Spatial Statistics

Guanxu Gleason WANG

University of Glasgow

版本：Past Paper of STAT5012 Spatial Statistics, UofG

更新：April 19, 2025

目录

1 地统计学 (Geostatistics)

1.1 变异图 (Variogram)

定义 1.1 (Weakly Stationary) A geostatistical process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is **weakly stationary** if

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

However, even stationary geostatistical processes are hard to estimate, so typically a further simplification is enforced in practice called **isotropy**. Essentially, isotropy means that the autocovariance (and autocorrelation) function is directionally invariant, so that only the distance between two points determines their autocovariance, and not the direction you have to travel to get from the first to the second.

然而，即使是静态的地质统计过程也很难估计，因此在实践中通常会进一步简化，这种简化被称为**各向同性**。本质上，各向同性意味着自协方差（以及自相关）函数在方向上是不变的，因此只有两点之间的距离决定了它们的自协方差，而不是你必须从第一点到第二点所走的路线的方向。

定义 1.2 (Isotropic) A weakly stationary geostatistical process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is **isotropic** if the autocovariance function $C(\boldsymbol{h})$ can be further simplified to

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

where $h = ||\boldsymbol{h}||$ denotes the length of the lag vector $\boldsymbol{h}$ as measured by its Euclidean distance. That is in two

dimensions for $\boldsymbol{h} = (h_1, h_2)$ we have that $||\boldsymbol{h}|| = \sqrt{(h_1^2 + h_2^2)}$.

定理 1.1 (Valid covariance function) Define mathematically the two conditions that a general function $C(h)$ must satisfy to be a valid covariance function for a weakly stationary process.

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

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

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

例 1.1 (via Note.2.2.4) Consider a weakly stationary and isotropic mean zero white noise process with autocovariance function

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

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 $(a_1, \cdots, a_n)$ and n locations $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n\}$, and let $h_{jk} = ||\boldsymbol{s}_j \boldsymbol{s}_k||$. Then

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

hence it is non-negative definite.

. . . . . .

定义 1.3 (Variogram) A **semi-variogram** of a geostatistical process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is a function denoted by $\gamma(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h})$, and measures the variance of the difference in the process at two spatial locations $\boldsymbol{s}$ and $\boldsymbol{s}+\boldsymbol{h}$. It is defined as:

$$\begin{aligned}\gamma(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h}) &= \frac{1}{2}\text{Var}\left[Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})\right] \\ &= \frac{1}{2}\text{Cov}\left[Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h}), Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})\right] \\ &= \frac{1}{2}\Big(\text{Cov}\left[Z(\boldsymbol{s}), Z(\boldsymbol{s})\right] - 2\text{Cov}\left[Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})\right] + \text{Cov}\left[Z(\boldsymbol{s}+\boldsymbol{h}), Z(\boldsymbol{s}+\boldsymbol{h})\right]\Big) \\ &= C(\boldsymbol{0}) - C(\boldsymbol{h}) \qquad \text{if weakly stationary}\end{aligned}$$

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

例 1.2 (via 2021.Q1(a)-iii) Derive the semi-variogram that corresponds to the covariance function

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

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

Thus we have that

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

例 1.3 Consider the following semi-variogram model for an isotropic geostatistical process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$,

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

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

解. 宽平稳性要求两个条件:

- **Mean constant:** mean model $\mu_Z(\boldsymbol{s}) = 0$ for all $\boldsymbol{s} \in D$.

- **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.

. . . . . .

命题 1.2 (Parameters) In summary, the following parameters are often used to describe variograms:

- ***Nugget*** (块金效应): is the limiting value of the **semi-variogram** as $h \to 0$, from the right.
  - 如果没有 Nugget, 即 Nugget $= 0$，则曲线从 $(0, 0)$ 平滑开始.
  - 如果存在 Nugget, 即 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*** (台阶高度): 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*** (影响范围): 使半变异函数达到 Sill 的距离.
- ***Partial sill*** $=$ Sill $-$ Nugget.

例 1.4 Suppose that $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ 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?

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

解. 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:***

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

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

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

3. ***Range:*** $\gamma(h)$ 在 $h\to\infty$ 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, $\lim\limits_{h\to 0^+}C(h)=\sigma^2$.

定义 1.4 (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 1.5 (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 1.6 (Spherical (球面) autocovariance function)

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

with semi-variogram

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

定义 1.7 (Wave autocovariance function)

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

with semi-variogram

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

定义 1.8 (分组经验半变异函数 (Binned Empirical Semi-variogram)) Recall that the semi-variogram for a spatial process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is defined by

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

Suppose we partition the range of distances between all pairs of points into K interval (bins)

$$I_k = (h_{k-1}, h_k], \quad k = 1, 2, \cdots, K,$$

where $0 = h_0 < h_1 < \cdots < h_K$. Let $h_k^m = (h_{k-1} + h_k)/2$ denote the midpoint of the kth interval. Then let

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

the set of pairs of points in I_k. Then the **binned empirical semi-variogram** is

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

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.

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)**。

1.2 Modelling geostatistical data

定义 1.9 (Monte Carlo envelope) Monte Carlo envelopes could be constructed by

1. generating a large number (say 1000) of simulated point patterns from 一个依题意的过程/分布; and
2. computing 需要的函数; and
3. computing the 2.5th and 97.5th percentiles of the distribution of 1000 values of 需要的函数 (95% CI), which would form the Monte Carlo envelope at that distance.

Thus if the 需要的函数 for the data lies completely inside these envelopes, then independence is likely. If it does not then spatial autocorrelation is likely.

2 Areal unit processes

2.1 定义

定义 2.1 (Areal process) An **areal process** (区域过程) is also called a **lattice process** (格过程), and is a stochastic process defined on regions that form a partition of some larger region of interest D. More formally, let $B_1, B_2, \cdots, B_n$ denote a **partition** (分区) of the domain D into n distinct regions, such that

- $\bigcup_{i=1}^{n} B_i = D$ and
- $B_i \cap B_j = \varnothing$ for each $i \neq j$.

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

2.2 Exploratory Analysis for Areal Unit Data

定义 2.2 (Moran's I) Moran's I statistic computed for data $\boldsymbol{z} = (z_1, \cdots, z_n)$ with a neighbourhood matrix $\boldsymbol{W}$ is defined as

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

where $\overline{z}$ is the mean of the data values. It measures the **strength of the linear association** in the areal data, suitably weighted for their spatial locations. As with other correlation coefficients, $-1 < I < 1$, and

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

例 2.1 (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.

定义 2.3 (Local Indicator of Spatial Association (LISA)) One approach is a local version of Moran's I, which for region i is given by

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

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

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

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

Then for area $\mathcal{A}_4$ we have

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

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 it 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）。

定义 2.4 (Geary's contiguity ratio) C is defined by

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

This is a positively valued statistic where:

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

2.3 Gaussian Markov Random Fields

Consider an areal unit process $\boldsymbol{Z} = (Z_1, \cdots, Z_n)$ relating to n areal units $\{\mathcal{A}_1, \cdots, \mathcal{A}_n\}$, and a corresponding symmetric neighbourhood matrix $\boldsymbol{W} = (w_{ij})$.

Q. 那老多 conditional autoregressive (CAR) model, 应该如何选择最适合他们数据的最优模型?

A. The researcher should fit multiple different CAR models to their data, and choose the one that best fits the data. 研究人员应该将多个不同的 *CAR* 模型拟合到他们的数据中，并选择最适合数据的模型。

For example, as these models are typically fitted in a Bayesian setting, they could measure the best model as the one that has the minimum Deviance Information Criterion (DIC) value. Fit to the data is a sensible criteria here as one typically does not do prediction with areal unit type data. 例如，由于这些模型通常在贝叶斯设置中拟合，它们可以测量最佳模型为具有最小 *DIC* 值的模型。在这里，拟合数据是一个合理的标准，因为通常不会使用面积单位类型的数据进行预测。

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

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

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

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

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

定理 2.1 (Precision matrix $Q = \Sigma^{-1}$) Consider partitioning the random vector $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$, where its joint distribution is given by

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

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

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

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

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

Taking the top row of this block matrix equation gives:

$$\begin{aligned} \boldsymbol{Q}_{11}\boldsymbol{\Sigma}_{11} + \boldsymbol{Q}_{12}\boldsymbol{\Sigma}_{21} = \boldsymbol{I} \quad &\Rightarrow \quad \boldsymbol{\Sigma}_{11} = \boldsymbol{Q}_{11}^{-1}\left(\boldsymbol{I} - \boldsymbol{Q}_{12}\boldsymbol{\Sigma}_{21}\right) \\ \boldsymbol{Q}_{11}\boldsymbol{\Sigma}_{12} + \boldsymbol{Q}_{12}\boldsymbol{\Sigma}_{22} = \boldsymbol{0} \quad &\Rightarrow \quad \boldsymbol{\Sigma}_{12} = -\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\Sigma}_{22} \end{aligned}$$

Thus subbing these into the mean and variance formula stated above that uses the variance matrix $\boldsymbol{\Sigma}$ gives:

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

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

例 2.2 (via 2021.Q2(b)-ii) The researcher decides to model the data with the following CAR model:

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

where the ijth element of precision matrix $\boldsymbol{Q}(\boldsymbol{W}, \rho)$ is given by

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

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

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

Q. 此 CAR 模型在 $w_{ij} = 0$（即如果区域 $\mathcal{A}_i, \mathcal{A}_j$ 不是邻居）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 $\mathrm{Corr}\left[Z_i, Z_j | \boldsymbol{Z}_{-ij}\right] = 0$.

3 Point Process

3.1 Exploratory analysis and models for point process data

定义 3.1 (Spatial point process) Consider a spatial domain $D \subset \mathbb{R}^2$, and let A be a subset of D. Then let $Z(A)$ denote the number of points in A. Then if the domain D is bounded and $Z(A)$ is finite for all $A \subset D$, then

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

is a **spatial point process**. 设一个空间区域 $D \subset \mathbb{R}^2$（例如一张地图的区域），再设 A 是 D 的一个子集（比如一个小区域）。定义 $Z(A)$ 为区域 A 内的“点的个数”（可以理解为事件的次数，比如树木的位置、病例出现的地点等）。如果：

- 区域 D 是有界的（bounded）；
- 对于所有 $A \subset D$，点的个数 $Z(A)$ 都是有限的；

那么我们称整个集合：

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

是一个空间点过程 spatial point process。

解释： 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。

注:

- The distribution of Z is completely determined by the joint distributions of $(Z(A_1), \ldots, Z(A_n))$ for any measurable sets $\{A_1, \ldots, A_n\}$ and for all positive integers n.
  点过程 Z 的概率分布可以被任意多个区域 $A_1, \ldots, A_n$ 的联合分布 $(Z(A_1), \ldots, Z(A_n))$ 完全刻画。
- If we only observe the locations of the data points this is an **unmarked point process**,
  如果我们只观测点的位置（如只知道树长在哪里），那么这个点过程称为**未标记点过程**,
  but if we additionally make a measurement at each data location then this is a **marked point process**.
  但如果我们在每个位置上还有附加信息（比如每棵树的高度），则这个过程是**带标记点过程**。
- An **unmarked spatial point pattern** is a realization of an unmarked spatial point process, and is a collection of n locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ in D where events of interest are observed.
  未标记空间点模式是一个未标记点过程的一次实现，即一组 n 个在区域 D 中出现的事件位置.
  解释： 点过程是一个随机模型，而“点模式”是它在某次观测中生成的数据。例如你看到地图上有 50 个病例点，那就是一次点模式。

定义 3.2 (Spatial dependence (空间依赖性)) In general, a spatial point process can exhibit three different types of spatial dependence:

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 3.1 (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) (particularly the south east) and a few visible gaps (易见的间隙) (particularly the north west). 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.

3.2 Point process theory

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)**。

Let $A \subset D$, and let $Z(A)$ be the random variable denoting the observed number of points that occurred in A. Then the expected number of points in A is given by

$$\mu(A) = \mathbb{E}\left[Z(A)\right] = \int_A \lambda(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}.$$

Here $\lambda(\boldsymbol{s})$ is called the first order intensity function, and measures the local contribution to the expected number

of points at location $\boldsymbol{s}$.

定义 3.3 (First order intensity function) Let $\mathrm{d}\boldsymbol{s}$ denote a small circular region centred around the point $\boldsymbol{s}$ with area $|\mathrm{d}\boldsymbol{s}|$. Then the **first order intensity function** is defined to be

$$\lambda(\boldsymbol{s}) = \lim_{|\mathrm{d}\boldsymbol{s}|\to 0} \frac{\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})\right]}{|\mathrm{d}\boldsymbol{s}|},$$

the expected number of points in the vicinity (附近) of location $\boldsymbol{s}$. 衡量某位置附近预期点数的函数.

定义 3.4 (Second order intensity function) Given 2 locations $(\boldsymbol{s}, \boldsymbol{t})$ the **second order intensity function** is defined to be

$$\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lim_{|\mathrm{d}\boldsymbol{s}|\to 0, |\mathrm{d}\boldsymbol{t}|\to 0} \frac{\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})Z(\mathrm{d}\boldsymbol{t})\right]}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{t}|},$$

and measures the spatial dependence in a spatial point process. 衡量空间点过程在两点间的空间依赖性. It is essentially (本质上) the point process analogue (模拟) of the expectation of a product of two random variables.

定义 3.5 (Covariance density & Pair correlation function) The **covariance density** is defined to be:

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

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

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

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。

- A spatial point process is **stationary** if probability statements about $Z(A_m)$ are invariant to arbitrarily moving (平移) A_m around in space.
- Similarly a spatial point process is **isotropic** if the distribution of $Z(A)$ is invariant to rotating (旋转) region A.

This means that if the process Z is stationary and isotropic then the (first order) intensity function is constant

$$\lambda(\boldsymbol{s}) = \lambda, \quad \boldsymbol{s} \in D,$$

and the second order intensity function only depends on the distance between $\boldsymbol{s}$ and $\boldsymbol{t}$:

$$\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lambda_2(||\boldsymbol{s} - \boldsymbol{t}||),$$

which in turn means that the pair correlation function also just depends on the distance apart, that is

$$\rho(\boldsymbol{s}, \boldsymbol{t}) = \rho(||\boldsymbol{s} - \boldsymbol{t}||).$$

3.3 Modelling and testing for complete spatial randomness

定义 3.6 (Homogeneous Poisson process (HPP)) A spatial point process Z in a spatial domain $D \subset \mathbb{R}^2$ is a **homogeneous Poisson process (HPP)** if:

1. For any measurable set $A \subset D$,
$$Z(A) \sim \text{Poisson}(\lambda|A|),$$
where $|A|$ is the area of A.
2. Observed points $\{s_1, \cdots, s_n\}$ in A are an independent random sample from the uniform distribution on A.
3. For any disjoint sub-regions $A, B \subset D$
$$Z(A) \perp Z(B).$$

This model corresponds to CSR because:

- The location of points are independently sampled over space from points 2 and 3 above.
- The first order intensity function is constant across the domain D because:

$$\lambda(\boldsymbol{s}) = \lambda(\boldsymbol{s}) = \lim_{|\mathrm{d}\boldsymbol{s}|\to 0} \frac{\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})\right]}{|\mathrm{d}\boldsymbol{s}|} = \frac{\lambda|\mathrm{d}\boldsymbol{s}|}{|\mathrm{d}\boldsymbol{s}|} = \lambda.$$

What's more, as $Z(\mathrm{d}\boldsymbol{s}) \perp Z(\mathrm{d}\boldsymbol{t})$ for $\boldsymbol{s} \neq \boldsymbol{t}$, then $\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})Z(\mathrm{d}\boldsymbol{t})\right] = \mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})\right]\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{t})\right]$ as long as $(\mathrm{d}\boldsymbol{s}, \mathrm{d}\boldsymbol{t})$ do not overlap. Hence we have that

$$\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lim_{|\mathrm{d}\boldsymbol{s}|\to 0, |\mathrm{d}\boldsymbol{t}|\to 0} \frac{\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})Z(\mathrm{d}\boldsymbol{t})\right]}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{t}|} = \lim_{|\mathrm{d}\boldsymbol{s}|\to 0, |\mathrm{d}\boldsymbol{t}|\to 0} \frac{\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{s})\right]\mathbb{E}\left[Z(\mathrm{d}\boldsymbol{t})\right]}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{t}|} = \lambda(\boldsymbol{s})\lambda(\boldsymbol{t}) = \lambda^2.$$

Thus its covariance density is given by:

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

and the pair correlation function is defined to be

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

定理 3.1 (Hypothesis test based on quadrat (方格) counts) A χ^2 test can be conducted comparing the observed and expected grid square counts, which has the following hypotheses:

- H_0: the point pattern is consistent with complete spatial randomness.
- H_1: the point pattern is not consistent with complete spatial randomness.

The test statistic is given by:

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

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

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

- If the point process is **clustered**, then $(n_{ij} - \overline{n})^2$ 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.

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

- If the point process is **regular**, then $(n_{ij} - \overline{n})^2$ 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.

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

定义 3.7 (Ripley's K-function) For homogeneous point process **Ripley's K-function** is defined to be

$$K(t) = 2\pi \int_0^t r\rho(r)\,\mathrm{d}r,$$

where $\rho(r)$ is the pair correlation function at a distance $r = ||\boldsymbol{s} - \boldsymbol{t}||$. Note, that under CSR the process is weakly stationary and isotropic and hence we have the simplification that for any points $(\boldsymbol{s}, \boldsymbol{t})$:

- $\lambda(\boldsymbol{s}) = \lambda$ for all $\boldsymbol{s} \in D$.
- $\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lambda_2(r = ||\boldsymbol{s} - \boldsymbol{t}||)$.
- $\rho(\boldsymbol{s}, \boldsymbol{t}) = \rho(r = ||\boldsymbol{s} - \boldsymbol{t}||)$.

定义 3.8 (Ripley's K-function (alternative)) An alternative definition for the $K(t)$ function is given by

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

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.

- $\lambda(\boldsymbol{s}) = \lambda$ for all $\boldsymbol{s} \in D$.
- $\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lambda_2(r = ||\boldsymbol{s} - \boldsymbol{t}||)$.
- $\rho(\boldsymbol{s}, \boldsymbol{t}) = \rho(r = ||\boldsymbol{s} - \boldsymbol{t}||)$.

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

$$K(t) = 2\pi \int_0^t r\rho_Z(r)\,\mathrm{d}r = 2\pi \int_0^t r \times 1\,\mathrm{d}r = \pi t^2.$$

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 $K(t) = \dfrac{\lambda\pi t^2}{\lambda} = \pi t^2$.

Ripley's K function for a **completely spatially random** process is given by πr^2 for a distance r, and for a **clustered process** it would be larger than this for small distances. Alternatively, the K function would be smaller than πr^2 for small distances r if the process was **regular**. Therefore we simply plot distance r vs $K(r) - \pi r^2$ and look for positive, negative or near zero values at small values of r.

4 Supplement

4.1 OLS

考虑一个空间线性模型:

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

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

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

计算它的 log-likelihood function,

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

计算极值,

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

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

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{z}.$$

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

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

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

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

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

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{x}^\top \boldsymbol{x}\right)^{-1} \boldsymbol{x}^\top \boldsymbol{z}$$

因此, the confidence interval with significance level of α is

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

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.