐标系统必须是经度和纬度.

2.3.4 Modelling the mean function (均值函数建模)

该过程与非空间数据相同, 建立如下模型

$$\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})] = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} = \beta_1 + x_2(\mathbf{s}_i)\beta_2 + \cdots + x_p(\mathbf{s}_i)\beta_p$$

其中, $x_i(\mathbf{s}_i)$ 可以是经度 (longitude), 纬度 (latitude), 海拔 (elevation), 以及等等感兴趣的变量.

2.3.5 Investigating the presence of spatial autocorrelation (调查空间自相关是否存在)

一种调查数据集或残差 (来自拟合趋势模型) $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))$ 中空间自相关存在的方法是估计其经验半变异函数 (empirical semi-variogram).

空间过程 $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ 的半变异函数为

$$\begin{aligned}\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \frac{1}{2} \text{Var}[Y(\mathbf{s}) - Y(\mathbf{s} + \mathbf{h})] \\ &= \frac{1}{2} \mathbb{E} \left[(Y(\mathbf{s}) - Y(\mathbf{s} + \mathbf{h}))^2 \right],\end{aligned}$$

这里我们假设趋势被去除, 那么 $\mathbb{E}[Y(\mathbf{s}) - Y(\mathbf{s} + \mathbf{h})] = 0$. 这种估计的直观给出为

$$\gamma_{ij} = \frac{1}{2} [y(\mathbf{s}_i) - y(\mathbf{s}_j)]^2,$$

这样可以针对每对数据点 $(y(\mathbf{s}_i), y(\mathbf{s}_j))$ for $i \neq j$ 进行计算. 假设过程是平稳且各向同性的, 那么 γ_{ij} vs. $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ 的图称为 **variogram cloud**.

但是这图很难去看, 因为会有太多噪点. 因此考虑分组进行.

定义 2.14 Binned empirical semi-variogram (分组经验半变异函数)

假设我们将 all pairs of points 之间的距离范围划分为 K 个区间 (interval or bins)

$$I_k = (h_{k-1}, h_k], \quad k = 1, \dots, K,$$

其中, $0 = h_0 < h_1 < \dots < h_K$. 令 $h_k^m = \frac{h_{k-1} + h_k}{2}$ 表示表示第 k 个区间的中点 (midpoint). 然后令

$$N(h_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\},$$

表示距离落在 I_k 这个区间的点对 $(\mathbf{s}_i, \mathbf{s}_j)$ 的集合. Then the **binned empirical semi-variogram** is

$$\hat{\gamma}(h_k^m) = \frac{1}{|2N(h_k)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(h_k)} [y(\mathbf{s}_i) - y(\mathbf{s}_j)]^2.$$

用途：检验各向同性 将经验半变异函数在不同方向上分别绘制 (见章末).

1. 如果曲线在所有方向上相似, 则数据是**各向同性 (isotropy)** 的.
2. 如果不同方向的曲线形状差异大, 则数据存在**各向异性 (anisotropy)**.

定义 2.15 Monte Carlo envelope (蒙特卡洛包络)

The **Monte Carlo envelope** for the binned empirical semi-variogram based on independence is computed by repeating the following steps a large number, say $j = 1, \dots, 1000$ times.

For $j = 1, \dots, 1000$,

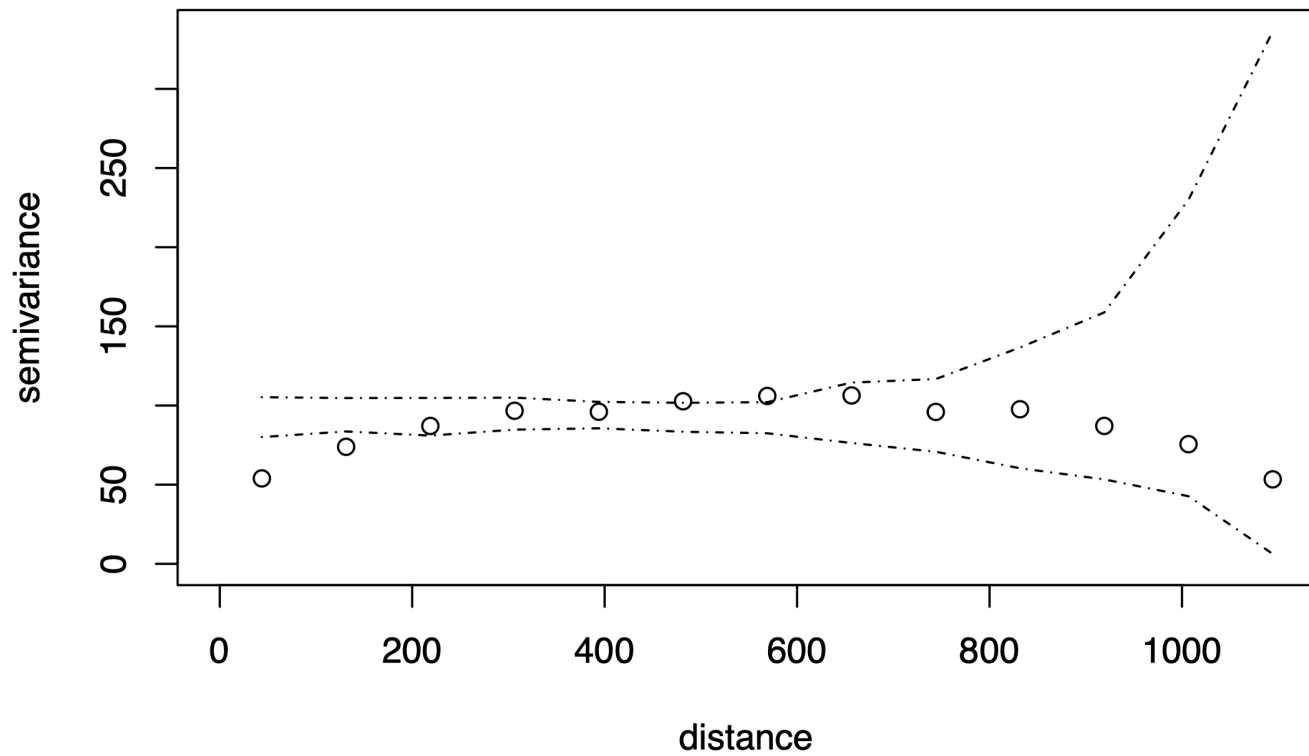
- (a) Create a pseudo data set $\mathbf{y}^{(j)} = (y^{(j)}(\mathbf{s}_1), \dots, y^{(j)}(\mathbf{s}_n))$ by randomly permuting (排列) the n data points to the n spatial locations, $\mathbf{s} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$.
- (b) Compute the binned empirical semi-variogram for each mid distance $(\hat{\gamma}^{(j)}(h_1^m), \dots, \hat{\gamma}^{(j)}(h_K^m))$ for this pseudo data set $\mathbf{y}^{(j)}$.

由于这些排列中的每一个都是 n 数据点到 n 位置的随机重新分配 (random re-allocation), 因此伪数据集 $\mathbf{y}^{(j)}$ 不应有空间自相关.

因此, 对于每个距离 h_i^m , 在假设**独立性**的情况下, 95% 的 $\hat{\gamma}(h_i^m)$ 包络计算为集合 $\{\hat{\gamma}^{(j)}(h_i^m)\}_{j=1}^{1000}$ 的 [2.5%, 97.5%] 百分位数.

.....
Thus if the empirical semi-variogram for the data lies **completely inside** these envelopes, then **independence** is likely. If it does **not** then **spatial autocorrelation** is likely.

There is clear evidence of **spatial autocorrelation** present because the empirical semi-variogram (dots) **lies outside** the Monte Carlo envelopes (lines) at the left side of the plot.



定义 2.16 Binned empirical covariogram

An estimate of the autocovariance function is called the **binned empirical covariogram** and is given by

$$\hat{C}(h_k^m) = \frac{1}{|2N(h_k)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(h_k)} [y(\mathbf{s}_i) - \bar{y}] [y(\mathbf{s}_j) - \bar{y}].$$

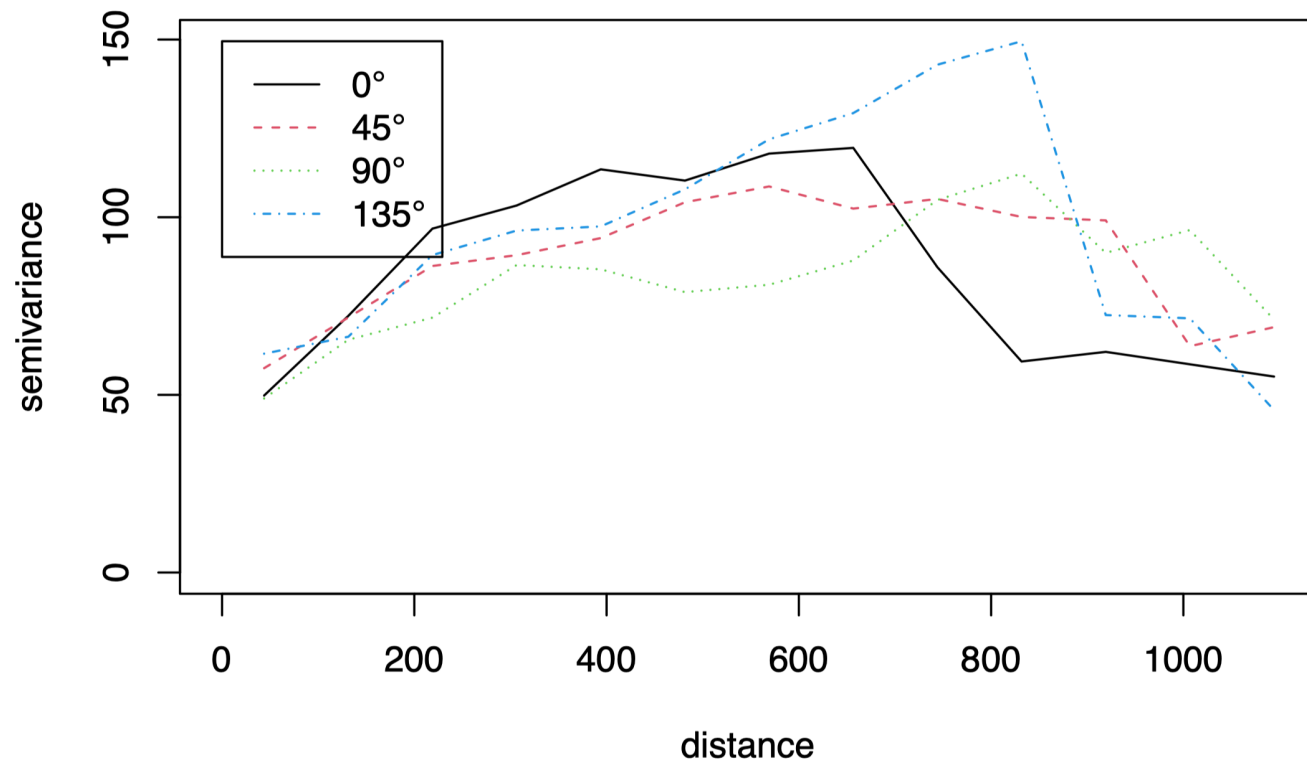
其中 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y(\mathbf{s}_i)$ 是数据的样本均值.

这个估计是有偏的, 但随着配对数量的增加, 偏差会减小.

到目前为止所有空间依赖性度量都假设**各向同性**, 即 autocovariance 或 semi-variogram 仅取决于数据位置之间的**距离**, 而与**方向**无关. 文献中有多种方法来**评估各向同性**是否是地质统计数据的合理假设. 最简单的方法是限制经验半变异图或协方差图中出现的数据点对, 以便我们仅测量沿某些方向 (或方向范围) 的依赖性. 我们称这种显示为**方向性变异图 (directional variograms, or covariograms)**. 如果过程是各向同性的, 那么无论我们朝哪个方向看, 变异图都应该相似 (同样需要注意我们没有不确定性度量).

我们可以计算拟合均值模型到数据残差的 4 个方向的 directional variograms, 如下所示. 这里的方向是: 0 度 (北到南 ↓), 45 度 (东北到西南 ↘); 90 度 (东到西 ←); 和 135 度 (西北到东南 ↙).

正如所见, 半变差图在左侧看起来有些相似 (右侧非常不确定), 因此各向同性在这里似乎是一个合理的假设.



2.4 Parameter estimation (参数估计)

对于我们使用的数据 $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))$ 有如下形式 (高斯模型):

$$\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta})),$$

因此, 有两种类型的参数需要估计:

- 均值模型中的参数 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$.
- 协方差模型中的参数 $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$.

参数估计有两种通用方法, 即 frequentist inference using maximum likelihood (or least squares) 和 Bayesian inference using Markov chain Monte Carlo simulation. 我们在此重点介绍最大似然估计的频率推断. 为了说明, 我们假设使用的是 exponential autocovariance 模型, 即

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I},$$

其中, 如前所述 $\mathbf{D} = (d_{ij})$ 是一个 $n \times n$ 的距离矩阵, 其值 $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, 并且 \mathbf{I} 是单位矩阵. 该多元高斯分布的似然函数由下式给出

$$\begin{aligned} f(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{z}) &= \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \\ &= (2\pi)^{-\frac{n}{2}} \left| \sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I} \right|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top \left[\sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I} \right]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right). \end{aligned}$$

Then after removing unnecessary constants $(2\pi)^{-\frac{n}{2}}$ the log-likelihood function is given by

$$\log [f(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{z})] = -\frac{1}{2} \log \left| \sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I} \right| - \frac{1}{2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top \left[\sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I} \right]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}).$$

由此很容易看出 (σ^2, τ^2) 被限制在一个逆矩阵和一个行列式内. 因此为了简化估计, 我们采用变换 $\xi^2 = \frac{\tau^2}{\sigma^2}$, 其中 ξ^2 被称为**信噪比 (noise to signal ratio)**. 用 $\tau^2 = \xi^2 \sigma^2$ 替换 τ^2 后, 对数似然函数此时为

$$\begin{aligned}\log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] &= -\frac{1}{2} \log \left| \sigma^2 \exp(-\mathbf{D}/\phi) + \xi^2 \sigma^2 \mathbf{I} \right| - \frac{1}{2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top \left[\sigma^2 \exp(-\mathbf{D}/\phi) + \xi^2 \sigma^2 \mathbf{I} \right]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log \left| \exp(-\mathbf{D}/\phi) + \xi^2 \mathbf{I} \right| - \frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top \left[\exp(-\mathbf{D}/\phi) + \xi^2 \mathbf{I} \right]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}).\end{aligned}$$

这是因为矩阵 \mathbf{A} 的行列式满足性质 $|\alpha \mathbf{A}| = \alpha^n |\mathbf{A}|$. 在以下内容中, 我们用 $\mathbf{V}(\phi, \xi^2) = \exp(-\mathbf{D}/\phi) + \xi^2 \mathbf{I}$ 表示, 以简化记号, 使得 $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \mathbf{V}(\phi, \xi^2)$, 并且对数似然函数给出

$$\log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log |\mathbf{V}(\phi, \xi^2)| - \frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}).$$

例 2.6 Estimation of $\boldsymbol{\beta}$ using MLE

仅保留对 $\boldsymbol{\beta}$ 依赖的似然函数部分, 我们得到

$$\begin{aligned}\log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] &= \frac{1}{2\sigma^2} \left(2\boldsymbol{\beta}^\top \mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{z} - \boldsymbol{\beta}^\top \mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{X}\boldsymbol{\beta} \right) \\ \frac{d}{d\boldsymbol{\beta}} \log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] &= \frac{1}{2\sigma^2} \left(2\mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{z} - 2\mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{X}\boldsymbol{\beta} \right).\end{aligned}$$

Setting the latter equal to zero and solving gives:

$$\hat{\boldsymbol{\beta}}(\phi, \xi^2) = \left(\mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top [\mathbf{V}(\phi, \xi^2)]^{-1} \mathbf{z}.$$

which depends on the covariance parameters (ϕ, ξ^2) but not on σ^2 .

例 2.7 Estimation of σ^2 using MLE

仅保留对 σ^2 依赖的似然函数部分, 我们得到

$$\begin{aligned}\log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{d}{d\sigma^2} \log [f(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z})] &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}).\end{aligned}$$

Setting the latter equal to zero and solving gives:

$$\hat{\sigma}^2(\boldsymbol{\beta}, \phi, \xi^2) = \frac{1}{n} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}).$$

虽然与标准线性模型一样, 但是 MLE 是有偏的, 因此使用替代方案

$$\hat{\sigma}^2(\boldsymbol{\beta}, \phi, \xi^2) = \frac{1}{n-p} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}),$$

其中 p 是均值模型中的参数数量.

例 2.8 Estimation of (ϕ, ξ^2)

用不了 MLE, 因为这俩参数都在矩阵里, 并进行了逆和行列式运算, 因此把 $\boldsymbol{\beta}$ 和 σ^2 的 MLE 代入.

$$\begin{aligned}\log [f(\phi, \xi^2|\mathbf{z})] &= -\frac{n}{2} \log(\hat{\sigma}^2) - \frac{1}{2} \log |\mathbf{V}(\phi, \xi^2)| - \frac{1}{2\hat{\sigma}^2} (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top [\mathbf{V}(\phi, \xi^2)]^{-1} (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= -\frac{n}{2} \log \left(\hat{\sigma}^2(\hat{\boldsymbol{\beta}}, \phi, \xi^2) \right) - \frac{1}{2} \log |\mathbf{V}(\phi, \xi^2)|.\end{aligned}$$

注意, 上述完整对数似然函数中的第三项 (二次型) 已被消去为常数. 这被称为 (ϕ, ξ^2) 的轮廓似然

(profile likelihood) 或简化似然 (reduced likelihood), 必须通过计算机使用数值优化方法来最大化.

因此, 在估计 (β, σ^2) 时, 不允许 $(\hat{\phi}, \hat{\xi}^2)$ 存在不确定性, 这可以被视为具有限制性. 顺便一提, 贝叶斯估计方法可以克服这一局限性.

例 2.9 95% confidence interval of $\hat{\beta}$

$$\begin{aligned}
 \text{Var} [\hat{\beta}] &= \text{Var} \left[(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{z} \right] \\
 &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \text{Var} [\mathbf{z}] \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \\
 &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \Sigma \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \\
 &\stackrel{\Sigma = \sigma^2 \mathbf{V}}{=} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} (\sigma^2 \mathbf{V}) \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \\
 &= \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \\
 &= \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}.
 \end{aligned}$$

from which 95% confidence intervals can be constructed as

$$\hat{\beta}_i \pm t_{0.975}(n-p) \times \sigma \sqrt{\left((\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \right)_{ii}}$$

- MLE: $\hat{\beta}(\hat{\phi}, \hat{\xi}^2) = \left(\mathbf{X}^\top [\mathbf{V}(\hat{\phi}, \hat{\xi}^2)]^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top [\mathbf{V}(\hat{\phi}, \hat{\xi}^2)]^{-1} \mathbf{z}$, 通过 $(\hat{\phi}, \hat{\xi}^2)$ 考虑空间自相关性.
- OLS: $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}$, 假设具有独立性.

这俩哪个好? 有三个标准指标来衡量这一点.

1. **Bias**: 真实值 β 与估计值之间的差异是多少.

$$\text{Bias}(\beta) = \mathbb{E} \left[\hat{\beta} \right] - \beta = \frac{1}{n} \sum_{j=1}^n \hat{\beta}_j - \beta.$$

2. **Root mean square error (RMSE, 均方误差)**: 估计值之间的变异程度是多少.

$$\text{RMSE}(\beta) = \mathbb{E} \left[\left(\hat{\beta} - \beta \right)^2 \right] = \sqrt{\frac{1}{n} \sum_{j=1}^n \left(\hat{\beta}_j - \beta \right)^2}.$$

3. **Coverage probability (覆盖率)**: 每个数据集都产生一个 β 的估计值和 95% 置信区间, 有多少百分比的 95% 置信区间包含真实值 β .

一般来说, **bias** 基本不受影响, MLE 的 **RMSE** 略好. 然而主要区别在于**覆盖率**, MLE 的覆盖率在 95% 左右基本正确. 基于独立性的 OLS 估计的覆盖率非常低, 表明其 95% 置信区间过于狭窄. 原因在于它假设数据是独立的, 因此你有 n 个独立的信息片段来估计 β . 实际上这些数据点是相关的, 因此你拥有的独立信息片段要少得多. 因此, 不确定性被低估了.

2.5 Model adequacy and comparison (模型适宜度与比较)

Overall model fit (整体模型拟合)

我们希望最小化 Kullback-Leibler 信息的估计值, 这是一种比较**真实模型**与每个我们拟合的**估计模型**的似然度量. 下是两种最常用的信息准则.

定义 2.17 Akaike information criterion (AIC)

$$\text{AIC} = -2\ell(\hat{\beta}, \hat{\sigma}^2, \hat{\phi}, \hat{\xi}^2) + 2q,$$

其中 q 是模型中的总参数数量, $\ell(\hat{\beta}, \hat{\sigma}^2, \hat{\phi}, \hat{\xi}^2)$ 是在参数估计处评估的对数似然函数。

定义 2.18 Bayesian information criterion (BIC)

$$\text{BIC} = -2\ell(\hat{\beta}, \hat{\sigma}^2, \hat{\phi}, \hat{\xi}^2) + \log(n)q.$$

Predictive performance using cross validation (使用交叉验证的预测性能)

上述信息准则衡量模型拟合**观测数据**的程度. 如果建模过程的目标是预测模型在未测量位置的预测能力, 那么拟合数据可能不是模型选择诊断的最佳选择.

Instead, 我们可以采用**交叉验证方法**, 这些方法量化了模型预测未测量位置上地质统计过程的能力. 交叉验证有多种形式, 这里描述的最简单的是**留一交叉验证 (leave-one-out cross validation)**.

Repeat the process for $i = 1, \dots, n$.

1. 构建数据集 $\mathbf{z}_{-i} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_{i-1}), z(\mathbf{s}_{i+1}), \dots, z(\mathbf{s}_n))$, 其中第 i 个观测值被移除.
2. 将所提出的地质统计模型应用于 \mathbf{z}_{-i} , 并使用它来预测缺失位置 \mathbf{s}_i 上的过程 (下一节将说明如何进行预测). 将此预测称为 $P_{\mathbf{z}_{-i}}(\mathbf{s}_i)$.

然后, 通过**均方根预测误差 (root mean square prediction error, RMSPE)** 总结模型的预测质量,

其定义为

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z(\mathbf{s}_i) - P_{z_{-i}}(\mathbf{s}_i))^2}.$$

最小化 RMSPE 的模型是最好的. 注意 RMSPE 是预测能力最常用的指标, 但也有许多其他指标.

2.6 Spatial prediction (空间预测)

地统计分析的主要目标之一是预测未测量位置 \mathbf{s}_0 处的过程. 如果这在未测量的规则网格 (regular grid) 上进行, 就可以生成研究区域感兴趣的底层过程 (underlying process) 的**地图 (map)**. 假设我们有以下数据模型

$$\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta})),$$

其中协方差矩阵 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ 由一个平稳且各向同性的自协方差函数 $C(h, \boldsymbol{\theta})$ 定义.

定义 2.19 Linear prediction operator (线性预测算子)

For some constant a_0 and prediction weights $\mathbf{a} = (a_1, \dots, a_n)$, the **linear prediction operator**, P , is defined as

$$P_{\mathbf{z}}(\mathbf{s}_0) = a_0 + \sum_{i=1}^n a_i z(\mathbf{s}_i).$$

预测的一个简单观点是, 由于数据在空间上是自相关的, 因此可以预期, 距离未测量位置 \mathbf{s}_0 更近的数据点的预测权重可能会更大.

定义 2.20 Inverse distance weighted predictor (反距离加权预测器)

根据距离更近的点受到的权重就更大的观点, 有

$$P_z(\mathbf{s}_0) = \frac{\sum_{i=1}^n \frac{z(\mathbf{s}_i)}{d_{i0}}}{\sum_{i=1}^n \frac{1}{d_{i0}}},$$

where $a_0 = 0$ and $a_i \propto \frac{1}{d_{i0}}$. Here $d_{i0} = \|\mathbf{s}_i - \mathbf{s}_0\|$, the distance between the i th data location \mathbf{s}_i and the prediction location \mathbf{s}_0 .

缺点:

- 权重的理论依据是什么 $\frac{1}{d_{i0}}$? 存在许多替代方案, 例如 $\exp(-d_{i0})$ 或 $\frac{1}{d_{i0}^2}$.
- 如何量化预测中的不确定性.

2.6.1 Kriging (克里金插值)

该方法基于推导出**最佳线性无偏预测 (Best Linear Unbiased Prediction, BLUP)**, 给定数据 $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))^T$. BLUP 可以通过选择 (a_0, \mathbf{a}) 来获得, 以最小化**均方预测误差 (mean squared prediction error, MSPE)**.

定义 2.21 Mean squared prediction error (MSPE)

$$\text{MSPE} = \mathbb{E} \left[\left(Z(\mathbf{s}_0) - P_z(\mathbf{s}_0) \right)^2 \right],$$

where $U(\mathbf{s}_0) = Z(\mathbf{s}_0) - P_z(\mathbf{s}_0)$ is the prediction error at \mathbf{s}_0 .

一般来说, 我们有一个随机变量向量 \mathbf{Y} (其现值已被观察到), 以及一个需要预测的随机变量 T (其现值未知). 此外, 设 $\hat{T} = t(\mathbf{Y})$ 是 T 的一个点预测器, 即任何关于观察到的随机变量 \mathbf{Y} 的函数. 那么, 如之前定义的, \hat{T} 的 MSPE 是

$$\text{MSPE}(\hat{T}) = \mathbb{E} \left[\left(\hat{T} - T \right)^2 \right],$$

其中期望是关于联合分布 (T, \mathbf{Y}) 的. 最优预测器的推导使用了以下两个结果。

定理 2.4

$\text{MSPE}(\hat{T})$ 在 $\hat{T} = \mathbb{E}[T|\mathbf{Y}]$ 时取最小值.

Proof. 根据全期望公式,

$$\text{MSPE}(\hat{T}) = \mathbb{E} \left[\left(\hat{T} - T \right)^2 \right] = \mathbb{E} \left[\mathbb{E} \left[\left(\hat{T} - T \right)^2 \middle| \mathbf{Y} \right] \right],$$

其中, 内层期望为

$$\mathbb{E} \left[\left(\hat{T} - T \right)^2 \middle| \mathbf{Y} \right] = \text{Var} \left[\hat{T} - T \middle| \mathbf{Y} \right] + \left(\mathbb{E} \left[\hat{T} - T \middle| \mathbf{Y} \right] \right)^2.$$

由于是 given \mathbf{Y} 的, 那么 $\hat{T} = t(\mathbf{Y})$ 则应为一常数. 因此上式简化为

$$\mathbb{E} \left[\left(\hat{T} - T \right)^2 \middle| \mathbf{Y} \right] = \text{Var} [T|\mathbf{Y}] + \left(\hat{T} - \mathbb{E} [T|\mathbf{Y}] \right)^2.$$

那么对上式取期望就得到了 MSPE, 即

$$\text{MSPE}(\hat{T}) = \mathbb{E}[\text{Var}[T|\mathbf{Y}]] + \mathbb{E}\left[\left(\hat{T} - \mathbb{E}[T|\mathbf{Y}]\right)^2\right].$$

第一项不依赖 predictor \hat{T} , 因此可以视为 constant, 那么它就是一个关于 \hat{T} 的函数. 因此若使 $\text{MSPE}(\hat{T})$ 最小, 则需使第二项 $\mathbb{E}\left[\left(\hat{T} - \mathbb{E}[T|\mathbf{Y}]\right)^2\right] = 0$, 即 $\hat{T} = \mathbb{E}[T|\mathbf{Y}]$.

引理 2.1 分块矩阵的逆 (非本课程内容 @ 高等代数)

分块矩阵 $\mathbf{X} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ 的逆矩阵 $\mathbf{X}^{-1} = \begin{pmatrix} \mathbf{F}_1 & -\mathbf{F}_1 \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{F}_1 & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{F}_1 \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \end{pmatrix}$,
其中, $\mathbf{F}_1 = (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1}$.

Proof. 假设 $\mathbf{X} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ 的逆矩阵 $\mathbf{B} = \mathbf{X}^{-1} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$, 且有 $\mathbf{AB} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$, 即

$$\left\{ \begin{array}{l} \mathbf{A}_{11} \mathbf{B}_{11} + \mathbf{A}_{12} \mathbf{B}_{21} = \mathbf{I} \\ \mathbf{A}_{11} \mathbf{B}_{12} + \mathbf{A}_{12} \mathbf{B}_{22} = \mathbf{0} \\ \mathbf{A}_{21} \mathbf{B}_{11} + \mathbf{A}_{22} \mathbf{B}_{21} = \mathbf{0} \\ \mathbf{A}_{21} \mathbf{B}_{12} + \mathbf{A}_{22} \mathbf{B}_{22} = \mathbf{I} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \mathbf{B}_{21} = -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{B}_{11} \\ \mathbf{B}_{22} = \mathbf{A}_{22}^{-1} (\mathbf{I} - \mathbf{A}_{21} \mathbf{B}_{12}) = \mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{B}_{12} \end{array} \right.$$

将 \mathbf{B}_{21} 代入第 1 个式子中, 得:

$$\begin{aligned} \mathbf{A}_{11} \mathbf{B}_{11} + \mathbf{A}_{12} (-\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{B}_{11}) &= \mathbf{I} \\ (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}) \mathbf{B}_{11} &= \mathbf{I} \end{aligned}$$

记 $\boxed{\boldsymbol{F}_1 := \boldsymbol{B}_{11} = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}}$, 则 $\boldsymbol{B}_{21} = \boxed{-\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{F}_1}$. 然后, 将 \boldsymbol{B}_{22} 代入第 2 个式子中, 得:

$$\begin{aligned}\boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}(\boldsymbol{A}_{22}^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{12}) &= \mathbf{0} \\ (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\boldsymbol{B}_{12} &= -\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{B}_{12} &= -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ &= \boxed{-\boldsymbol{F}_1\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}} = \boldsymbol{B}_{21}^{\top}\end{aligned}$$

最后, 将 \boldsymbol{B}_{12} 代入 \boldsymbol{B}_{22} 中, 得:

$$\begin{aligned}\boldsymbol{B}_{22} &= \boldsymbol{A}_{22}^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{12} \\ &= \boldsymbol{A}_{22}^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(-\boldsymbol{F}_1\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}) \\ &= \boxed{\boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{F}_1\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}}\end{aligned}$$

得证.

定理 2.5

令 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^\top$ 为长度为 n 的向量分为两个子向量的长度 $(q, n - q)$. 假设 \mathbf{X} 服从多元高斯分布, 即

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right).$$

Then the conditional distribution of $\mathbf{X}_1|\mathbf{X}_2$ is given by

$$\mathbf{X}_1|\mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

很显然, $\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, $\mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, 并且 $\text{Cov}[\mathbf{X}_1, \mathbf{X}_2] = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^\top$. 因此

- $\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2] = \text{原均值} + \text{协方差占比} \times (\text{观测值} - \text{对应均值})$.
- $\text{Var}[\mathbf{X}_1|\mathbf{X}_2] = \text{原方差} - \text{可被 } \mathbf{X}_2 \text{ 解释的部分}$.

Proof. 根据多元正态密度函数公式 (joint p.d.f.):

$$f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \right).$$

记 $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{D} \end{pmatrix}^{-1}$, 其中 $\begin{cases} \mathbf{A} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}, \\ \mathbf{B} = -\mathbf{A}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}, \\ \mathbf{D} = \boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{A}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}. \end{cases}$ 化简指数内的二次型得:

$$(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \mathbf{A}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + 2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \mathbf{B}(\mathbf{x}_2 - \boldsymbol{\mu}_2) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \mathbf{D}(\mathbf{x}_2 - \boldsymbol{\mu}_2).$$

在条件密度中, \mathbf{x}_2 是已知常量, 所以我们只看这是一个关于 \mathbf{x}_1 的二次函数. 然后再配方得,

$$\begin{aligned} & (\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \mathbf{A}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + 2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \mathbf{B}(\mathbf{x}_2 - \boldsymbol{\mu}_2) + \text{const.} \\ &= \left((\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{A}^{-1}\mathbf{B}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right)^\top \mathbf{A} \left((\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{A}^{-1}\mathbf{B}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right) + \text{const.} \end{aligned}$$

因此可知

$$\begin{aligned} \mathbb{E}[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] &= \boldsymbol{\mu}_1 - \mathbf{A}^{-1}\mathbf{B}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\mu}_1 - \mathbf{A}^{-1}(-\mathbf{A}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1})(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \\ \text{Var}[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] &= \mathbf{A}^{-1} \\ &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$

得证.

.....
现在, 我们有 $\mathbf{Z} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, 则随机向量 $\mathbf{Z}^* = (Z(\mathbf{s}_0), \mathbf{Z})$ 具有以下分布

$$\mathbf{Z}^* = \begin{pmatrix} Z(\mathbf{s}_0) \\ \mathbf{Z} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{x}(\mathbf{s}_0)^\top \boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} C(0, \boldsymbol{\theta}) & \mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta})^\top \\ \mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta}) & \boldsymbol{\Sigma}(\boldsymbol{\theta}) \end{pmatrix} \right),$$

其中

- $\mathbf{x}(\mathbf{s}_0)$ 是未观测 (预测) 位置 \mathbf{s}_0 的协变量向量.
- $C(0, \boldsymbol{\theta}) = \text{Var}[\mathbf{Z}(\mathbf{s}_0)]$ 是预测位置的方差.

- $\mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta}) = \left(C(\|\mathbf{s}_1 - \mathbf{s}_0\|, \boldsymbol{\theta}), \dots, C(\|\mathbf{s}_n - \mathbf{s}_0\|, \boldsymbol{\theta}) \right)$ 是预测位置与 n 个数据位置之间的协方差向量.

然后利用这些定理的结果, 在 MSPE 意义上给出了最优预测器, 即

$$\mathbb{E}[Z(\mathbf{s}_0)|\mathbf{Z}] = \mathbf{x}(\mathbf{s}_0)^\top \boldsymbol{\beta} + \mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}),$$

这被称为 **Universal Kriging Predictor (通用克里金预测器)**. 预测器的方差是

$$\text{Var}[Z(\mathbf{s}_0)|\mathbf{Z}] = C(0, \boldsymbol{\theta}) - \mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{C}(\mathbf{s}_0, \boldsymbol{\theta}),$$

它量化了预测中的不确定性.

在实践中 $(\boldsymbol{\beta}, \boldsymbol{\theta})$ 是未知的, 因此实际上在高斯过程假设下, 我们有一个在 MSPE 意义下的最优预测器是

$$Z(\mathbf{s}_0)|\mathbf{Z} \sim \mathcal{N}\left(\mathbb{E}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right], \text{Var}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right]\right),$$

其中

- $\mathbb{E}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right] = \mathbf{x}(\mathbf{s}_0)^\top \hat{\boldsymbol{\beta}} + \mathbf{C}(\mathbf{s}_0, \hat{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}).$
- $\text{Var}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right] = C(0, \hat{\boldsymbol{\theta}}) - \mathbf{C}(\mathbf{s}_0, \hat{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{C}(\mathbf{s}_0, \hat{\boldsymbol{\theta}}).$

这直接导致了一个 95% 预测区间,

$$\mathbb{E}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right] \pm 1.96\sqrt{\text{Var}\left[\widehat{Z(\mathbf{s}_0)}|\mathbf{Z}\right]}$$

如果均值 $\mu(\mathbf{s}) = \mu$ 是常数, 那么它被称为 **Ordinary Kriging 普通克里金**, 但如果均值不是常数, 则称为 **Universal Kriging (通用克里金)**.

3 Areal unit processes (区域单元过程)

定义 3.1 Areal unit process (区域单元/面元过程)

Areal unit process is also known as **lattice process** (格点过程), 空间域 D 被划分为 n 个不重叠 (non-overlapping, 不相交, disjoint) 的 **面元 (areal unit)** 或 **分区 (partition)**, 记为

$$D = \{B_1, B_2, \dots, B_n\}.$$

Then the **areal process** is a stochastic process $\mathbf{Z} = (Z(B_1), Z(B_2), \dots, Z(B_n))$.

这一组 areal units 可以形成一个规则的 grid (lattice), 也可以是不规则的形状单元, 但必须满足

- $\bigcup_{i=1}^n B_i = D$ (即区域被划分成的子域没有遗漏), 和
- $B_i \cap B_j = \emptyset$ for each $i \neq j$ (即不重叠).

3.1 Exploratory Analysis for Areal Unit Data (EDA)

3.1.1 Mapping (制图)

探索性数据分析的第一阶段是绘制数据的空间地图, 以及区域单元的边界 (B_1, \dots, B_n) . 在大多数情况下, 这需要获得两个不同的量:

- 每个 areal unit 需要映射 (mapped) 的数据.

- 一个包含每个 areal unit 多边形轮廓 (outlines of the polygons) 的 **shapefile**. Shapefile 是一组具有相同名称但扩展名不同的不同文件类型, 包括 *.shp*, *.dbf*, *.shx* 等.

3.1.2 Defining spatial closeness (空间邻近性)

在区域数据中定义和建模空间依赖性的一个重要概念是 **neighbourhood (邻域)** 或 **adjacency matrix (邻接矩阵)**, W , 它定义了不同的区域彼此之间的空间邻近程度. 该矩阵的值可以是

- **Binary**: 哪些区域在空间上接近其他区域.
- **Continuous**: 测量区域对之间空间接近程度的范围.

令 $W = (w_{ij})$ 表示一个非负的 $n \times n$ 矩阵, 其中 (i, j) 项 w_{ij} 表示区域 B_i 和 B_j 之间的空间接近程度.

- 如果 $w_{ij} = 0$, 则区域 (B_i, B_j) 在空间上不接近. 显然, 对角线上元素 $w_{ii} = 0$.
- 如果 $w_{ij} > 0$ 则区域 (B_i, B_j) 在空间上接近, 并且 w_{ij} 越大, 两个区域就越接近.

定义 3.2 Binary W matrix

为 W 提供二元规范, 即 w_{ij} 的值为 0 或 1. 如果区域 (B_i, B_j) 被定义为 **neighbour**, 则 $w_{ij} = 1$.

指定 neighbour 的方法有很多, 最常见的是:

- 如果区域 (B_i, B_j) 公用一个边界.
- 如果区域 B_j 是按其质心 (centroid, centre point) 之间的距离而言, 区域 B_i 最近的 k 个区域之一. 尽管 k 的选择并不明显.

定义 3.3 Non-binary \mathbf{W} matrix

通常基于距离, 两个区域相距越远, 权重越小.

以下是基于距离的 \mathbf{W} 的常见规范, 并依赖于区域 (B_i, B_j) 的质心 $(\mathbf{s}_i, \mathbf{s}_j)$.

- $w_{ij} = \frac{1}{\|\mathbf{s}_i - \mathbf{s}_j\|}.$
- $w_{ij} = \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|).$

这里 $\|\cdot\|$ 表示欧几里得距离. 对于这两种规范, \mathbf{W} 不是一个 sparse (稀疏, 即有很多零) 的矩阵, 这使得模型拟合速度大大降低。

3.1.3 Estimating the extent of spatial autocorrelation (估计空间自相关程度)

定义 3.4 Moran's I

Moran's I statistic computed for data $\mathbf{z} = (z_1, \dots, z_n)$ with a neighbourhood matrix \mathbf{W} is defined as

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (z_i - \bar{z})^2}.$$

它测量了区域数据中 **线性关联强度 (strength of the linear association)**, 并根据其空间位置进行了适当的加权. 与 acf 和 Pearson 一致, $-1 < I < 1$, and

$$I = \begin{cases} -1, & \text{perfect dispersion (分散),} \\ 0, & \text{a random arrangement,} \\ 1, & \text{perfect autocorrelation.} \end{cases}$$

定义 3.5 Permutation test (置换检验)

置换检验通常用于评估空间自相关的显著性. The hypotheses for this test are:

$$H_0 : \text{no spatial autocorrelation} \quad \text{vs.} \quad H_1 : \text{positive spatial autocorrelation.}$$

Moran's I 统计量用于 K 组不同数据随机排列. 该检验的 p 值为

$$p = \frac{1}{K+1} \sum_{k=1}^K \mathbb{1}_{\{I_k > I_{\text{obs}}\}}.$$

例 3.1 Bookwork 3.2.3

考虑一个简单的 1 维空间数据, 有 4 个区域按顺序为 $[A|B|C|D]$, 并有一个相应的邻域矩阵

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

so that the only neighbour pairs are (A, B) , (B, C) and (C, D) . Then suppose that $z(A) = 3$, $z(B) = 4$, $z(C) = 5$ and $z(D) = 6$. Compute the **Moran's I** statistic.

Solution.

- First, $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = 6$, and $\bar{z} = 4.5$.

- Second, $\sum_{i=1}^n (z_i - \bar{z})^2 = (3 - 4.5)^2 + (4 - 4.5)^2 + (5 - 4.5)^2 + (6 - 4.5)^2 = 5.$

- Third,

$$\begin{aligned} \sum_{j=1}^n w_{ij}(z_i - \bar{z})(z_j - \bar{z}) &= 2 \times \left((z(A) - \bar{z})(z(B) - \bar{z}) + (z(B) - \bar{z})(z(C) - \bar{z}) + (z(C) - \bar{z})(z(D) - \bar{z}) \right) \\ &= 2 \times ((3 - 4.5)(4 - 4.5) + (4 - 4.5)(5 - 4.5) + (5 - 4.5)(6 - 4.5)) \\ &= 2.5. \end{aligned}$$

Hence Moran's I is given by

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (z_i - \bar{z})^2} = \frac{4 \times 2.5}{6 \times 5} = \frac{1}{3} = 0.3333.$$

定义 3.6 Geary's contiguity ratio (邻接比)

C is defined by

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{2 \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (z_i - \bar{z})^2} \Rightarrow C_i = \sum_{j=1}^n w_{ij} (z_i - z_j)^2.$$

This is a positively valued statistic where:

$$C \begin{cases} \rightarrow 0^+, & \text{strong positive autocorrelation.}, \\ \approx 1, & \text{perfect independence,} \\ > 1, & \text{negative autocorrelation.} \end{cases}$$

定义 3.7 Local Indicator of Spatial Association (LISA, 局部空间关联指标)

One approach is a local version of Moran's I , which for region i is given by

$$I_i = \frac{n(z_i - \bar{z}) \sum_{j=1}^n w_{ij} (z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2}.$$

If area i is an outlier that has a much different value to all of its neighbours, then its LISA will be negative.

例 3.2 via 2021.Q2(a)

Consider a simple 1-dimensional spatial data set with 4 regions ordered as $[\mathcal{A}_1|\mathcal{A}_2|\mathcal{A}_3|\mathcal{A}_4]$, with a corresponding neighbourhood matrix

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Then suppose that $Z_1 = 100$, $Z_2 = 101$, $Z_3 = 100$ and $Z_4 = 90$. Compute the **local indicators of spatial association** based on Moran's I statistic for areal units \mathcal{A}_2 and \mathcal{A}_4 , and interpret what these statistics tell you about the data.

Solution. First $n = 4$ and $\bar{Z} = 97.75$ and we have $\sum_{i=1}^4 (Z_i - \bar{Z})^2 = 80.75$. Then for area \mathcal{A}_2 we have

$$I_2 = \frac{4(101 - 97.75) [1 \times (100 - 97.75) + 0 \times (101 - 97.75) + 1 \times (100 - 97.75) + 0 \times (90 - 97.75)]}{80.75} = 0.7244.$$

Then for area \mathcal{A}_4 we have

$$I_4 = \frac{4(90 - 97.75) [0 \times (100 - 97.75) + 0 \times (101 - 97.75) + 1 \times (100 - 97.75) + 0 \times (90 - 97.75)]}{80.75} = -0.864.$$

As I_2 is positive it tells us that the data value Z_2 is positively correlated (similar) to its neighbours, but as I_4 is negative it tells us that the data value Z_4 is negatively correlated (very different) to its neighbours.

由于 I_2 为正, 这告诉我们数据值 Z_2 与其邻居呈正相关 (相似), 但 I_4 为负, 这告诉我们数据值 Z_4 与其邻居呈负相关 (非常不同).

3.2 Gaussian Markov random field theory (高斯马尔可夫随机场理论)

高斯马尔可夫随机场 (Gaussian Markov random fields, GMRFs) 作为一种机制, 来构建区域过程 $\mathbf{Z} = (Z_1, \dots, Z_n)$ 中随机变量之间的依赖关系. 这类模型包括许多不同类型的模型, 而我们在这里考虑的特殊情况是条件自回归 (Conditional Autoregressive, CAR) 模型. CAR 模型是目前在区域单元数据中模拟空间自相关的最常用方法, 并且它们是时间序列 AR 模型的一个直接扩展.

例 3.3

考虑 AR(1) 模型

$$X_t = \alpha X_{t-1} + Z_t, \quad Z_t \sim \mathcal{N}(0, \sigma^2).$$

- $\alpha = 0$: 数据是独立的.
- $\alpha = 1$: 数据具有强的时间自相关性, 并且过程是非平稳的. 这是一个一阶随机游走过程.
- $\alpha \in (0, 1)$: 过程是自相关的, 均值和方差一定, 即宽平稳.

因此, AR(1) 过程是一系列条件分布

$$f(X_t | X_{t-1}) = \mathcal{N}(\alpha X_{t-1}, \sigma^2).$$

为了确定 $\mathbf{X} = (X_1, \dots, X_n)$ 的有效的联合概率分布, 根须条件概率公式 $f(A, B) = f(A|B)f(B)$, 有:

$$\begin{aligned}
 f(X_1, \dots, X_n) &= f(X_1) \times f(X_2|X_1) \times f(X_3|X_2, X_1) \times \dots \times f(X_n|X_{n-1}, \dots, X_1) \\
 &= f(X_1) \prod_{t=2}^n f(X_t|X_{t-1}, \dots, X_1) \\
 &= f(X_1) \prod_{t=2}^n f(X_t|X_{t-1}) \\
 &= \mathcal{N}(X_1|0, \sigma^2) \prod_{t=2}^n \mathcal{N}(X_t|\alpha X_{t-1}, \sigma^2).
 \end{aligned}$$

我们称该过程为 **一阶时间马尔可夫 (Markov-in-time of order one)**. AR(1) 过程只是高斯马尔可夫随机场 (GMRF) 的一个特例.

在空间数据案例中, 我们根据两个数据点是否被定义为彼此的邻居 (如 \mathbf{W} 所定义) 来衡量它们的空间邻近度. 因此, 令

$$N_{Z_i} = \{j | w_{ij} > 0\},$$

为区域 i 的邻域集合. 那么, 对于建模空间数据 \mathbf{Z} 的类似结果 (analogous result) 将是

$$f(Z_1, \dots, Z_n) = \prod_{t=1}^n f(Z_i | \{Z_j : j \in N_{Z_i}\}),$$

每个 Z_i 在仅定义为其邻居的其他随机变量的条件下, 是单变量**全条件分布 (full conditional distributions)** 的乘积. 当上式中所有条件分布和联合分布都是高斯分布时, 该过程称为 **高斯马尔可夫随机场 (Gaussian Markov random field, GMRF)**.

3.2.1 Hammersley-Clifford theorem (GMRF 的充分条件)

这个定理旨在告诉我们, 联合分布 $f(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ 必须采取的什么样的形式才能成为高斯马尔可夫随机场.

定义 3.8 Clique (团)

在整体域 D 中的 **团 (clique)** 用 \mathcal{C} 表示, 它是 n 个区域的一个子集, 使得团 \mathcal{C} 中所有的区域对都是彼此的邻居. 对于不满足此条件的区域 j , 团可能是一个单独的区域.

比如, 如果区域 (B_1, B_2, B_3) 是团, 那么需使 $w_{12} = w_{21} > 0$, $w_{13} = w_{31} > 0$, 且 $w_{23} = w_{32} > 0$.

定义 3.9 Potential function (势函数)

一个 size 为 k 的团 \mathcal{C} 的 **势函数 (potential function)** 或 **势 (potential)** $f_{\mathcal{C}}(\mathbf{Z})$ 是一个关于团中 k 个元素的函数, 其参数是可交换 (对称) 的.

比如,

- $f_{\mathcal{C}}(Z_i, Z_j) = (Z_i - Z_j)^2$: 在 GMRF 模型中最有可能用, 因为其出现在正态分布的指数中.
- $f_{\mathcal{C}}(Z_i, Z_j) = Z_i Z_j$: 出现在二元空间数据的 Ising 模型 (不知道是啥) 中.
- $f_{\mathcal{C}}(Z_i, Z_j) = Z_i Z_j + (1 - Z_i)(1 - Z_j)$.

定理 3.1 Hammersley-Clifford theorem

一个对于 $\mathbf{Z} = (Z_1, \dots, Z_n)$ 的 GMRF 模型的联合分布可以表示为:

$$f(Z_1, \dots, Z_n) = \frac{1}{K} \exp \left(\gamma \sum_{C \in cl(D)} f_C(\mathbf{Z}) \right),$$

其中 $cl(D)$ 表示在域 D 上的所有团的集合, K 是一个归一化常数.

这个定理告诉我们, 在将提出的全条件分布相乘后, 你必须能够将乘积写成团集上势能的指数和的形式. 这种联合分布被称为**吉布斯分布 (Gibbs distribution)**. 这些势函数定义了**局部**依赖结构, 进而定义了 GMRF 的联合**全局**分布.

这个定理告诉我们如何从一个全条件分布集合推导出有效的联合分布, 但反过来呢? 如果我写下联合多元高斯分布, 我如何获得它的全条件分布, 以知道它们是否仅在邻域集上具有马尔可夫性质?

如定理 2.5 所述, 若将随机向量分割为 $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)^\top$, 它们的联合分布为

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

then the conditional distribution of $\mathbf{Z}_1 | \mathbf{Z}_2$ is given by

$$\mathbf{Z}_1 | \mathbf{Z}_2 \sim \mathcal{N} \left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{Z}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right).$$

This can be applied with $\mathbf{Z}_1 = Z_j$ and $\mathbf{Z}_2 = \mathbf{Z}_{-j}$, where $\mathbf{Z}_{-j} = (Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_n)$ denotes all elements in \mathbf{Z} except Z_j .

然而, 在 GMRF 模型中, 通常使用精度矩阵 (precision matrix) $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ 而不是 $\boldsymbol{\Sigma}$, 引出以下定理.

定理 3.2 Precision matrix $\mathbf{Q} = \Sigma^{-1}$

Consider partitioning the random vector $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, where its joint distribution is given by

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \Sigma = \mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}^{-1} \right)$$

Then the conditional distribution of $\mathbf{Z}_1 | \mathbf{Z}_2$ is given by

$$\mathbf{Z}_1 | \mathbf{Z}_2 \sim \mathcal{N} (\boldsymbol{\mu}_1 - \mathbf{Q}_{11}^{-1} \mathbf{Q}_{12} (\mathbf{Z}_2 - \boldsymbol{\mu}_2), \mathbf{Q}_{11}^{-1})$$

Proof. Given that $\mathbf{Q} = \Sigma^{-1}$, it must be that

$$\begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix} \times \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix}.$$

Taking the top row of this block matrix equation gives:

$$\mathbf{Q}_{11} \Sigma_{11} + \mathbf{Q}_{12} \Sigma_{21} = \mathbf{I} \quad \Rightarrow \quad \Sigma_{11} = \mathbf{Q}_{11}^{-1} (\mathbf{I} - \mathbf{Q}_{12} \Sigma_{21})$$

$$\mathbf{Q}_{11} \Sigma_{12} + \mathbf{Q}_{12} \Sigma_{22} = \mathbf{0} \quad \Rightarrow \quad \Sigma_{12} = -\mathbf{Q}_{11}^{-1} \mathbf{Q}_{12} \Sigma_{22}$$

Thus subbing these into the mean and variance formula stated above that uses the variance matrix Σ gives:

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}_1|\mathbf{Z}_2] &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\mu}_1 + (-\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\Sigma_{22})\Sigma_{22}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\mu}_1 - \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}(\mathbf{Z}_2 - \boldsymbol{\mu}_2) \\
\text{Var}[\mathbf{Z}_1|\mathbf{Z}_2] &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \mathbf{Q}_{11}^{-1}(\mathbf{I} - \mathbf{Q}_{12}\Sigma_{21}) - (-\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\Sigma_{22})\Sigma_{22}^{-1}\Sigma_{21} \\
&= \mathbf{Q}_{11}^{-1} - \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\Sigma_{21} + \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\Sigma_{21} \\
&= \mathbf{Q}_{11}^{-1}
\end{aligned}$$

这里一定要注意一件事: 将向量 \mathbf{Z} 拆成两部分 $(\mathbf{Z}_1, \mathbf{Z}_2)$, 其中 \mathbf{Z}_1 是我们感兴趣的子向量, \mathbf{Z}_2 是已知变量, 那么 \mathbf{Q}_{11} 实际是从整个 \mathbf{Q} 中提取的矩阵. 换句话说, 如果 $\mathbf{Z} = (Z_i, Z_j, Z_k, \dots)$, 那么 \mathbf{Q}_{11} 应该是由 \mathbf{Q} 的第 i, j, k, \dots 行和第 i, j, k, \dots 列提取出来的子阵.

3.2.2 Conditional autoregressive (CAR) models (条件自回归模型)

定义 3.10 Conditional autoregressive (CAR) models (条件自回归模型)

考虑数据 $\mathbf{Z} = (Z_1, \dots, Z_n)$, 具有完全条件分布的一类条件自回归模型 (conditional autoregressive models) 满足

$$Z_i|\mathbf{Z}_{-i} \sim \mathcal{N}\left(\sum_{j=1}^n b_{ij}Z_j, \tau_i^2\right), \quad \text{where } b_{ii} = 0.$$

然后这些完全条件分布与联合分布兼容, 该联合分布为

$$f(Z_1, \dots, Z_n) \propto \exp \left(-\frac{1}{2} \mathbf{Z}^\top \mathbf{D}^{-1} (\mathbf{I} - \mathbf{B}) \mathbf{Z} \right),$$

其中, $\mathbf{B} = (b_{ij})$ 是一个包含回归系数的 $n \times n$ 矩阵, 而 $\mathbf{D} = \text{diag}(\tau_1^2, \dots, \tau_n^2)$ 是一个包含方差的对角矩阵. 这个联合概率密度函数看起来像

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q} = \mathbf{\Sigma}^{-1} = \mathbf{D}^{-1}(\mathbf{I} - \mathbf{B})),$$

即 $\mathbf{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{D}$. 在我们能够说 $\mathbf{\Sigma}$ 是一个有效的协方差矩阵之前, 我们需要确保 $\mathbf{Q} = \mathbf{D}^{-1}(\mathbf{I} - \mathbf{B})$ 是对称的, 即只需保证 $\mathbf{D}^{-1} \mathbf{B}$ 是对称的. 如果 \mathbf{Q} 不是对称的, 那么它不是一个有效的精度矩阵, 因此 $\mathbf{\Sigma} = \mathbf{Q}^{-1}$ 不是一个有效的协方差矩阵. 确保这必须成立的简单条件是

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}, \quad \forall i, j.$$

这意味着 \mathbf{B} 不需要是对称矩阵. 满足此约束的最简单规定是设置

$$b_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}}, \quad \tau_i^2 = \frac{\tau^2}{\sum_{j=1}^n w_{ij}},$$

在邻域矩阵 \mathbf{W} 对称的条件下, 这满足此约束.

定义 3.11 Intrinsic conditional autoregressive (ICAR) model (内在条件自回归模型)

最简单的条件自回归模型是 **内在条件自回归 (ICAR) 模型**, 其给定了一个非标准的多元高斯分布:

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]^-),$$

其中, \mathbf{A}^- 表示矩阵 \mathbf{A} 的广义逆 (generalised inverse).

精度矩阵 $\mathbf{Q} = \frac{1}{\tau^2} [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]$ 是秩亏 (rank deficient) 的 (秩为 $n - 1$ 而不是 n), 因此是奇异 (singular, 不满秩的方阵, 导致行列式为 0) 的. 这很容易发现, 因为 \mathbf{Q} 每行之和均为零, 因为

$$\begin{aligned}
\mathbf{Q} &= \frac{1}{\tau^2} [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}] = \frac{1}{\tau^2} \left\{ \text{diag} \left[\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right] - \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \right\} \\
&= \frac{1}{\tau^2} \left\{ \text{diag} \begin{pmatrix} \sum_{j=1}^n w_{1j} \\ \sum_{j=1}^n w_{2j} \\ \vdots \\ \sum_{j=1}^n w_{nj} \end{pmatrix} - \begin{pmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{pmatrix} \right\} \quad \text{due to } w_{ii} = 0 \\
&= \frac{1}{\tau^2} \begin{pmatrix} \sum_{j=1}^n w_{1j} & -w_{12} & \cdots & -w_{1n} \\ -w_{21} & \sum_{j=1}^n w_{2j} & \cdots & -w_{2n} \\ \vdots & \vdots & & \vdots \\ -w_{n1} & -w_{n2} & \cdots & \sum_{j=1}^n w_{nj} \end{pmatrix}.
\end{aligned}$$

尽管 ICAR 模型的联合分布是非标准的, 但完全条件分布 $Z_i | \mathbf{Z}_{-i}$ 是标准的, 并且可以使用上述多元高

斯结果推导出来, 即

$$\begin{aligned}
\mathbb{E}[Z_i | \mathbf{Z}_{-i}] &= \boldsymbol{\mu}_1 - Q_{ii}^{-1} \mathbf{Q}_{i,-i} (\mathbf{Z}_{-i} - \boldsymbol{\mu}_{-i}) \\
&= 0 - \left(\frac{1}{\tau^2} \sum_{j=1}^n w_{ij} \right)^{-1} \left(\frac{1}{\tau^2} \sum_{j \neq i} (-w_{ij})(Z_j - 0) \right) \\
&= \frac{\sum_{j \neq i} w_{ij} Z_j}{\sum_{j=1}^n w_{ij}} = \frac{\sum_{j=1}^n w_{ij} Z_j}{\sum_{j=1}^n w_{ij}} \quad \text{due to } d_{ii} = 0 \\
\text{Var}[Z_i | \mathbf{Z}_{-i}] &= \mathbf{Q}_{ii}^{-1} \\
&= \left(\frac{1}{\tau^2} \sum_{j=1}^n w_{ij} \right)^{-1} \\
&= \frac{\tau^2}{\sum_{j=1}^n w_{ij}}
\end{aligned}$$

因此我们有:

$$Z_i | \mathbf{Z}_{-i} \sim \mathcal{N} \left(\frac{\sum_{j=1}^n w_{ij} Z_j}{\sum_{j=1}^n w_{ij}}, \frac{\tau^2}{\sum_{j=1}^n w_{ij}} \right).$$

利用上述多元高斯结果, 我们可以计算 $(Z_i, Z_j) | \mathbf{Z}_{-ij}$ 之间的偏 (条件) 自相关 (partial (conditional) autocorrelation). 利用上述定理从联合分布中获得条件分布, 得到以下结果, 其中 \mathbf{Q}_{ij} 是 \mathbf{Q} 的对应于第 i

行和第 j 列的 2×2 子矩阵.

$$\begin{aligned}\text{Var} [(Z_i, Z_j)|\mathbf{Z}_{-ij}] &= \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]_{ij}^{-1} = \mathbf{Q}_{ij}^{-1} \\ &= \tau^2 \begin{pmatrix} \sum_{k=1}^n w_{ik} & -w_{ij} \\ -w_{ji} & \sum_{k=1}^n w_{jk} \end{pmatrix}^{-1} = \frac{\tau^2}{d_{ij}} \begin{pmatrix} \sum_{k=1}^n w_{jk} & w_{ij} \\ w_{ji} & \sum_{k=1}^n w_{ik} \end{pmatrix},\end{aligned}$$

其中, d_{ij} 为行列式, 即 $d_{ij} = \left(\sum_{k=1}^n w_{ik}\right) \left(\sum_{k=1}^n w_{jk}\right) - w_{ij}^2$, 二阶矩阵的逆“主对调, 副变号”.

因此, 偏 (条件) 自相关为

$$\begin{aligned}\text{Corr} [(Z_i, Z_j)|\mathbf{Z}_{-ij}] &= \frac{\text{Cov} [(Z_i, Z_j)|\mathbf{Z}_{-ij}]}{\sqrt{\text{Var} [Z_i|\mathbf{Z}_{-ij}] \text{Var} [Z_j|\mathbf{Z}_{-ij}]}} \\ &= \frac{w_{ij}}{\sqrt{\left(\sum_{k=1}^n w_{ik}\right) \left(\sum_{k=1}^n w_{jk}\right)}}\end{aligned}$$

这意味着,

- 对于相邻区域, 即 $w_{ij} \neq 0$ 时, 数据 (Z_i, Z_j) 是**部分 (条件) 相关的 (partially (conditionally) correlated)**;
- 对于非相邻区域, 即 $w_{ij} = 0$ 时, 数据 (Z_i, Z_j) 是**条件独立的 (conditionally independent)**.

因此, 邻域矩阵 \mathbf{W} 中的 0 意味着精度矩阵 \mathbf{Q} 中的 0, 这意味着给定其余数据值, 数据是条件独立的.

例 3.4 via 2021.Q2(b)-ii

The researcher decides to model the data with the following CAR model (后面定义3.15):

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho)^{-1}),$$

where the ij th element of precision matrix $\mathbf{Q}(\mathbf{W}, \rho)$ is given by

$$\mathbf{Q}(\mathbf{W}, \rho)_{ii} = \rho \sum_{j=1}^n w_{ij} + 1 - \rho \quad \text{and} \quad \mathbf{Q}(\mathbf{W}, \rho)_{ij} = -\rho w_{ij}$$

Using the multivariate Gaussian theory that was presented in the lectures, we have that

$$\begin{aligned} \text{Var}[Z_i, Z_j | \mathbf{Z}_{-ij}] &= \tau^2 \mathbf{Q}_{ij}^{-1} = \tau^2 \begin{pmatrix} \mathbf{Q}_{ii} & \mathbf{Q}_{ij} \\ \mathbf{Q}_{ji} & \mathbf{Q}_{jj} \end{pmatrix}^{-1} \\ &= \frac{\tau^2}{\mathbf{Q}_{ii}\mathbf{Q}_{jj} - \mathbf{Q}_{ij}^2} \begin{pmatrix} \mathbf{Q}_{jj} & -\mathbf{Q}_{ij} \\ -\mathbf{Q}_{ji} & \mathbf{Q}_{ii} \end{pmatrix} \\ \text{Corr}[Z_i, Z_j | \mathbf{Z}_{-ij}] &= \frac{\text{Cov}[Z_i, Z_j | \mathbf{Z}_{-ij}]}{\sqrt{\text{Var}[Z_i | \mathbf{Z}_{-ij}] \text{Var}[Z_j | \mathbf{Z}_{-ij}]}} = \frac{\mathbf{Q}_{ij}}{\sqrt{\mathbf{Q}_{ii}\mathbf{Q}_{jj}}}. \end{aligned}$$

Q. 此 CAR 模型在 $w_{ij} = 0$ (即区域 $\mathcal{A}_i, \mathcal{A}_j$ 不相邻) 的情况下, 对 (Z_i, Z_j) 之间的依赖性有何假设?

A. If $w_{ij} = 0$ then this model assumes that (Z_i, Z_j) are **conditionally independent** (the conditional correlation in the previous question is clearly zero) given the remaining data points, that is $\text{Corr}[Z_i, Z_j | \mathbf{Z}_{-ij}] = 0$.

3.3 Modelling areal unit data (建模)

3.3.1 Modelling framework (建模框架)

建模 areal unit data 的最简单方法是基于联合数据似然模型进行估计

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]).$$

这可以通过调整模型来扩展, 包括协变量

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]),$$

但是这并不常用, 因为精度矩阵是奇异的, 则方差矩阵亦不存在; 此外, 该模型只适用于连续数据, 不适用离散数据.

因此, 通常将 CAR 模型指定为一组随机效应的先验分布, 而不是直接指定为数据似然模型. 这克服了上述两个限制, 因为先验分布可以是非正态的, 而数据模型则可以自由地采用任何适当的分布.

定义 3.12 贝叶斯方法 – 分层模型 (hierarchical model)

$$\left. \begin{aligned} Z_i &\sim f(Z_i|\mu_i), \quad i = 1, \dots, n \\ g(\mu_i) &= \mathbf{x}_i\boldsymbol{\beta} + \phi_i \end{aligned} \right\} \text{generalised linear mixed model}$$

$$\phi_i | \boldsymbol{\phi}_{-1} \sim \mathcal{N}\left(\frac{\sum_{j=1}^n w_{ij} Z_j}{\sum_{j=1}^n w_{ij}}, \frac{\tau^2}{\sum_{j=1}^n w_{ij}}\right)$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$$

$$\tau^2 \sim \text{Inverse-gamma}(a, b)$$

假设数据在空间自相关的随机效应 (random effects, ϕ) 下是条件独立的, 因此数据似然模型为

$$f(\mathbf{Z}|\boldsymbol{\beta}, \phi, \tau^2) = \prod_{i=1}^n f(Z_i|\boldsymbol{\beta}, \phi),$$

联合先验分布为

$$\begin{aligned} f(\boldsymbol{\beta}, \phi, \tau^2) &= f(\boldsymbol{\beta})f(\phi|\tau^2)f(\tau^2) \\ &= \mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \mathbf{V}_\beta) \times \text{ICAR}(\phi|\tau^2, \mathbf{W}) \times \text{Inverse-gamma}(\tau^2|a, b), \end{aligned}$$

其中, $(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta, a, b)$ 被称为 **hyperparameters (超参数)**, 人为选择, 通常 $(\boldsymbol{\mu}_\beta = \mathbf{0}, \mathbf{V}_\beta = 100,000\mathbf{I}, a = 1, b = 0.01)$ 确保先验信息微弱.

最常用的 generalised linear mixed model 有如下三种:

1. **Binomial logistic model (二项逻辑模型)**: $Z_i \sim \text{Binomial}(N_i, \theta_i)$

$$\log\left(\frac{\theta_i}{1 - \theta_i}\right) = \mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i.$$

这里 N_i 是区域 i 的试验次数, θ_i 是每次试验的成功概率. 当然, 也可以使用正态分布 (probit, 概率单位) 而不是逻辑回归.

2. **Gaussian linear model (高斯线性模型)**: $Z_i \sim \mathcal{N}(\mu_i, \sigma^2)$

$$\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i.$$

此处, 数据误差方差 σ^2 将被赋予与 τ^2 相同弱信息的 (weakly informative) inverse-gamma 先验分布。

3. **Poisson log-linear model (泊松对数线性模型)**: $Z_i \sim \text{Poisson}(\mu_i)$

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \phi_i.$$

3.3.2 Parameter estimation (参数估计)

如前所述, 参数估计将在贝叶斯框架下进行, 目标是估计后验分布

$$\begin{aligned} f(\Theta|\mathbf{Z}) &\propto f(\mathbf{Z}|\Theta)f(\Theta) \\ &= \left[\prod_{i=1}^n f(Z_i|\beta, \phi_i) \right] \mathcal{N}(\beta|\mu_\beta, \mathbf{V}_\beta) \times \text{ICAR}(\phi|\tau^2, \mathbf{W}) \times \text{Inverse-gamma}(\tau^2|a, b) \times \text{Inverse-gamma}(\sigma^2|a, b), \end{aligned}$$

其中 $\Theta = (\beta, \phi, \tau^2, \sigma^2)$ 表示所有模型参数 (如果数据模型是高斯的, 则只需要 σ^2). 这通常使用马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 模拟来完成, 这是一种迭代算法, 从该联合后验分布中抽取样本。

3.3.3 An alternative estimation approach (参数估计的替代办法)

对于这些层次高斯马尔可夫随机场模型, 存在一种贝叶斯推理的替代方法. 它被称为**集成嵌套拉普拉斯近似 (Integrated Nested Laplace Approximations, INLA)**, 通过拉普拉斯分布来近似每个单变量参数的后验分布, 而不是从中抽取随机样本.

- 优点:
 - 它非常快, 比 MCMC 快得多.
 - 有一个 R package *INLA*, 因此实现这些模型所需的编码很少.
- 缺点: 它具有局限性 (restrictive), 因为它只能适用于特定类型的模型.

3.4 Extensions to the ICAR model (内在条件自回归模型的扩展)

- Multivariate form (多元形式):

$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}]^-).$$

- Conditional form (条件形式):

$$\phi_i | \boldsymbol{\phi}_{-i} \sim \mathcal{N}\left(\frac{\sum_{j=1}^n w_{ij} \phi_j}{\sum_{j=1}^n w_{ij}}, \frac{\tau^2}{\sum_{j=1}^n w_{ij}}\right).$$

这是 CAR 类中首次提出的模型, 但它存在以下缺陷:

- 它是时间序列一阶随机游走过程的空间类比, 因此其在空间上是非平稳的.
- 它只模拟强自相关, 有时会对随机效应施加过多的空间平滑. 这是因为它不包含平滑或自相关参数, 这允许从数据中估计自相关的强度. 条件方差参数 τ^2 并不控制过程中的平滑程度, 因为如果你将 ϕ 乘以 10, 那么方差 τ^2 将明显增加, 因为 $\text{Var}[aX] = a^2 \text{Var}[X]$, 但过程的光滑度并没有改变.
- 其联合分布的多变量高斯表示是不恰当的, 因为精度矩阵是奇异的.

因此, 已提出许多该模型的扩展, 它们克服了其中一些或所有所有问题.

定义 3.13 Convolution (卷积) 或 BYM (Besag-York-Mollié)

The **Convolution or BYM** CAR model for ϕ is given by the following linear combination:

$$\begin{aligned}\phi_i &= \phi_i^{(1)} + \phi_i^{(2)} \\ \phi_i^{(1)} | \phi_{-i}^{(1)} &\sim N \left(\frac{\sum_{j=1}^n w_{ij} \phi_j^{(1)}}{\sum_{j=1}^n w_{ij}}, \frac{\tau_1^2}{\sum_{j=1}^n w_{ij}} \right) \\ \phi_i^{(2)} &\sim \mathcal{N}(0, \tau_2^2).\end{aligned}$$

- 该模型是实践中最常用的 CAR 模型.
- BYM 模型通过添加一组具有零均值和恒定方差的 i.i.d. 随机效应来扩展 ICAR 模型. 因此, 它是由相关且独立过程 (或求和) 的卷积组成的.
- ϕ 中元素的空间平滑程度由两个方差 $\frac{\tau_1^2}{\tau_2^2}$ 的比值决定.
- 因此, 克服了 ICAR 模型过于平滑且不让数据选择空间自相关强度的问题.

Disadvantages:

- 第一个分量 $\phi^{(1)}$ 是 ICAR 先验, 仍然具有不适当的多变量高斯分布.
- 在拟合该模型时, 人们试图为每个数据点 Z_i 估计两个随机效应 $(\phi_i^{(1)}, \phi_i^{(2)})$, 并且只有它们的和 $\phi_i^{(1)} + \phi_i^{(2)}$ 可以从数据中可靠地估计出来. 对于两个方差参数 (τ_1^2, τ_2^2) 也存在类似的问题.

定义 3.14 Proper CAR

The **proper CAR** model 引入了 ρ 来控制空间自相关的程度, 即

$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \tau^2 [\text{diag}(\mathbf{W}\mathbf{1}) - \rho\mathbf{W}]^{-1}).$$

如果 $\rho \in [0, 1)$, 则精度矩阵 $\mathbf{Q} = \frac{1}{\tau^2} [\text{diag}(\mathbf{W}\mathbf{1}) - \rho\mathbf{W}]$ 是可逆的. 这意味着对于 $\rho \in [0, 1)$, 此模型指定了一个适当的多变量高斯分布. 应用关于条件分布的多变量高斯结果, 得到完全条件分布为

$$\phi_i | \boldsymbol{\phi}_{-i} \sim \mathcal{N}\left(\rho \frac{\sum_{j=1}^n w_{ij} \phi_j}{\sum_{j=1}^n w_{ij}}, \frac{\tau^2}{\sum_{j=1}^n w_{ij}}\right).$$

从这个条件规范中很容易看出, ρ 控制了数据中空间自相关的程度,

- $\rho = 0$: independence in space.
- $\rho = 1$: strong spatial dependence.

此模型因此是 AR(1) 过程的空间类似物, 因为它可以写成

$$\phi_i | \boldsymbol{\phi}_{-i} = \rho \bar{\phi}_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \tau_i^2),$$

其中 $\bar{\phi}_i = \frac{\sum_{j=1}^n w_{ij} \phi_j}{\sum_{j=1}^n w_{ij}}$, 邻近区域的随机效应的均值.

空间背景下的区域单元不是规则的, 因此条件方差 $\text{Var}[\phi_i | \boldsymbol{\phi}_{-i}] = \frac{\tau^2}{\sum_{j=1}^n w_{ij}}$ 会随着从区域到区域的移动而变化. 这在时间序列背景下并不存在这个问题, 因此我们希望方差是恒定的.

定义 3.15 Leroux CAR

The **Leroux CAR** model has a multivariate representation given by

$$\boldsymbol{\phi} \sim \mathcal{N}\left(\mathbf{0}, \tau^2 [\rho(\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}) + (1 - \rho)\mathbf{I}]\right).$$

如果 $\rho \in [0, 1)$, 则精度矩阵 $\mathbf{Q} = \frac{1}{\tau^2} [\rho(\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}) + (1 - \rho)\mathbf{I}]$ 是可逆的. 这意味着对于 $\rho \in [0, 1)$, 此模型指定了一个适当的多变量高斯分布. 此时, \mathbf{Q} 的元素为:

- $Q_{ij} = -\frac{1}{\tau^2} \rho w_{ij}$.
- $Q_{ii} = \frac{1}{\tau^2} \left(\rho \sum_{j=1}^n w_{ij} + 1 - \rho \right)$.

因此应用关于条件分布的多元高斯结果, 得到完整的条件分布:

$$\phi_i | \boldsymbol{\phi}_{-i} \sim \mathcal{N}\left(\frac{\rho \sum_{j=1}^n w_{ij} \phi_j}{\rho \sum_{j=1}^n w_{ij} + (1 - \rho)}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{ij} + (1 - \rho)}\right).$$

从这个条件规范中很容易看出, ρ 再次控制了数据中空间自相关的程度,

- $\rho = 0$: independence in space, 即 $\phi_i \sim \mathcal{N}(0, \tau^2)$, 其中均值为零且方差为常数.
- $\rho = 0$: strong spatial dependence, 即 **ICAR 模型**.

因此, 这通常是一个用于捕捉全球空间自相关的非常有用的 CAR 模型.

4 Point Process (点过程)

定义 4.1 Point process (点过程)

考虑一个空间域 $D \subset \mathbb{R}^2$, 并令 $A \subset D$. 然后令 $Z(A)$ 表示在 A 中出现的点的数量. 因此集合

$$Z = \{Z(A) : A \subset D\},$$

是一个空间点过程.

空间点过程有两种类型:

- 对于**无标记 (unmarked) 点过程**, 仅记录数据的位置.
- 对于**标记 (marked) 点过程**, 记录数据的位置, 以及每个位置的相关测量值.

空间点过程本质上就是一个随机过程, 描述的是在一个二维区域 (空间) 中, 随机“落点”的方式和数量. 这些“点”可以是: 病例, 星星, 雷达信号等.

定义 4.2 Spatial dependence (空间依赖性)

通常, 空间点过程可以表现出三种不同类型的空间依赖性:

1. **Complete spatial randomness (CSR)** 完全空间随机

The points are randomly scattered across the study region D , which corresponds to independence in space.

点在研究区域 D 中随机分布, 这对应于空间上的独立性。

2. **Clustered process** 聚集过程

The points are clustered together in groups, which corresponds to positive spatial dependence. 点聚集在一起形成群组, 这对应于正空间依赖性。

3. **Regular process** 规则过程

The points stay far away from other points, which corresponds to negative spatial dependence. 这些点彼此距离很远, 这对应于负空间依赖性。

例 4.1 via 2021.Q3(c)

Q. 如何描述一组数据在空间图上的分布?

A. The trees appear to be fairly evenly (相当均匀) spread across the region, with a few visible clusters (易见的簇) and a few visible gaps (易见的间隙). Therefore initially, **complete spatial randomness** does not look unreasonable. There also does not appear to be any pattern to the species of the trees, as both species types are found in most parts of the forest.

4.1 Point process theory (理论)

在统计学中, 兴趣通常在于估计一个均值函数, 该函数量化了感兴趣变量的预期响应。在空间点过程设置中, 均值函数对应于在任何给定区域内过程的预期点数, 这被称为其**强度 (intensity)**。

定义 4.3 First order intensity function (一阶强度函数)

设 $A \subset D$, 并设 $Z(A)$ 表示在 A 中出现的观测点的数量. 那么 A 中的点数期望为

$$\mu(A) = \mathbb{E}[Z(A)] = \int_A \lambda(\mathbf{s}) \, d\mathbf{s},$$

其中 $\lambda(\mathbf{s})$ 被称为**一阶强度函数 (first order intensity function)**, 用于衡量位置 \mathbf{s} 处点数期望的局部贡献 (local contribution).

设 $d\mathbf{s}$ 表示以点 \mathbf{s} 为中心, 面积为 $|d\mathbf{s}|$ 的小圆形区域. 那么**一阶强度函数**定义为

$$\lambda(\mathbf{s}) = \lim_{|d\mathbf{s}| \rightarrow 0} \frac{\mathbb{E}[Z(d\mathbf{s})]}{|d\mathbf{s}|},$$

为在位置 \mathbf{s} 附近点的预期数量.

定义 4.4 Second order intensity function (二阶强度函数)

Given 2 locations (\mathbf{s}, \mathbf{t}) the **second order intensity function** is defined to be

$$\lambda_2(\mathbf{s}, \mathbf{t}) = \lim_{|\mathbf{ds}| \rightarrow 0, |\mathbf{dt}| \rightarrow 0} \frac{\mathbb{E}[Z(\mathbf{ds})Z(\mathbf{dt})]}{|\mathbf{ds}||\mathbf{dt}|},$$

并测量空间点过程中的空间依赖性. 它本质上是两个随机变量乘积期望的空间点过程的类似 (analogue).

定义 4.5 Covariance density (协方差密度) & Pair correlation function (对关联函数)

The **covariance density** is defined to be:

$$\gamma(\mathbf{s}, \mathbf{t}) = \lambda_2(\mathbf{s}, \mathbf{t}) - \lambda(\mathbf{s})\lambda(\mathbf{t}),$$

and the **pair correlation function** is defined to be

$$\rho(\mathbf{s}, \mathbf{t}) = \frac{\lambda_2(\mathbf{s}, \mathbf{t})}{\lambda(\mathbf{s})\lambda(\mathbf{t})}.$$

与地统计学一样, 建模一般空间依赖性是困难的, 因此可以做出以下简化假设.

定义 4.6

- **Stationary (平稳):** 将区域 A 平移到空间任意位置, $Z(A)$ 的概率分布保持不变.
- **Isotropic (各向同性):** 将区域 A 旋转 (rotating), $Z(A)$ 的概率分布保持不变.

如果过程 Z 是平稳 且各向同性 的, 那么

- (一阶) 强度函数恒定, 即 $\lambda(\mathbf{s}) = \lambda, \mathbf{s} \in D$.
- 二阶强度函数仅取决于 \mathbf{s} 和 \mathbf{t} 之间的距离, 即 $\lambda_2(\mathbf{s}, \mathbf{t}) = \lambda_2(\|\mathbf{s} - \mathbf{t}\|)$.
- 对关联函数也仅取决于彼此的距离, 即 $\rho(\mathbf{s}, \mathbf{t}) = \rho(\|\mathbf{s} - \mathbf{t}\|)$.

4.2 Modelling and testing for complete spatial randomness (CSR) (完全随机的建模与测试)

4.2.1 A model for complete spatial randomness (CSR 的建模)

定义 4.7 Homogeneous Poisson process (HPP, 齐次泊松过程)

一个空间点过程 Z 在一个空间域 $D \subset \mathbb{R}^2$ 中是 **homogeneous Poisson process (HPP)**, 如果

1. For any measurable set $A \subset D$,

$$Z(A) \sim \text{Poisson}(\lambda|A|),$$

where $|A|$ is the area of A .

2. 在 A 中的观测点 $\{s_1, \dots, s_n\}$ 是从 A 上的均匀分布中独立抽取的随机样本.
3. 对于任何不相交的子区域 (disjoint sub-regions) $A, B \subset D$

$$Z(A) \perp Z(B), \quad \text{i.e. } Z(A) \text{ is independent of } Z(B).$$

This model corresponds to complete spatial randomness (CSR) because:

- 点的位置在空间上独立地从上述第 2 点和第 3 点中抽取.
- 一阶强度函数在整个域 D 上是恒定的, 因为:

$$\lambda(\mathbf{s}) = \lambda(\mathbf{s}) = \lim_{|\mathrm{d}\mathbf{s}| \rightarrow 0} \frac{\mathbb{E}[Z(\mathrm{d}\mathbf{s})]}{|\mathrm{d}\mathbf{s}|} = \frac{\lambda|\mathrm{d}\mathbf{s}|}{|\mathrm{d}\mathbf{s}|} = \lambda.$$

例 4.2 Bookwork 4.4.1: deriving the data likelihood

估计恒定一阶强度函数 λ , 这可以通过 MLE 来完成, 这里有两个随机量:

- 空间点过程中的点的数量: $Z(D) = n$.
- 空间点过程中的点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 的位置.

Solution. 完整的数据似然为:

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n) = f(Z(D) = n) \times f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\} | Z(D) = n)$$

1. 从齐次泊松过程的定义来看, $Z(D) \sim \text{Poisson}(\lambda|D|)$.
2. 在已知总共有 $Z(D) = n$ 个点的条件下, 空间位置点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 的联合概率密度是

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\} | Z(D) = n) = \prod_{i=1}^n f(\mathbf{s}_i | Z(D) = n) = \prod_{i=1}^n \frac{\lambda}{\lambda|D|} = \left(\frac{1}{|D|}\right)^n.$$

因此, 完整数据的似然函数为

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n) = \frac{\lambda |D|^n}{n!} \exp(-\lambda |D|) \times \left(\frac{1}{|D|} \right)^n.$$

Log-likelihood:

$$\log [f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n)] = n \log(\lambda |D|) - \log(n!) - \lambda |D| - n \log(|D|).$$

Differentiate w.r.t. λ :

$$\frac{d}{d\lambda} \log [f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n)] = n \frac{1}{\lambda |D|} |D| - |D| = \frac{n}{\lambda} - |D|.$$

Setting equal to zero and solving yields:

$$\hat{\lambda} = \frac{n}{|D|}.$$

例 4.3 Bookwork 4.4.1: covariance density & pair correlation function

考虑一个齐次泊松点过程, 然后我们可以计算其**协方差密度**和**对关联函数**.

首先, 当 $\mathbf{s} \neq \mathbf{t}$ 时, $Z(d\mathbf{s}) \perp Z(d\mathbf{t})$, 即只要 $(d\mathbf{s}, d\mathbf{t})$ 不重叠, 就有 $\mathbb{E}[Z(d\mathbf{s})Z(d\mathbf{t})] = \mathbb{E}[Z(d\mathbf{s})] \mathbb{E}[Z(d\mathbf{t})]$. 因此, 二阶强度函数为

$$\lambda_2(\mathbf{s}, \mathbf{t}) = \lim_{|d\mathbf{s}| \rightarrow 0, |d\mathbf{t}| \rightarrow 0} \frac{\mathbb{E}[Z(d\mathbf{s})Z(d\mathbf{t})]}{|d\mathbf{s}||d\mathbf{t}|} = \lim_{|d\mathbf{s}| \rightarrow 0, |d\mathbf{t}| \rightarrow 0} \frac{\mathbb{E}[Z(d\mathbf{s})] \mathbb{E}[Z(d\mathbf{t})]}{|d\mathbf{s}||d\mathbf{t}|} = \lambda(\mathbf{s})\lambda(\mathbf{t}) = \lambda^2.$$

Thus its **covariance density** is given by:

$$\gamma(\mathbf{s}, \mathbf{t}) = \lambda_2(\mathbf{s}, \mathbf{t}) - \lambda(\mathbf{s})\lambda(\mathbf{t}) = \lambda^2 - \lambda \times \lambda = 0,$$

and the **pair correlation function** is defined to be

$$\rho(\mathbf{s}, \mathbf{t}) = \frac{\lambda_2(\mathbf{s}, \mathbf{t})}{\lambda(\mathbf{s})\lambda(\mathbf{t})} = \frac{\lambda^2}{\lambda \times \lambda} = 1.$$

因此, 一个独立过程的对关联函数为 1, 因此 ‘correlation’ 这个名字并不是这个函数的最佳名称. 因此, 我们可以将泊松点过程视为空间统计中白噪声过程类似的东西.

4.2.2 Testing for complete spatial randomness (CSR 的测试)

定理 4.1 Hypothesis test based on quadrat (方格) counts

首先将一个由大小相等且不重叠的正方形 (称为 quadrat) 组成的矩形网格叠加到域 D 上, 使得它被划分为不重叠的子区域. 假设在这个规则网格中有 r 行和 c 列, 总共形成 $r \times c$ 个网格方格. 然后对于第 i 行第 j 列的网格方格:

- 令 n_{ij} 表示第 i 行和第 j 列的网格方格中的点数, 其中显然 $\sum_{i=1}^r \sum_{j=1}^c n_{ij} = n$.
- 基于 CSR, 所有网格方块都将有相同数量的预期点数, 因为强度函数 $\lambda(s) = \lambda$ 在 D 上是常数. 因此, 由于有 $r \times c$ 个网格方块, 每个网格方块中的预期点数是 $\bar{n} = \frac{n}{rc}$.

然后可以进行一个 χ^2 检验, 比较观测到的和预期的网格正方形计数, 该检验具有以下假设:

- H_0 : the point pattern is **consistent** with CSR.
- H_1 : the point pattern is **not consistent** with CSR.

The test statistic is given by:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \bar{n})^2}{\bar{n}} \sim \chi_{rc-1}^2.$$

- Under H_0 of **complete spatial randomness** the test statistic $\chi^2 \sim \chi_{rc-1}^2$, i.e.

$$F_{rc-1}^{-1}(2.5\%) < \chi^2 < F_{rc-1}^{-1}(97.5\%) \quad \Leftrightarrow \quad 0.025 < p < 0.975 \quad \Rightarrow \quad \text{fail to reject } H_0.$$

- If the point process is **clustered**, then $(n_{ij} - \bar{n})^2$ will be larger than under the random Poisson assumption, and hence you will reject H_0 by χ^2 being too **large**, p -value being too **small**, i.e.

$$\chi^2 > F_{rc-1}^{-1}(97.5\%) \quad \Leftrightarrow \quad p < 0.025 \quad \Rightarrow \quad \text{reject } H_0 \text{ in favour of } H_1.$$

- If the point process is **regular**, then $(n_{ij} - \bar{n})^2$ will be smaller than under the random Poisson assumption, and hence you will reject H_0 by χ^2 being too **small**, p -value being too **large**, i.e.

$$\chi^2 < F_{rc-1}^{-1}(2.5\%) \quad \Leftrightarrow \quad p > 0.975 \quad \Rightarrow \quad \text{reject } H_0 \text{ in favour of } H_1.$$

定义 4.8 Ripley's K -function

For homogeneous point process **Ripley's K -function** is defined to be

$$K(t) = 2\pi \int_0^t r \rho(r) \, dr,$$

其中 $\rho(r)$ 是距离为 $r = \|\mathbf{s} - \mathbf{t}\|$ 的对关联函数.

注意, 在 CSR 条件下, 过程是弱平稳且各向同性的, 因此我们有简化公式, 对于任意点 (\mathbf{s}, \mathbf{t}) :

- $\lambda(\mathbf{s}) = \lambda, \forall \mathbf{s} \in D.$
- $\lambda_2(\mathbf{s}, \mathbf{t}) = \lambda_2(r = \|\mathbf{s} - \mathbf{t}\|).$
- $\rho(\mathbf{s}, \mathbf{t}) = \rho(r = \|\mathbf{s} - \mathbf{t}\|).$

定义 4.9 Ripley's K -function (alternative)

An alternative definition for the $K(t)$ function is given by

$$K(t) = \frac{\mathbb{E} [\text{Number of points within a circle of radius } t \text{ centred at a current point}]}{\lambda},$$

圆心位于当前点, 半径为 t 的圆内点的期望数量与 λ 的比值.

总体思路是我们可以对观察到的空间点模式估计 $K(t)$, 然后将其与在完全空间随机性 (complete spatial randomness, CSR) 下预期的结果进行比较。这可以做到, 因为,

$$K(t) \begin{cases} < K_{\text{CSR}}(t), & \text{clustered process,} \\ > K_{\text{CSR}}(t), & \text{regular processes,} \end{cases} \quad \text{at a shorter distance } t.$$

例 4.4 Spatial point pattern comparison

For a homogeneous Poisson process $Z(\mathcal{C}) \sim \text{Poisson}(\lambda|\mathcal{C}|)$ we have:

$$K(t) = 2\pi \int_0^t r \rho_Z(r) \, dr = 2\pi \int_0^t r \times 1 \, dr = \pi t^2.$$

This gives the same answer as under the **alternative expectation definition** given above, because for a homogeneous Poisson process with intensity λ , the expected number of points within a circle \mathcal{C} of radius t , is $\lambda|\mathcal{C}| = \lambda\pi t^2$. Hence $K(t) = \frac{\lambda\pi t^2}{\lambda} = \pi t^2$.

在 Ripley's K function 中,

- **CSR:** 距离 r 的值为 πr^2 .
- **Clustered:** 在较小的距离 r 上会大于 πr^2 .
- **Regular:** 在较小的距离 r 上会小于 πr^2 .

因此, 我们可以绘制 $K(r) - \pi r^2$ 关于距离 r 的图, 并在 r 值较小的情况下寻找正值, 负值或接近 0 的值.

定义 4.10 Estimate K-function using method of moment estimation (MME, 矩估计)

K-function 测量的是距离过程任意点 t 处点的期望数量. 基于观测数据 $\{s_1, \dots, s_n\}$, 可以使用矩估计法估计 K-function, 其表达式为:

$$\hat{K}(t) = \frac{|D|}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} w_{ij} \mathbb{1}_{\{\|s_i - s_j\| \leq t\}}.$$

使用 R 中的 `Kest()` 函数来估计 K-function. 通过绘制 $\hat{K}(t) - \pi t^2$ vs. t 图, 来查看与 CSR 的偏差, 即在短距离 t 下:

- **Clustering:** $\hat{K}(t) - \pi t^2 > 0$.
- **Regularity:** $\hat{K}(t) - \pi t^2 < 0$.

4.3 Modelling spatially varying intensities (空间变化强度的建模)

4.3.1 Model and data likelihood (模型与数据似然)

定义 4.11 Inhomogeneous Poisson process (IPP, 非齐次泊松过程)

在空间域 $D \subset \mathbb{R}^2$ 中的空间点过程 Z 是一个非齐次泊松过程 (Inhomogeneous Poisson process, IPP), 其强度函数为 $\lambda(\mathbf{s})$, 如果:

1. 对于 D 中的任何可测集 A , 有

$$Z(A) \sim \text{Poisson}(\mu(A)), \quad \text{where } \mu(A) = \int_A \lambda(\mathbf{s}) \, d\mathbf{s},$$

其中, $\mu(A)$ 是集合 A 中预期的点的数量.

2. 观测点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 在 A 中是从一个与 $\lambda(\mathbf{s})$ 成比例的分布 A 上抽取的独立随机样本.
3. 对于任何 $A, B \subset D$, 若 $A \cap B = \emptyset$, 则 $Z(A) \perp Z(B)$.

例 4.5 Bookwork 4.5.1: deriving the data likelihood

IPP 的数据似然与 HPP 的数据似然具有相同的结构, 即存在两个随机量:

- 空间点过程中的点的数量: $Z(D) = n$.
- 空间点过程中的点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 的位置.

Solution. 完整的数据似然为:

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n) = f(Z(D) = n) \times f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\} | Z(D) = n)$$

1. 从 IPP 的定义来看, $Z(D) \sim \text{Poisson} \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right)$.
2. 在已知总共有 $Z(D) = n$ 个点的条件下, 空间位置点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 的联合概率密度是

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\} | Z(D) = n) = \prod_{i=1}^n f(\mathbf{s}_i | Z(D) = n) = \prod_{i=1}^n \frac{\lambda(\mathbf{s}_i)}{\int_D \lambda(\mathbf{s}) \, d\mathbf{s}}.$$

这既源于 IPP 中点的独立性, 也源于一个点在任何单个位置 \mathbf{s}_i 的概率与其一阶强度函数 $\lambda(\mathbf{s}_i)$ 成正比这一事实.

因此, 完整数据的似然函数为

$$f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n) = \frac{\left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right)^n}{n!} \exp \left(- \int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right) \times \prod_{i=1}^n \frac{\lambda(\mathbf{s}_i)}{\int_D \lambda(\mathbf{s}) \, d\mathbf{s}}.$$

Log-likelihood:

$$\begin{aligned} & \log [f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n)] \\ &= n \log \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right) - \log(n!) - \int_D \lambda(\mathbf{s}) \, d\mathbf{s} + \sum_{i=1}^n \left[\log(\lambda(\mathbf{s}_i)) - \log \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right) \right] \\ &= n \log \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right) - \log(n!) - \int_D \lambda(\mathbf{s}) \, d\mathbf{s} + \sum_{i=1}^n \log(\lambda(\mathbf{s}_i)) - n \log \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right) \\ &\propto \sum_{i=1}^n \log(\lambda(\mathbf{s}_i)) - \int_D \lambda(\mathbf{s}) \, d\mathbf{s} \end{aligned}$$

4.3.2 Non-parametric estimation of $\lambda(\mathbf{s})$ (强度函数的非参数估计)

如果不假设 $\lambda(\mathbf{s})$ 的任何模型或结构, 那么可以用非参数平滑方法来估计它. 最简单的方法是基于 quadrat 计数进行估计, 但这会导致一个非平滑 (马赛克状, pixelated) 的估计. 作为替代, 使用核密度估计来根据观测数据 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 来估计 $\lambda(\mathbf{s})$.

定义 4.12 Kernel density estimate (核密度估计)

一阶强度函数 (first order intensity function) 的**核密度估计 (kernel density estimate)** 为:

$$\hat{\lambda}(\mathbf{s}) = \frac{\sum_{i=1}^n k(\mathbf{s} - \mathbf{s}_i|b)}{\int_D k(\mathbf{s} - \mathbf{u}|b) d\mathbf{u}},$$

其中 $k(\cdot|b)$ 是一个**核函数 (kernel function)**, 它根据数据位置 \mathbf{s}_i 与估计点 \mathbf{s} 之间的距离来估计任何点的强度. 它是关于距离 $\|\mathbf{s}_i - \mathbf{s}\|$ 的减函数, 因此:

- 对于靠近数据点 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 的位置 \mathbf{s} , 强度会很大.
- 对于远离数据点的位置 \mathbf{s} , 强度将较小.

核函数是一个关于原点对称的双变量 (bivariate) 概率密度函数. 参数 b , 即带宽 (bandwidth) 控制平滑的程度.

- b 的值越小, 估计结果越粗糙.
- b 的值越大, 估计结果越平滑.

需要在估计中过多偏差和过多变异性之间进行权衡, 并提出了一系列不同的估计方法. 在上面强度函

数的估计中, 归一化常数

$$K(\mathbf{s}) = \int_D k(\mathbf{s} - \mathbf{u}|b) \, d\mathbf{u},$$

是一种边缘校正 (edge correction), 仅确保强度 $\hat{\lambda}_Z(\mathbf{s})$ 在正确的尺度上. 可以使用多种不同的核函数, 其中一种是高斯核函数 (Gaussian kernel function), 其表达式为:

$$k(\mathbf{s} - \mathbf{s}_i|b) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{\|\mathbf{s} - \mathbf{s}_i\|^2}{2b}\right).$$

4.3.3 Parametric models for $\lambda(\mathbf{s})$ (强度函数的参数模型)

定义 4.13 Log-linear model (对数线性模型)

一个适用于非均匀泊松过程 (IPP) 的通用**对数线性模型 (log-linear model)** 为:

$$\begin{aligned} Z(D) &\sim \text{Poisson}\left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s}\right) \\ \log(\lambda(\mathbf{s})) &= \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} \end{aligned}$$

其中 $\mathbf{x}(\mathbf{s})$ 是一个 $p \times 1$ 的协变量向量 (包括截距项的 1), 而 $\boldsymbol{\beta}$ 是 $p \times 1$ 的待估参数向量.

该模型的数据似然性, 使用 $\lambda(\mathbf{s}) = \exp(\mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta})$ 给出

$$\begin{aligned} \log[f(\{\mathbf{s}_1, \dots, \mathbf{s}_n\}, Z(D) = n)] &\propto \sum_{i=1}^n \log(\lambda(\mathbf{s}_i)) - \int_D \lambda(\mathbf{s}) \, d\mathbf{s} \\ &= \sum_{i=1}^n \log(\mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta}) - \int_D \exp(\mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}) \, d\mathbf{s}. \end{aligned}$$

无法获得封闭形式的解, 因此需要使用数值优化算法, 例如 Newton-Raphson 算法.

4.4 Overview of models for clustered point processes (聚类点过程的模型)

4.4.1 Processes with offspring (后代)

考虑以下生成空间点模式的方法:

- 从强度为 λ_P 的齐次泊松过程 (HPP) 中模拟父点 (parent).
- 对于每个父点, 独立同分布地生成 S 个后代, 其后代位置相对于父点的分布是独立同分布的, 且与父点的距离存在某种密度 $h(\cdot)$, 该密度随距离的增加而减小.

然后我们研究的点模式仅仅是后代, 这被称为**泊松聚类过程 (Poisson cluster process)**.

4.4.2 Processes generated by thinning (稀疏/减薄/抽稀)

当我们对空间点过程进行**抽稀 (thin)** 时, 会根据某种规则随机移除点. 我们讨论**独立性抽稀 (independent thinning)**. 对于每个点 s_i , 我们独立地以概率 p 保留它. 通过独立抽稀 Z , 一个强度为 λ 的齐次泊松过程 (HPP), 我们得到一个强度为 $p\lambda$ 的 HPP.

在**依赖性抽稀 (dependent thinning)** 中, 我们以随空间变化的概率保留点. 设 Z 为一个点过程, 并设 $\{0 < \pi(s) < 1 : s \in D\}$ 表示一个随机概率过程, 称之为**抽稀场 (thinning field)**, 它表示在位置 s 保留点的概率. 然后, **依赖性抽稀过程 (dependently thinned process)** Y 包含在以概率 $\{\pi(s)\}$ 抽样后保留的 Z 的实现. 也就是说, 属于过程 Z 的点 $s_i \in D$ 是:

- 以概率 $\pi(s_i)$ 保留在抽样过程 Y 中.
- 以概率 $1 - \pi(s_i)$ 从抽样过程 Y 中移除.

4.4.3 Log Gaussian Cox processes

定义 4.14 Log Gaussian Cox processes

一个定义在空间域 D 的 **log Gaussian Cox processes** 由下式给出:

$$Z(D) \sim \text{Poisson} \left(\int_D \lambda(\mathbf{s}) \, d\mathbf{s} \right)$$
$$\log(\lambda(\mathbf{s})) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + \psi(\mathbf{s})$$

其中 $\psi(\mathbf{s})$ 是一个零均值的高斯过程, 并且其空间协方差函数为 $C(\cdot)$.

它反映了空间点模式中协变量无法解释的空间结构. 如果需要, 可以移除协变量.

空间协方差函数可以是之前我们见过的一种. 比如, 一个平稳且各向同性的零点核指数模型是

$$C(h = \|\mathbf{s} - \mathbf{u}\|) = \sigma^2 \exp(-h/\phi).$$

在这个模型下, 可以分两个阶段模拟聚类数据.

1. 使用高斯过程估计一阶强度函数的自然对数.
2. 使用估计的强度函数来模拟点过程.

5 Supplement

5.1 OLS

考虑一个空间线性模型：

$$\mathbf{z} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}).$$

既然 $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, 那么 $\mathbf{z} \sim \mathcal{N}(\mathbf{x}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, 那么

$$f(\mathbf{z}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta})\right).$$

计算它的 log-likelihood function,

$$\ell(\boldsymbol{\beta}) = \log f(\mathbf{z}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta}),$$

计算极值,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left(-\frac{1}{2}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta}) \right) \\ &= -\frac{1}{2} \cdot 2\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta}) \\ &= \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{x}\boldsymbol{\beta}) \\ &= -\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{z} + \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\boldsymbol{\beta}. \end{aligned}$$

令 $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$ 得,

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{z}.$$

由于 $\mathbf{z} \sim \mathcal{N}(\mathbf{x}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, 那么

$$\begin{aligned}\text{Var} [\hat{\boldsymbol{\beta}}] &= \left((\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{\Sigma} \left((\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \right)^\top \\ &= (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \mathbf{x} (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \\ &= (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x} (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \\ &= (\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1}.\end{aligned}$$

因此, the confidence interval with significance level of 95% is

$$\left(\hat{\boldsymbol{\beta}} \pm 1.96 \sqrt{\text{Var} [\hat{\boldsymbol{\beta}}]} \right) = \left(\hat{\boldsymbol{\beta}} \pm 1.96 \sqrt{(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1}} \right).$$

如果考虑协方差为 $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, 即观测值之间独立 (忽略了空间相关性), 则

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^\top \mathbf{x})^{-1} \mathbf{x}^\top \mathbf{z}$$

因此, the confidence interval with significance level of α is

$$\left(\hat{\boldsymbol{\beta}} \pm 1.96 \cdot \sigma \sqrt{(\mathbf{x}^\top \mathbf{x})^{-1}} \right).$$

The point estimate of $\boldsymbol{\beta}$ will be unbiased but its 95% confidence interval will be too narrow. 出现这种情况的原因是: 在独立性的假设下, 错误地认为你有 n 个独立的信息片段来估计 $\boldsymbol{\beta}$. 然而, 如果数据相关, 实际上你拥有的独立信息片段少于 n 个, 导致更大的不确定性, 从而使得置信区间更宽.