中国矿业大学（北京）

China University of Mining & Technology, Beijing

硕士学位论文摘要

空间广义线性混合效应模型及其应用

作　　者：黄湘云

学　　院：理学院

学　　号：TSP150701029

学科专业：统计学

导　　师：李再兴

2018 年 6 月

中图分类号：_____ 单位代码：_____

密　　级：_____

硕 士 学 位 论 文 摘 要

中文题目：空间广义线性混合效应模型及其应用

英文题目：Spatial Generalized Linear Mixed Models and Its Applications

作　　者：黄湘云　　学　　号：TSP150701029

学科专业：统计学　　研究方向：数据分析与统计计算

导　　师：李再兴　　职　　称：教授

论文提交日期：2018 年 10 月 22 日 论文答辩日期：2018 年 10 月 25 日

学位授予日期：2018 年 11 月 7 日

中国矿业大学（北京）

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国矿业大学或其他教学机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：[signature] 日期：2018年11月7日

关于论文使用授权的说明

本人完全了解中国矿业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅或借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

作者签名：[signature] 导师签名：李再兴 日期：2018.11.7

摘 要

空间广义线性混合效应模型（简称 SGLMM）在地质勘探、流行病预测和环境污染分析中扮演着重要的角色，因此实现其参数估计的算法对解决实际问题有直接的帮助，也一直是研究的重要方向。

本文充分运用空间随机过程、蒙特卡罗积分、拉普拉斯近似、极大似然估计等理论知识，研究了空间广义线性混合效应模型的参数估计及其算法实现问题，特别是空间随机效应给参数估计带来的难以处理的高维积分问题，它直接导致估计没有显式表达式，对计算造成很大的挑战，归纳了文献中存在的基于似然函数的拉普拉斯近似极大似然算法（简称 LAML）、蒙特卡罗极大似然算法（简称 MCML）和基于贝叶斯方法的 Langevin-Hastings 算法。系统总结和分析了基于似然函数的估计方法的优缺点，这类算法收敛速度快，但是要求参数初值接近真值，否则容易陷入局部极值，对此提出样本变差图和剖面似然轮廓图来帮助选择参数初值，并在小麦数据和核污染数据分析中加以应用，取得显著的效果。而基于贝叶斯方法的 Langevin-Hastings 算法迭代次数多，调参过程复杂，而且运行时间长，对此提出基于 Stan 实现的汉密尔顿蒙特卡罗算法（简称 Stan-HMC）弥补了这些不足，在响应变量分别服从泊松分布和二项分布的两组模拟实验中得到了印证。

第二章主要介绍本文中涉及到的指数族、最小二乘估计、极大似然估计、平稳高斯过程、先验后验分布和常用贝叶斯估计等相关基础知识，这里不再赘述。

第三章首先回顾了简单线性模型、广义线性模型和广义线性混合效应模型，然后引出空间广义线性混合效应模型，它是随机效应来自平稳空间高斯过程的广义线性混合效应模型，接着推导了添加空间随机效应后模型的协方差结构。SGLMM 模型的具体形式如下：

$$g(\mu_i) = d(x_i)^\top \boldsymbol{\beta} + S(x_i) + Z_i \tag{1}$$

为方便起见，记 $T_i = d(x_i)^\top \boldsymbol{\beta} + S(x_i) + Z_i$，平稳空间高斯过程 $\mathcal{S} = S(x), x \in \mathbb{R}^2$ 的理论变差 $V(x, x')$ 和模型 (1) 中 T_i 的变差 $V_T(u_{ij})$ 分别为

$$\begin{aligned} V(x,x') &= \frac{1}{2}\mathsf{Var}\{S(x) - S(x')\} \\ &= \frac{1}{2}\mathsf{Cov}(S(x) - S(x'), S(x) - S(x')) \\ &= \frac{1}{2}\{\mathsf{E}[S(x) - S(x')][S(x) - S(x')] - [\mathsf{E}(S(x) - S(x'))]^2\} \\ &= \sigma^2 - \mathsf{Cov}(S(x), S(x')) = \sigma^2\{1 - \rho(u)\} \\ V_T(u_{ij}) &= \frac{1}{2}\mathsf{Var}\{T_i(x) - T_j(x)\} \\ &= \frac{1}{2}\mathsf{E}[(T_i - T_j)^2] = \tau^2 + \sigma^2(1 - \rho(u_{ij})) \end{aligned} \tag{2}$$

根据协方差定义可推知随机向量 $\mathbf{T}=(T_1,T_2,\ldots,T_n)$ 的协方差矩阵结构如下：

$$\begin{aligned}\mathsf{Cov}(T_i(x),T_i(x))&=\mathsf{E}[S(x_i)]^2+\mathsf{E}Z_i^2=\sigma^2+\tau^2\\ \mathsf{Cov}(T_i(x),T_j(x))&=\mathsf{E}[S(x_i)S(x_j)]=\sigma^2\rho(u_{ij})\end{aligned}\tag{3}$$

平稳空间高斯过程 $\mathcal{S}$ 的自相关函数 $\rho(u)$ 为

$$\rho(u)=\{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa}\mathcal{K}_{\kappa}(u/\phi),u>0\tag{4}$$

(4) 式中函数 $\mathcal{K}_{\kappa}(\cdot)$ 的具体形式为

$$\begin{aligned}I_{-\kappa}(u)&=\sum_{m=0}^{\infty}\frac{1}{m!\Gamma(m+\kappa+1)}\Big(\frac{u}{2}\Big)^{2m+\kappa}\\ \mathcal{K}_{\kappa}(u)&=\frac{\pi}{2}\frac{I_{-\kappa}(u)-I_{\kappa}(u)}{\sin(\kappa\pi)}\end{aligned}\tag{5}$$

其中 $u\geq 0$，$\kappa\in\mathbb{R}$，如果 $\kappa\in\mathbb{Z}$，则取该点的极限值。

第四章首先介绍了目前估计 SGLMM 模型参数的算法，依次是拉普拉斯近似算法、蒙特卡罗极大似然算法、贝叶斯 Langevin-Hastings 算法和低秩近似算法。其中，有四部分补充文献中的内容：其一，推导了 SGLMM 模型似然函数的一般形式；其二，详细给出了剖面似然的思想和计算过程；其三，以卡方分布为例阐述了拉普拉斯近似的思想和计算方法，用以近似似然函数中关于空间随机效应的高维积分；其四，在贝叶斯 Langevin-Hastings 算法的基础上，提出 Stan 程序库实现的汉密尔顿蒙特卡罗算法（简称 Stan-HMC）。并分四小节进行详细介绍：第一小节，以计算 n 维超球体积的积分过程给出蒙特卡罗积分的原理和局限；第二小节，给出 Stan-HMC 算法提出的背景和意义；第三小节，归纳总结了 Stan 的历史，并以 Eight Schools 数据集为例介绍 Stan 的使用，在原来文献上补充了代码注解，平稳性检验的全过程；第四小节，给出实现 Stan-HMC 算法的过程。下面分两部分给出其重要内容，分别是在 SGLMM 模型下极大似然估计的主要推导过程和 Stan-HMC 算法的主要实现步骤。

设研究区域 $D\subseteq\mathbb{R}^2$，对于第 i 次观测，s_i 表示区域 D 内的位置，$y(s_i)$ 表示响应变量，$\mathbf{x}(s_i),i=1,\ldots,n$ 是一个 p 维的固定效应，定义如下的 SGLMM 模型：

$$\mathrm{E}[y(s_i)|u(s_i)]=g^{-1}[\mathbf{x}(s_i)^{\top}\boldsymbol{\beta}+\mathbf{u}(s_i)],\quad i=1,\ldots,n$$

其中，$g(\cdot)$ 是实值可微的逆联系函数，$\boldsymbol{\beta}$ 是 p 维的回归参数向量，代表 SGLMM 模型的固定效应。随机过程 $\{U(\mathbf{s}):\mathbf{s}\in D\}$ 是平稳的空间高斯过程，其均值为 $\mathbf{0}$，自协方差函数 $\mathsf{Cov}(U(\mathbf{s}),U(\mathbf{s}'))=C(\mathbf{s}-\mathbf{s}';\boldsymbol{\theta})$，$\boldsymbol{\theta}$ 表示其中的参数向量。$\mathbf{u}=(u(s_1),u(s_2),\ldots,u(s_n))^{\top}$ 是平稳空间高斯过程 $U(\cdot)$ 的一个实例。给定 $\mathbf{u}$ 的情况

下，观察值 $\mathbf{y}=(y(s_1),y(s_2),\ldots,y(s_n))^\top$ 是相互独立的。

给定 $u_i=u(s_i),i=1,\ldots,n$ 的条件下，$y_i=y(s_i)$ 的条件概率密度函数是

$$f(y_i|u_i;\boldsymbol{\beta})=\exp[a(\mu_i)y_i-b(\mu_i)]c(y_i)$$

其中，$\mu_i=\mathsf{E}(y_i|u_i)$，$a(\cdot),b(\cdot)$ 和 $c(\cdot)$ 是特定的函数，具体的情况视所服从的分布而定。SGLMM 模型的边际似然函数

$$L(\boldsymbol{\psi};\mathbf{y})=\int\prod_{i=1}^{n}f(y_i|u_i;\boldsymbol{\beta})\phi_n(\mathbf{u};0,\Sigma_{\boldsymbol{\theta}})\mathrm{d}\mathbf{u} \tag{6}$$

记号 $\boldsymbol{\psi}=(\boldsymbol{\beta},\boldsymbol{\theta})$ 表示 SGLMM 模型的全部参数，$\phi_n(\cdot;0,\Sigma_{\boldsymbol{\theta}})$ 表示 n 元正态密度函数，其均值为 $\mathbf{0}$，协方差矩阵为 $\Sigma_{\boldsymbol{\theta}}=(c_{ij})=(C(s_i-s_j;\boldsymbol{\theta})),i,j=1,\ldots,n$。边际似然函数 (6) 几乎总是卷入一个难以处理的积分，这是主要面临的问题，并且计算量随观测 y_i 的数量增加，此积分的维数等于观测点的个数。

再从贝叶斯方法的角度来看 SGLMM 模型，令 $\mathbf{y}=(y(s_1),\ldots,y(s_n))^\top$ 表示观测值，$\pi(\boldsymbol{\psi})$ 表示模型参数的联合先验密度，那么联合后验密度为

$$\begin{aligned}\pi(\boldsymbol{\psi},\mathbf{u}|\mathbf{y})&=\frac{f(\mathbf{y}|\mathbf{u},\boldsymbol{\psi})\phi_n(\mathbf{u};0,\Sigma_{\boldsymbol{\theta}})\pi(\boldsymbol{\psi})}{m(\mathbf{y})}\\ m(\mathbf{y})&=\int f(\mathbf{y}|\mathbf{u},\boldsymbol{\psi})\phi_n(\mathbf{u};0,\Sigma_{\boldsymbol{\theta}})\pi(\boldsymbol{\psi})\mathrm{d}\mathbf{u}\mathrm{d}\boldsymbol{\psi}\end{aligned} \tag{7}$$

同样遭遇难以处理的高维积分问题，所以 $m(\mathbf{y})$ 亦不会有显式表达式。

极大似然估计是一种被广泛接受的参数估计方法，因其优良的大样本性质，在宽松的正则条件下，极大似然估计服从渐近正态分布，满足无偏性，而且是有效的估计。为了叙述方便，似然函数能有显式表达式，考虑空间线性混合效应模型，即响应变量服从正态分布的情况，以此来介绍剖面似然估计。

$$\mathbf{Y}\sim\mathcal{N}(D\boldsymbol{\beta},\sigma^2\mathbf{R}(\phi)+\tau^2\mathbf{I}) \tag{8}$$

其中，D 是 $n\times p$ 的观测数据矩阵，$\boldsymbol{\beta}$ 是 $p\times 1$ 维的回归参数向量，$\mathbf{R}$ 依赖于 ϕ，这里 ϕ 可能含有多个参数。模型 (8) 的对数似然函数

$$\begin{aligned}L(\boldsymbol{\beta},\tau^2,\sigma^2,\phi)=&-0.5\{n\log(2\pi)+\log\{|(\sigma^2\mathbf{R}(\phi)+\tau^2\mathbf{I})|\}\\&+(\mathbf{Y}-D\boldsymbol{\beta})^\top(\sigma^2\mathbf{R}(\phi)+\tau^2\mathbf{I})^{-1}(\mathbf{Y}-D\boldsymbol{\beta})\}\end{aligned} \tag{9}$$

极大化 (9) 式就是求模型 (8) 参数的极大似然估计，极大化对数似然的过程分步如下：

1. 重参数化 $\nu^2=\tau^2/\sigma^2$，令 $V=\mathbf{R}(\phi)+\nu^2\mathbf{I}$；

2. 给定 V，对数似然函数 (9) 在

$$\begin{aligned}\hat{\boldsymbol{\beta}}(V) &= (D^\top V^{-1} D)^{-1} D^\top V^{-1} \mathbf{Y} \\ \hat{\sigma}^2(V) &= n^{-1}\{\mathbf{Y} - D\hat{\boldsymbol{\beta}}(V)\}^\top V^{-1}\{\mathbf{Y} - D\hat{\boldsymbol{\beta}}(V)\}\end{aligned} \tag{10}$$

取得极大值；

3. 将 (10) 式代入对数似然函数 (9) 式，可获得一个简化的对数似然

$$L_0(\nu^2, \phi) = -0.5\{n\log(2\pi) + n\log\hat{\sigma}^2(V) + \log|V| + n\} \tag{11}$$

4. 关于参数 ν^2, ϕ 极大化 (11) 式，获得参数 ν^2, ϕ 的估计值，再将其回代 (10) 式，获得估计值 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\sigma}^2$。

在空间线性混合效应模型的设置下，上述极大化似然函数的过程可能与自协方差函数的类型有关，如在使用 Matérn 型自协方差函数的时，平滑参数 κ 也卷入到 ϕ 中，导致识别问题。因此，让 κ 分别取 $0.5, 1.5, 2.5$，使得平稳空间高斯过程 $\mathcal{S}$ 覆盖到不同程度的均方可微性。原则上，极大似然估计的变化情况可以通过观察对数似然函数的曲面来分析，但是，似然曲面的维数往往不允许直接观察。在这种情形下，另一个基于似然的想法是剖面似然。一般地，假定有一个模型含有参数 (α, ϕ)，其参数的似然函数表示为 $L(\alpha, \phi)$。则关于 α 的剖面似然函数定义为

$$L_p(\alpha) = L(\alpha, \hat{\psi}(\alpha)) = \max_{\psi}(L(\alpha, \psi)) \tag{12}$$

即考虑似然函数随 α 的变化情况，对每一个 α （保持 α 不变），指定 ψ 的值使得对数似然取得最大值。剖面似然就是让我们可以观察到关于 α 的似然曲面，显然，其维数比完全似然曲面要低，与只有一个参数的对数似然一样，它也可以用来计算单个参数的置信区间。现在，注意到简化的对数似然 (11) 其实可以看作模型 (8) 关于 (ν^2, ϕ) 的剖面对数似然。

在估计 SGLMM 模型参数的过程中，Stan-HMC 算法先从条件分布 $S|\boldsymbol{\theta}, \boldsymbol{\beta}, Y$ 抽样，然后从条件分布 $\boldsymbol{\theta}|S$ 抽样，最后从条件分布 $\boldsymbol{\beta}|S, Y$ 抽样，具体步骤如下：

1. 选择初始值 $\boldsymbol{\theta}, \boldsymbol{\beta}, S$，如 $\boldsymbol{\beta}$ 的初始值来自正态分布，$\boldsymbol{\theta}$ 的初始值来自对数正态分布；
2. 更新参数向量 $\boldsymbol{\theta}$：
   (i) 从指定的先验分布中均匀抽取新的 $\boldsymbol{\theta}'$；
   (ii) 以概率 $\Delta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{\frac{p(S|\boldsymbol{\theta}')}{p(S|\boldsymbol{\theta})}, 1\right\}$ 接受 $\boldsymbol{\theta}'$，否则不改变 $\boldsymbol{\theta}$。
3. 更新高斯过程 S 的取值：

(i) 抽取新的值 S_i'，向量 S 的第 i 值来自一元条件高斯密度 $p(S_i'|S_{-i},\boldsymbol{\theta})$，$S_{-i}'$ 表示移除 S 中的第 i 个值；

(ii) 以概率 $\Delta(S_i,S_i')=\min\left\{\frac{p(y_i|s_i',\beta)}{p(y_is_i,\beta)},1\right\}$ 接受 S_i'，否则不改变 S_i；

(iii) 重复 (i) 和 (ii) $\forall i=1,2,\ldots,n$。

4. 更新模型系数 $\boldsymbol{\beta}$：从条件密度 $p(\boldsymbol{\beta}'|\boldsymbol{\beta})$ 以概率

$$\Delta=\min\Big\{\frac{\prod_{j=1}^n p(y_i|s_i,\boldsymbol{\beta}')p(\boldsymbol{\beta}|\boldsymbol{\beta}')}{\prod_{j=1}^n p(y_i|s_i,\boldsymbol{\beta})p(\boldsymbol{\beta}'|\boldsymbol{\beta})},1\Big\}$$

接受 $\boldsymbol{\beta}'$，否则不改变 $\boldsymbol{\beta}$；

5. 重复步骤 2，3，4 既定的次数，获得参数 $\boldsymbol{\beta},\boldsymbol{\theta}$ 的迭代序列，直到参数的迭代序列平稳，然后根据后续的平稳序列采样，获得各参数后验分布的样本，再根据样本估计参数值。

第五章首先模拟了一维和二维情形下平稳空间高斯过程，在二维情形下，又分别模拟了响应变量服从二项分布和泊松分布的 SGLMM 模型，比较了贝叶斯 Langevin-Hastings 算法和我们提出的贝叶斯 Stan-HMC 算法，在获得相似估计效果的情形下，我们提出的算法所需迭代次数少，迭代初值可以随机生成，运行时间短。下面就一维和二维情形下，给出平稳空间高斯过程的模拟过程和二项型 SGLMM 模型的模拟过程，数值实验获得的图表不再赘述。

一维情形下，平稳高斯过程 $S(x)$ 的自协方差函数采用幂指数型，见公式 (14)。特别地，当 $\kappa=1$ 时，自协方差函数为指数型，见公式 (13)。下面分 $\kappa=1$ 和 $\kappa=2$，模拟两个一维平稳空间高斯过程，协方差参数均为 $\sigma^2=1$，$\phi=0.15$，均值向量都是 $\mathbf{0}$，在 $[-2,2]$ 的区间上，产生 2000 个服从均匀分布的随机数，由这些随机数的位置和协方差函数公式 (13) 或 (14) 计算得到 2000 维的高斯分布的协方差矩阵 G，为保证协方差矩阵的正定性，在矩阵对角线上添加扰动 1×10^{-12}，然后即可根据 Cholesky 分解该对称正定矩阵，得到下三角块 L，使得 $G=L\times L^{\top}$，再产生 2000 个服从标准正态分布的随机向量 η，而 $L\eta$ 即为所需的服从平稳高斯过程的一组随机数。

$$\mathsf{Cov}(S(x_i),S(x_j))=\sigma^2\exp\Big\{-\frac{|x_i-x_j|}{\phi}\Big\} \tag{13}$$

$$\mathsf{Cov}(S(x_i),S(x_j))=\sigma^2\exp\Big\{-\Big(\frac{|x_i-x_j|}{\phi}\Big)^{\kappa}\Big\},0<\kappa\leq2 \tag{14}$$

二维情形下，在规则平面上模拟平稳高斯过程 $\mathcal{S}=S(x),x\in\mathbb{R}^2$，其均值向量为零向量 $\mathbf{0}$，协方差函数为指数型，见公式 (13)，协方差参数 $\phi=1,\sigma^2=1$。在单位平面区域为 $[0,1]\times[0,1]$ 模拟服从上述二维平稳空间高斯过程，不妨将此区域划分为 6×6 的小网格，而每个格点作为采样的位置，共计 36 个采样点，在这些采样点上的观察值

即为目标值 $S(x)$。类似模拟一维平稳空间过程的步骤，首先根据采样点位置坐标和协方差函数 (13) 计算得目标空间过程的 $\mathcal{S}$ 协方差矩阵 G，然后使用 R 包 MASS 提供的 `mvrnorm` 函数产生多元正态分布随机数，与模拟一维平稳空间过程不同的是这里采用特征值分解，即 $G = L\Lambda L^{\top}$，与 Cholesky 分解相比，特征值分解更稳定些，但是 Cholesky 分解更快，Stan 即采用此法，后续过程与一维模拟一致。

响应变量服从二项分布 $Y_i \sim \text{Binomial}(m_i, p(x_i))$，即在位置 x_i 处以概率 $p(x_i)$ 重复抽取了 m_i 个样本，总样本数 $M = \sum_{i=1}^{N} m_i$，N 是采样点的个数，模拟二项型空间广义线性混合效应模型为 (15)，联系函数为 $g(\mu_i) = \log\{\frac{p(x_i)}{1-p(x_i)}\}$，$S(x)$ 是均值为 $\mathbf{0}$，协方差函数为 $\mathsf{Cov}(S(x_i), S(x_j)) = \sigma^2\Big\{2^{\kappa-1}\Gamma(\kappa)\Big\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi), \kappa = 0.5$ 的平稳空间高斯过程。

$$g(\mu_i) = \log\Big\{\frac{p(x_i)}{1-p(x_i)}\Big\} = \alpha + S(x_i) \tag{15}$$

其中，固定效应参数 $\alpha = 0$，协方差参数为 $\boldsymbol{\theta} = (\sigma^2, \phi) = (0.5, 0.2)$，采样点数目为 $N = 64$，每个采样点抽取的样本数 $m_i = 4, i = 1, 2, \ldots, 64$，则 Y_i 的取值范围为 $0, 1, 2, 3, 4$。首先模拟平稳空间高斯过程 $S(x)$，在单位区域 $[0,1] \times [0,1]$ 划分为 8×8 的网格，格点选为采样位置，用 geoR 包提供的 `grf` 函数产生协方差参数为 $\boldsymbol{\theta} = (\sigma^2, \phi) = (0.5, 0.2)$ 的平稳空间高斯过程，由公式 (15) 可知 $p(x_i) = \exp[\alpha + S(x_i)]/\{1 + \exp[\alpha + S(x_i)]\}$，即每个格点处二项分布的概率值，然后依此概率，由 `rbinom` 函数产生服从二项分布的观察值 Y_i。模拟响应变量服从泊松分布的情形与此类似，不再赘述。

第六章给出了两个案例分析，分别是基于空间线性混合效应模型的小麦数据分析和基于泊松型空间广义线性混合效应模型的核污染数据分析，我们发现基于样本变差图和剖面似然轮廓图等可视化辅助手段可以获得非常好的模型参数初始值，这对于基于似然函数的参数估计算法选初值很有帮助。

关键词：空间随机效应，拉普拉斯近似，蒙特卡罗方法，Stan-HMC

Abstract

The spatial generalized linear mixed-effects model (SGLMM) plays an important role in geological exploration, epidemic prediction and environmental pollution analysis. Therefore, the algorithm to estimate its parameters has a direct help to solve practical problems, and it has always been an vital direction of research.

The spatial stochastic process, Monte Carlo integral, Laplace approximation, maximum likelihood estimation and other theoretical knowledge are used to estimate the parameters and algorithms of SGLMM in the paper. The high-dimensional integral from spatial random effect is analytically intractable in general, which gives a great challenge to the calculation.

The purpose of the paper is to conclude the Laplace approximation maximum likelihood (LAML), Monte Carlo maximum likelihood (MCML) and Langevin-Hastings algorithm in the literature and summarizes systematically the advantages and disadvantages of parameter estimations. The algorithms standing on the shoulders of the likelihood function have a fast convergence speed, but they require really good initial values of the parameters, otherwise they usually fall into the local extremum. So, the sample variation diagram and the profile likelihood contour are proposed to select the initial values of the parameters, and they are also applied in the analysis of wheat data and nuclear pollution data, respectively. Langevin-Hastings algorithm must iterate many times, tune parameters repeatedly and run for a long time while the Hamilton Monte Carlo algorithm programmed with Stan (Stan-HMC) relieves the shortcomings significantly.

The second chapter mainly introduces the related basic knowledge of the exponential family, least squares estimation, maximum likelihood estimation and stationary Gaussian process which are involved in this paper and will not be repeated here.

The third chapter derives the covariance structure of the SGLMM with spatial random effects. The specific form of the SGLMM is as follows:

$$g(\mu_i) = d(x_i)^\top \boldsymbol{\beta} + S(x_i) + Z_i \tag{16}$$

Let $T_i = d(x_i)^\top \boldsymbol{\beta} + S(x_i) + Z_i$ and $\mathcal{S} = S(x), x \in \mathbb{R}^2$. Then the theoretical variograms of T_i and $S(x)$ are $V_T(u_{ij})$ and $V(x, x')$ respectively and detailed by

$$\begin{aligned} V(x,x') &= \frac{1}{2}\mathsf{Var}\{S(x) - S(x')\} = \frac{1}{2}\mathsf{Cov}(S(x) - S(x'), S(x) - S(x')) \\ &= \frac{1}{2}\{\mathsf{E}[S(x) - S(x')][S(x) - S(x')] - [\mathsf{E}(S(x) - S(x'))]^2\} \\ &= \sigma^2 - \mathsf{Cov}(S(x), S(x')) = \sigma^2\{1 - \rho(u)\} \end{aligned} \tag{17}$$

$$\begin{aligned} V_T(u_{ij}) &= \frac{1}{2}\mathsf{Var}\{T_i(x) - T_j(x)\} \\ &= \frac{1}{2}\mathsf{E}[(T_i - T_j)^2] = \tau^2 + \sigma^2(1 - \rho(u_{ij})) \end{aligned} \tag{18}$$

According to the definition of covariance, the structure of covariance matrix of random vector $\mathbf{T} = (T_1, T_2, \ldots, T_n)$ can be inferred by

$$\begin{aligned} \mathsf{Cov}(T_i(x), T_i(x)) &= \mathsf{E}[S(x_i)]^2 + \mathsf{E}Z_i^2 = \sigma^2 + \tau^2 \\ \mathsf{Cov}(T_i(x), T_j(x)) &= \mathsf{E}[S(x_i)S(x_j)] = \sigma^2\rho(u_{ij}) \end{aligned} \tag{19}$$

where $\rho(u)$ denotes the Matérn family of correlation functions of stationary spatial Gaussian process $\mathcal{S}$ and detailed by

$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa}\mathcal{K}_{\kappa}(u/\phi), u > 0 \tag{20}$$

in which $\mathcal{K}_{\kappa}(\cdot)$ is the modified Bessel function of the second kind.

$$\begin{aligned} I_{-\kappa}(u) &= \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m + \kappa + 1)}\Big(\frac{u}{2}\Big)^{2m+\kappa} \\ \mathcal{K}_{\kappa}(u) &= \frac{\pi}{2}\frac{I_{-\kappa}(u) - I_{\kappa}(u)}{\sin(\kappa\pi)} \end{aligned} \tag{21}$$

where $u \geq 0$, $\kappa \in \mathbb{R}$. Specifically, $\mathcal{K}_{\kappa}(u)$ gets the limit value when $\kappa \in \mathbb{Z}$.

The fourth chapter introduces the algorithms to estimate the parameters of SGLMM. There are four parts to supplement the contents of the literature. Firstly, the general form of the likelihood function of the SGLMM model is derived. Secondly, the idea and calculation process of the profile likelihood are given in detail. Thirdly, the chi-square distribution is taken as an example for explaining the idea and calculation method behind Laplace approximation which is expounded to approximate the high-dimensional integral of the spatial random effect in the likelihood function. Fourthly, based on the Bayesian Langevin-Hastings algorithm, Hamilton Monte Carlo algorithm (Stan-HMC) is proposed.

Stan-HMC is divided into four sections for detailed introduction: in the first section, the integration process of calculating the n dimension hypersphere volume gives the principle and limitation of Monte Carlo integration. The second section introduces the background and significance of the Stan-HMC. The third section summarizes the history of Stan and details the use of Stan with the Eight Schools dataset. It supplements the code annotation and the whole process of stationarity test compared with the original literature. The fourth section gives the implementation of Stan-HMC. The following is the main derivation process of the maximum likelihood estimation and iteration steps of the Stan-HMC.

Let $D \subseteq \mathbb{R}^2$ be the region of interest and for the ith observation, let s_i be some spatial location within D and $y(s_i)$ be the response variable, $\mathbf{x}(s_i), i = 1, \ldots, n$ be a $q-$dimensional vector of covariates for the fixed effects. The SGLMM as follows:

$$\mathrm{E}[y(s_i)|u(s_i)] = g^{-1}[\mathbf{x}(s_i)^\top \boldsymbol{\beta} + \mathbf{u}(s_i)], \quad i = 1, \ldots, n$$

where, $g(\cdot)$ is a real-value differentiable and invertible link function and $\boldsymbol{\beta}$ is a vector of p regression parameters. $\{U(\mathbf{s}) : \mathbf{s} \in D\}$ is stationary spatial Gaussian process with zero mean and spatial covariance function $\mathsf{Cov}(U(\mathbf{s}), U(\mathbf{s}')) = C(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of correlation parameters. $\mathbf{u} = (u(s_1), u(s_2), \ldots, u(s_n))^\top$ is a realization of $U(\cdot)$. Conditionally on $\mathbf{u}$, $\mathbf{y} = (y(s_1), y(s_2), \ldots, y(s_n))^\top$ are mutually independent. The form of conditional density function of $y_i = y(s_i)$ given $u_i = u(s_i), i = 1, \ldots, n$ is

$$f(y_i|u_i; \boldsymbol{\beta}) = \exp[a(\mu_i)y_i - b(\mu_i)]c(y_i)$$

where $\mu_i = \mathsf{E}(y_i|u_i)$, $a(\cdot), b(\cdot)$ and $c(\cdot)$ are specific functions. Then, the marginal likelihood function of SGLMM will be

$$L(\boldsymbol{\psi}; \mathbf{y}) = \int \prod_{i=1}^{n} f(y_i|u_i; \boldsymbol{\beta}) \phi_n(\mathbf{u}; 0, \Sigma_{\boldsymbol{\theta}}) \mathrm{d}\mathbf{u} \tag{22}$$

where $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ are the parameters of the model and $\phi_n(\cdot; 0, \Sigma_{\boldsymbol{\theta}})$ is the $n-$variate normal density function with zero mean and covariance matrix $\Sigma_{\boldsymbol{\theta}} = (c_{ij}) = (C(s_i - s_j; \boldsymbol{\theta})), i, j = 1, \ldots, n$. Here the calculation of marginal function (22) nearly always involves intractable integrals, which is the main obstacle. Also the computational burden increases with the number of observations because the dimension of covariance matrix is equal to the number of observations.

Now, Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))^\top$ be the observed data and $\pi(\boldsymbol{\psi})$ be the joint prior density of the parameters. Then the joint posterior density is defined by

$$\begin{aligned} \pi(\boldsymbol{\psi}, \mathbf{u}|\mathbf{y}) &= \frac{f(\mathbf{y}|\mathbf{u}, \boldsymbol{\psi}) \phi_n(\mathbf{u}; 0, \Sigma_{\boldsymbol{\theta}}) \pi(\boldsymbol{\psi})}{m(\mathbf{y})} \\ m(\mathbf{y}) &= \int f(\mathbf{y}|\mathbf{u}, \boldsymbol{\psi}) \phi_n(\mathbf{u}; 0, \Sigma_{\boldsymbol{\theta}}) \pi(\boldsymbol{\psi}) \mathrm{d}\mathbf{u} \mathrm{d}\boldsymbol{\psi} \end{aligned} \tag{23}$$

which is not available in closed form because of the same intractable integrals that cause trouble in the likelihood function.

Maximum likelihood estimation is a widely accepted statistical method, with well-known optimality properties in large samples. Under mild regularity conditions, the

maximum likelihood estimator is asymptotically normally distributed, unbiased and fully efficient.

For convenience of description, the likelihood function can have an explicit expression, and the spatial linear mixed-effect model is considered, that is, the response variable obeys the normal distribution.

$$\mathbf{Y} \sim \mathcal{N}(D\boldsymbol{\beta}, \sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I}) \tag{24}$$

where D is and $n \times p$ matrix of covariates, $\boldsymbol{\beta}$ is the corresponding vector of regression parameters, and $\mathbf{R}$ depends on a scalar or vector-valued parameter ϕ. The log-likelihood function of the model is

$$\begin{aligned} L(\boldsymbol{\beta}, \tau^2, \sigma^2, \phi) = & -0.5\{n\log(2\pi) + \log\{|(\sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I})|\} \\ & + (\mathbf{Y} - D\boldsymbol{\beta})^\top(\sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I})^{-1}(\mathbf{Y} - D\boldsymbol{\beta})\} \end{aligned} \tag{25}$$

maximisation of which yields the maximum likelihood estimates of the model parameters. An algorithm for maxmisation of log-likelihood proceeds as follows.

1. We reparameterise to $\nu^2 = \tau^2/\sigma^2$ and let $V = \mathbf{R}(\phi) + \nu^2\mathbf{I}$;
2. Given V, the log-likelihood function is maximised at

$$\begin{aligned} \hat{\boldsymbol{\beta}}(V) &= (D^\top V^{-1} D)^{-1} D^\top V^{-1}\mathbf{Y} \\ \hat{\sigma}^2(V) &= n^{-1}\{\mathbf{Y} - D\hat{\boldsymbol{\beta}}(V)\}^\top V^{-1}\{\mathbf{Y} - D\hat{\boldsymbol{\beta}}(V)\} \end{aligned} \tag{26}$$

3. By substituting the above expressions for $\hat{\boldsymbol{\beta}}(V)$ and $\hat{\sigma}^2(V)$ into the log-likelihood function, we obtain a concentrated log-likelihood

$$L_0(\nu^2, \phi) = -0.5\{n\log(2\pi) + n\log\hat{\sigma}^2(V) + \log|V| + n\} \tag{27}$$

4. This must then be optimised numerically with respect to ν^2 and ϕ, followed by back substitution to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$.

The practical details of the optimisation may depend on the particular family under consideration. For example, when using the Matérn covariance function, the shape parameter κ is ofen poorly identified. Therefore, let κ from a discrete set, for example $\{0.5, 1.5, 2.5\}$, cover different degrees of mean-square differentiability of the underlying stationary spatial Gaussian process.

In principle, the variability of maximum likelihood estimates can be investigated by inspection of the log-likelihood surface. However, the typical dimension of this surface does not allow direct inspection. Another generic likelihood-based idea which is useful in this

situation is that profile likelihood. Suppose, in general, that we have a model with parameters (α, ϕ) and denote its likelihood by $L(\alpha, \phi)$. We define the profile likelihood for α by

$$L_p(\alpha) = L(\alpha, \hat{\psi}(\alpha)) = \max_{\psi}(L(\alpha, \psi)) \tag{28}$$

In other words, we consider how the likelihood varies with respect to α when, for each value of α we assign to ψ the value which maximises the log-likelihood with α held fixed. The profile log-likelihood allows us to inspect a likelihood surface for α, which is of lower dimension than the full likelihood surface. It can also be used to calcute confidence intervals for individual parameters, exactly as in the case of ordinary log-likelihood for a single parameter model. Note that the concentrated log-likelihood (27), which we introduced as a computational device for maximum likelihood estimation, can now be seen to be the profile log-likelihood surface for (ν^2, ϕ) in the model (24).

For esitmating parameters of SGLMM, and noting that the data Y are fixed, a single cycle of the Stan-HMC algorithms involves first sampling from $S|\boldsymbol{\theta}, \boldsymbol{\beta}, Y$, then from $\boldsymbol{\theta}|S$, and finally from $\boldsymbol{\beta}|S, Y$, detailed iteration steps as follow:

1. Choose initial values for $\boldsymbol{\theta}, \boldsymbol{\beta}, S.$ such as $\boldsymbol{\beta}$ from normal distribution and $\boldsymbol{\theta}$ from log-normal distribution;
2. Update all the components of the parameter vector $\boldsymbol{\theta}$:

   (i) choose a new proposed value $\boldsymbol{\theta}'$ by sampling uniformly from the parameter space specified by the prior;

   (ii) accept $\boldsymbol{\theta}'$ with probability $\Delta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{\frac{p(S|\boldsymbol{\theta}')}{p(S|\boldsymbol{\theta})}, 1\right\}$, otherwise leave $\boldsymbol{\theta}$ unchanged.

3. Update S:

   (i) choose a new proposed value, S_i', for ith component of S from the univariate Gaussian conditional probability density $p(S_i'|S_{-i}, \boldsymbol{\theta})$, where S'_{-i} denotes S with its ith element removed;

   (ii) accept S_i' with probability $\Delta(S_i, S_i') = \min\left\{\frac{p(y_i|s_i', \beta)}{p(y_i s_i, \beta)}, 1\right\}$, otherwise leave S_i unchanged;

   (iii) repeat (i) and (ii) $\forall i = 1, 2, \ldots, n$.

4. Update all the elements of the regression parameter $\boldsymbol{\beta}$:

   (i) choose a new proposed value $\boldsymbol{\beta}'$ from conditional density $p(\boldsymbol{\beta}'|\boldsymbol{\beta})$;

(ii) accept $\boldsymbol{\beta}'$ with probability

$$\Delta = \min\Big\{\frac{\prod_{j=1}^{n} p(y_i|s_i,\boldsymbol{\beta}')p(\boldsymbol{\beta}|\boldsymbol{\beta}')}{\prod_{j=1}^{n} p(y_i|s_i,\boldsymbol{\beta})p(\boldsymbol{\beta}'|\boldsymbol{\beta})},1\Big\},$$

otherwise leave $\boldsymbol{\beta}$ unchanged.

5. repeat step 2，3，4 with given iteration numbers to gain chains of all parameters $\boldsymbol{\beta},\boldsymbol{\theta}$ so as to guarantee convergence of the chains to the required equilibrium distribution. According to the stationary subsequence, parameter estimations are gained by sampling samples of the posterior distribution of each parameter.

The fifth chapter firstly simulates the stationary spatial Gaussian process in one-dimensional and two-dimensional cases. In the two-dimensional case, the SGLMM with response variables obeying binomial distribution and Poisson distribution is simulated respectively. The proposed Bayesian Stan-HMC algorithm compared with the Bayesian Langevin-Hastings algorithm hold similar estimation results. The Stan-HMC algorithm requires fewer iterations, and its initial values can be randomly generated and its running time is shorter. In the following one-dimensional and two-dimensional cases, the simulation process of the stationary spatial Gaussian process and the simulation process of the binomial SGLMM are given. The figures and tables obtained by numerical experiments are not showed here.

In the one-dimensional case, the covariance function of the stationary Gaussian process $S(x)$ is exponentiated quadratic, see the formula (30). In particular, when $\kappa = 1$, the auto-covariance function is exponential, see the formula (29).

The following divides $\kappa = 1$ and $\kappa = 2$ to simulate two one-dimensional stationary spatial Gaussian processes. The covariance parameters are $\sigma^2 = 1$ and $\phi = 0.15$ with the mean vectors are both $\mathbf{0}$. In the interval of $[-2, 2]$, we generate 2000 random numbers from uniform distribution. On the basis of the locations and covariance function formula (29) or (30), we get the covariance matrix G of the 2000-dimensional Gaussian distribution.

To ensure the positive definiteness of the covariance matrix G, we add the noise, 1×10^{-12}, on the diagonal of the matrix. So, the symmetric positive definite matrix G can be decomposed using Cholesky and the lower triangular block L is obtained simultaneously because of $G = L \times L^{\top}$. And then we generate 2000 random numbers denoted by η from normal distribution with mean 0 and standard deviation 1. $L\eta$ is the random vector required from the stationary Gaussian process.

$$\mathsf{Cov}(S(x_i), S(x_j)) = \sigma^2 \exp\Big\{-\frac{|x_i - x_j|}{\phi}\Big\} \tag{29}$$

$$\mathsf{Cov}(S(x_i), S(x_j)) = \sigma^2 \exp\Big\{ -\Big(\frac{|x_i - x_j|}{\phi}\Big)^{\kappa}\Big\}, 0 < \kappa \leq 2 \tag{30}$$

In the two-dimensional case, the stationary Gaussian process $\mathcal{S} = S(x), x \in \mathbb{R}^2$ is simulated on the regular plane with zero mean vector $\mathbf{0}$ and covariance parameter $\phi = 1, \sigma^2 = 1$. $S(x)$ obeys two-dimensional stationary spatial Gaussian process in the unit plane area $[0, 1] \times [0, 1]$ where divided into grid of 6×6 and each grid point as the sampling locations. There are a total of 36 sampling points and the corresponding observed value is the target value $S(x)$.

Similar to the step of simulating a one-dimensional stationary spatial process. We first calculate the $\mathcal{S}$ covariance matrix G of the target spatial process using the sampling point coordinates and the covariance function (29). Then use the `mvrnorm` function provided by the R package MASS to generate random numbers from multivariate normal distribution. Unlike the simulated one-dimensional stationary spatial process, the eigenvalue decomposition is used here with $G = L\Lambda L^{\top}$. Compared with Cholesky decomposition, eigenvalue decomposition is more stable, however, Cholesky decomposition which adopted by Stan is faster. The subsequent steps are consistent with one-dimensional simulation.

Response variable obeys binomial distribution $Y_i \sim \mathrm{Binomial}(m_i, p(x_i))$, that is to say, repeatedly extracted m_i samples with probility $p(x_i)$ at location x_i. So, total number of samples $M = \sum_{i=1}^{N} m_i$ where N denotes the number of locations.

Here, we simulate binomial SGLMM (31) with link function $g(\mu_i) = \log\{\frac{p(x_i)}{1-p(x_i)}\}$ and $S(x)$ is two-dimensional stationary spatial Gaussian process with zero mean $\mathbf{0}$ and covariance function $\mathsf{Cov}(S(x_i), S(x_j)) = \sigma^2\Big\{2^{\kappa-1}\Gamma(\kappa)\Big\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi), \kappa = 0.5$.

$$g(\mu_i) = \log\Big\{\frac{p(x_i)}{1 - p(x_i)}\Big\} = \alpha + S(x_i) \tag{31}$$

where the fixed effect parameter $\alpha = 0$, the covariance parameter $\boldsymbol{\theta} = (\sigma^2, \phi) = (0.5, 0.2)$ and the number of sampling points $N = 64$. The number of samples each location is $m_i = 4, i = 1, 2, \ldots, 64$. Then the range of Y_i is $0, 1, 2, 3, 4$.

Firstly we simulate the stationary spatial Gaussian process $S(x)$ in the unit area $[0, 1] \times [0, 1]$ where divided into 8×8 grid point selected as the sampling locations with the `grf` function provided by geoR which generates stationary spatial Gaussian process with covariance parameters $\boldsymbol{\theta} = (\sigma^2, \phi) = (0.5, 0.2)$. By the formula (31), we know that $p(x_i) = \exp[\alpha + S(x_i)]/\{1 + \exp[\alpha + S(x_i)]\}$ is the probability value of the binomial distribution. According to this probability, Y_i is generated by the `rbinom` function. The response variable obeys the Poisson distribution is similar to this and will not be described again.

The sixth chapter gives two case studies, namely wheat data analysis with spatial linear

mixed effect model and nuclear pollution data analysis with Poisson spatial generalized linear mixed effect model. We can obtain very good initial values of model parameters by sample variogram and profile likelihood contour which is helpful for parameter estimation algorithms using likelihood function.

Key words：spatial random effects, laplace approximation, monte carlo methods, Stan-HMC