

非线性混合效应模型统一标准形式及应用

献给严士健教授 90 华诞

符利勇, 唐守正*

中国林业科学研究院资源信息研究所, 国家林业和草原局森林经营与生长模拟重点实验室, 北京 100091

E-mail: fuly@ifrit.ac.cn, stang@caf.ac.cn

收稿日期: 2018-07-01; 接受日期: 2019-08-21; 网络出版日期: 2020-01-06; * 通信作者
国家自然科学基金 (批准号: 31570628) 资助项目

摘要 非线性混合效应模型 (nonlinear mixed effect model, NLMEM) 对分析重复调查数据有着突出优势并且得到广泛应用. 但是现有的混合模型仅包含单水平和逐级嵌套多水平的随机效应分析, 没有包括交互作用及其他组合类型随机效应分析, 因此, 现有的 NLMEM 没有一个统一的标准表达式. 本文的目的在于提出一个包括所有随机效应类型 NLMEM 的统一表达式, 并且给出正态分布 NLMEM 的一种参数计算方法 (线性逼近 - 逐步 2 次规划). 通过落叶松实例和数值模拟验证表明, 线性逼近 - 逐步 2 次规划算法具有较高的计算精度和较快的计算速度. 该算法已在 ForStat 2.2 版本上实现. 除此之外, NLMEM 统一表达式还可推广到组变量情形, 即考虑组变量的 NLMEM.

关键词 交错随机效应 嵌套随机效应 一阶条件期望 非线性混合效应模型 逐步 2 次规划算法

MSC (2010) 主题分类 60J60

1 引言

非线性混合效应模型 (nonlinear mixed effect model, NLMEM) 是依据回归函数依赖于固定效应参数和随机效应参数的非线性关系而建立的 (参见文献 [1]). 它是分析纵向数据、多水平数据及重复调查数据的近代统计学方法, 能反映总体的平均变化趋势, 又能刻画个体差异 (参见文献 [2, 3]). 近几年, NLMEM 受到越来越多的学者关注, 并且被引入到多门学科, 如医学、工学、农学和林学等 (参见文献 [3-8]).

Pinheiro 和 Bates^[3] 依据随机效应因子 (产生随机效应的因子) 个数把 NLMEM 分为单水平 (single-level) NLMEM 和嵌套多水平 (multi-level) NLMEM. (为了含义确切, 本文把 multi-level 译为嵌套多水平, 它是 2 个以上逐级嵌套随机因素所产生的效应. 因为多因素产生的随机效应除嵌套效应外

英文引用格式: Fu L Y, Tang S Z. A general formulation of nonlinear mixed effect models and its application (in Chinese). *Sci Sin Math*, 2020, 50: 15-30, doi: 10.1360/N012018-00173

还包括其他多种类型, 本文将研究所有这些类型的随机效应. 以后本文中的多因素不仅包括嵌套多水平也包括其他类型.) 以单水平 NLMEM 为例, 表达式如下:

$$\begin{cases} y_{ij} = f(\phi_{ij}, v_{ij}) + \varepsilon_{ij}, & i = 1, \dots, M, \quad j = 1, \dots, n, \\ \phi_{ij} = A_{ij}\beta + B_{ij}u_i, & u_i \sim N(0, \Psi), \end{cases} \quad (1.1)$$

其中 M 为研究对象个数; n_i 为第 i 个研究对象重复观测次数; y_{ij} 为第 i 个研究对象第 j 次观测的因变量值; f 是关于参数向量 ϕ_i 和连续变量 v_i 的非线性函数; β 为 $p \times 1$ 维固定效应参数; u_i 为 $q \times 1$ 维的随机效应参数, 假定服从期望为 0 方差为 Ψ 的正态分布; ϕ_{ij} 为形式参数, 它与 β 及 u_i 呈线性函数关系; A_{ij} 和 B_{ij} 分别为 β 和 u_i 的设计矩阵; ε_{ij} 为随机误差项, 假定对于任意 i ($i = 1, \dots, M$), $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})$ 都服从期望为 0 方差为 R 的正态分布, 同时还假定随机效应参数 u_i 与误差项 ε_{ij} 之间相互独立. 类似可以推广到逐级嵌套的更多水平 (两水平、三水平等), 例如, 两水平

$$\begin{cases} y_{ijk} = f(\phi_{ijk}, v_{ijk}) + \varepsilon_{ijk}, & i = 1, \dots, M, \quad j = 1, \dots, M_i, \quad k = 1, \dots, n_{ij}, \\ \phi_{ijk} = A_{ijk}\beta + B_{ijk}u_i + B_{ijk}u_{ij}, & u_i \sim N(0, \Psi_1), \quad u_{ij} \sim N(0, \Psi_2). \end{cases} \quad (1.2)$$

到目前为止, 已有多种算法求解单水平和嵌套多水平 NLMEM^[3,9,10]. 典型的计算软件有 SAS (Nlmmix 宏和 Nlmixed 模块)^[11,12] 和 S-Plus (nlme 函数)^[3].

近几年, 由于 NLMEM 应用不断完善及深化, 单水平和嵌套多水平 NLMEM 随之已不能满足实际的需要. 例如, 在林业上分析树高生长随年龄的变化关系时, 如果考虑这种变化受林分密度和立地质量的随机干扰, 由于林分密度和立地质量不是相互嵌套而是相互交错关系, 故利用 Pinheiro 和 Bates^[3] 定义的嵌套多水平 NLMEM 无法分析此类问题. 如果更进一步考虑林分密度和立地质量的随机干扰的方差受海拔高度影响 (海拔高度叫作组变量) 等, 则这一系列问题已有的模型和算法无法解决. 因此, 如何求解基于随机效应因子间相互交错的 NLMEM 是当前混合模型领域急需解决的难题. 为直观起见, 本研究用一个例子来说明.

例 1.1^[8] 假设树高 H 与树木直径 D 的关系是

$$H = 1.3 + \exp\left(\phi_1 + \frac{\phi_2}{D}\right) + \varepsilon, \quad (1.3)$$

其中 ϕ_1 和 ϕ_2 为形式参数, ε 为误差项.

实验数据来自吉林省汪清林业局, 共设置 5 个区组 (Block), 每个区组包含 4 个样地 (Plot), 分别测定每个样地的树高 (H) 和直径 (D) 共 1,945 次. 环境因子测定海拔 (A)、林分密度 (SD) 和地位级指数 (SI) 作为分类变量. 海拔按照大于等于 700 m 和小于 700 m 分为两个等级 (分别记为 1 和 2). 林分密度 (SD) 按区间 100 株/ha 划分为 6 个等级 (分别记为 1, 2, ..., 6). 地位级指数 (SI) 从 15 m 开始, 按间隔 0.5 m 划分为 5 个等级 (分别记为 1, 2, ..., 5).

表 1 为落叶松样地概况一览表, 如果只考虑单个因素海拔、区组、样地、林分密度和地位级指数对树高生长的影响 (单水平 NLMEM), 或者考虑区组以及嵌套在区组中的样地对树高生长的影响 (Block + Block \times Plot, 嵌套两水平 NLMEM), 或者考虑海拔、嵌套在海拔中的区组以及嵌套在区组中的样地对树高生长的影响 ($A + A \times$ Block + $A \times$ Block \times Plot, 嵌套三水平 NLMEM), 则可以利用现有的单水平和嵌套多水平 NLMEM 进行分析. 如果要考虑主效应林分密度和地位级指数对树高生长影响, 或同时考虑主效应林分密度和地位级指数以及它们之间交互效应对树高生长影响, 或进一步考

表 1 落叶松样地概况一览表

海拔	区组	样地	林分密度	地位级指数	观测数	海拔	区组	样地	林分密度	地位级指数	观测数
1	1	301	3	2	47	2	3	311	2	3	124
1	1	302	6	1	51	2	3	312	3	4	106
1	1	303	4	2	57	2	4	313	1	4	100
1	1	304	3	3	51	2	4	314	2	1	104
1	2	305	4	1	129	2	4	315	6	1	98
1	2	306	6	3	205	2	4	316	4	1	72
1	2	307	4	1	193	2	5	317	5	2	70
1	2	308	1	5	100	2	5	318	2	3	67
2	3	309	5	5	119	2	5	319	2	4	75
2	3	310	2	3	94	2	5	320	6	3	83

考虑交互作用的方差受海拔的影响等, 由于随机效应因子之间相互交错以及考虑交互作用, 现有的模型 (模型 (1.1) 和 (1.2)) 不能分析.

为此, 本研究主要目的是提出一个比 Pinheiro 和 Bates^[3] 所定义的 NLMEM 更为一般的模型表达式, 它应该是包括单水平、嵌套多水平、独立多因素 (不考虑交互作用且因素间相互交错) 和多因素的交互作用 (考虑交互作用且因素间相互交错) 的 NLMEM, 同时给出 10 种常见随机效应 (或误差项) 方差类型满足正定或半正定的条件以及模型参数估计方法. 该模型能推广到含组变量的 NLMEM.

2 非线性混合效应模型统一标准形式

在样本的 N 个观测值中, 研究的问题涉及 m 个分类变量, 分别是 V_1, \dots, V_m , 每个分类变量的等级个数记为 c_1, \dots, c_m , 对分类变量取值完全相同的观测点归为一类, 称为一个对象. 全部对象用集合 Θ 表示, Θ 中总对象个数为 M . 设第 i 个对象对应的分类变量的值为向量 ξ_i ($\xi_i \in \Theta$), 例如, $\xi_i = (v_1, \dots, v_m)$. 设第 i 个对象所含的观测点数记为 n_i , 每个观测点的下标是 i 和 j ($j = 1, \dots, n_i$). 用 Ω 记产生随机效应的变量 (随机构造变量) 集合, 它的元素是变量 (产生主效应) 或变量乘积 (产生交互作用或相嵌效应), 例如, $\Omega = (V_1, V_2, V_1 \times V_2)$. 设某个随机构造变量为 E ($E \in \Omega$), 它的等级数等于 E 所涉及的分变量等级个数的乘积, 记为 M_E , $u_i^{(E)}$ 表示构造变量 E 在第 i 个对象上附加的随机效应. 非线性混合效应模型的标准形式是

$$\begin{cases} y_{ij} = f(\phi_i, v_{ij}) + \varepsilon_{ij}, & i = 1, \dots, M, \quad j = 1, \dots, n_i, \\ \phi_i = A_i \beta + \sum_{E \in \Omega} B_i^{(E)} u_i^{(E)}, \\ E(u_i^{(E)}) = 0, \quad \text{var}(u_i^{(E)}) = \Psi^{(E)}, \quad \text{cov}(u_i^{(E)}, u_i^{(F)}) = 0, \quad F \in \Omega, \quad E \neq F, \\ \varepsilon_i \sim N(0, R_i), \quad \varepsilon = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T, \end{cases} \quad (2.1)$$

其中 y_{ij} 和 v_{ij} 分别为第 i 个对象第 j 次观测的 1 维因变量值和 s 维自变量值; ϕ_i 为 p 维形式参数向量; β 为 $p_0 \times 1$ 维固定效应参数向量; $u_i^{(E)}$ 为构造变量 E 在第 i 个对象上产生的 $q^{(E)} \times 1$ 维随机效应参数向量; $\Psi^{(E)}$ 为 $u_i^{(E)}$ 的协方差矩阵; 对于不同的 $E \in \Omega$, $u_i^{(E)}$ 相互独立; A_i 和 $B_i^{(E)}$ 分别为 β

和 $u_i^{(E)}$ 的设计矩阵; ε_{ij} 为误差项, 对于不同的 j , ε_{ij} 相互独立, 同时还假定 ε_{ij} 与所有 $u_i^{(E)}$ 相互独立. 当各种随机效应参数 $u_i^{(E)}$ ($E \in \Omega$) 和 ε_{ij} 都服从正态分布时, 模型 (2.1) 称为正态 NLMEM.

2.1 形式参数随机效应构造的类型

按照随机效应因素个数和因素间的逻辑关系 (独立、嵌套或交错), 把模型 (2.1) 中的形式参数 $\phi_i = (\phi_{i1}, \dots, \phi_{ip})$ 某个分量 ϕ_{it} ($t = 1, \dots, p$) 的构造分为以下 4 种类型. 直观起见, 用例 1.1 来说明.

类型 1 单因素情形.

分量 ϕ_{it} 受一个随机变量 (如林分密度) 的影响, 因此只有一个构造变量. 例如, $\phi_{it} = \beta_t + u_{it}^{(\text{SD})}$, 其中 $u_{it}^{(\text{SD})}$ 是由林分密度 (SD) 作用在形式参数 ϕ_{it} 上的随机效应, $\Omega = \{\text{SD}\}$.

模型中如果所有混合形式参数 (含有随机效应的形式参数) 的随机效应构造都为单因素情形, 如 SD (等级数为 M), 并且对应的构造变量也相同, 集合 Θ 的元素是变量 SD 取值, 构造变量集合 $\Omega = \{\text{SD}\}$ 只含一个元素, 根据 (2.1), 模型写为

$$\begin{cases} y_{ij} = f(\phi_i, v_{ij}) + \varepsilon_{ij}, & i = 1, \dots, M, \quad j = 1, \dots, n_i, \\ \phi_i = A_i\beta + B_i^{(\text{SD})}u_i^{(\text{SD})}, & u_i \sim N(0, \Psi^{(\text{SD})}), \end{cases} \quad (2.2)$$

这就是 Pinheiro 和 Bates^[3] 定义的单水平 NLMEM (模型 (1.1)).

类型 2 嵌套多因素情形.

ϕ_{it} 受逐级嵌套的随机效应因子影响, 例如, 以嵌套两水平为例, 随机效应因子为样地 (Plot) 和区组 (Block), 其中样地嵌套在区组内, 相应的形式参数 ϕ_{it} 的构造为

$$\phi_{it} = \beta_t + u_{it}^{(\text{Block})} + u_{it}^{(\text{Block} \times \text{Plot})},$$

其中 $\Omega = \{\text{Block}, \text{Block} \times \text{Plot}\}$. 该类型要求被嵌套的因子不能单独出现为构造变量, 例如, $u_{it}^{(\text{Plot})}$ 不作为随机效应, 因为没有区组的样地在本例中没有意义.

如果所有混合形式参数的构造形式都为嵌套多因素情形, 并且对应的构造变量也相同, 则由这些形式参数所构成的模型称为嵌套多因素 NLMEM^[3]. 以嵌套 2 因素 Block 和 Plot 为例, 假设 Block 中含有 M 个等级, Block 的第 a 个等级中对应 Plot 的等级数为 M_a ($a = 1, \dots, M$), 对象集合 Θ 的元素个数为 $\sum_{a=1}^M M_a$, 取构造变量集合 $\Omega = \{\text{Block}, \text{Block} \times \text{Plot}\}$, 根据 (2.1), 模型写为

$$\begin{cases} y_{ij} = f(\phi_i, v_{ij}) + \varepsilon_{ij}, & i = 1, \dots, \sum_{a=1}^M M_a, \quad j = 1, \dots, n_i, \\ \phi_i = A_i\beta + B_i^{(\text{Block})}u_i^{(\text{Block})} + B_i^{(\text{Block} \times \text{Plot})}u_i^{(\text{Block} \times \text{Plot})}, \\ u_i^{(1)} \sim N(0, \Psi^{(\text{Block})}), \quad u_i^{(\text{Block} \times \text{Plot})} \sim N(0, \Psi^{(\text{Block} \times \text{Plot})}). \end{cases} \quad (2.3)$$

在模型 (2.3) 中, 集合 Θ 的元素是变量 Block 和 Plot 取值, 由于 Plot 嵌套在 Block 中, $\text{Block} \times \text{Plot}$ 的每一个等级就是一个对象. 因此, 模型 (2.3) 与 (1.2) 完全相同.

对于多个变量逐级嵌套, 类似 2 个变量嵌套情形, 与 Pinheiro 和 Bates^[3] 定义的嵌套多水平模型完全相同. 但本研究给出的模型表达式较为简单 (模型中始终只含有两个下标), 避免了 Pinheiro 和 Bates^[3] 定义的嵌套多水平 NLMEM 下标繁多的缺点.

类型 3 含交互作用多因素情形.

ϕ_{it} 受多个因素 (因素间相互交错) 及其相互作用的影响. 以 2 个因素 SD 和 SI 变量为例, 形式参数 ϕ_{it} 的构造为 $\phi_{it} = \beta_t + u_{it}^{(SD)} + u_{it}^{(SI)} + u_{it}^{(SD \times SI)}$, 其中 $SD \times SI$ 表示林分密度与地位级指数的交互, $\Omega = \{SD, SI, SD \times SI\}$. 对于 3 因素以上, 可以考虑主因素部分或全部 2 因素交互, 以及 3 因素或多因素交互, 或其中部分项之和.

类型 4 组合情形.

ϕ_{it} 可能受多个随机效应因子的影响, 都可以分解成上述 3 种情况的组合. 例如, 形式参数 ϕ_{it} 考虑由林分密度 (SD) 和地位级指数 (SI) 产生的交互效应以及由海拔高度 (A) 和林分密度 (SD) 产生的嵌套效应, 相应的构造写为 $\phi_{it} = \beta_t + u_{it}^{(SD)} + u_{it}^{(SI)} + u_{it}^{(SD \times SI)} + u_{it}^{(A)} + u_{it}^{(A \times SD)}$, 其中 $\Omega = \{SD, SI, SD \times SI, A, A \times SD\}$.

模型 (2.1) 中的每一个形式参数都可以为上面 4 种情形中的一种, 因此, 模型 (2.1) 几乎包含所有随机效应类型组合的 NLMEM.

2.2 形式参数固定效应构造的类型

形式参数中固定效应 β 每一个分量 (β_i) 的构造有 3 种类型.

类型 1 含有固定参数, 其取值在所有观测点上相同.

不受其他因素影响, 固定效应参数在所有对象上取相同值. 例如, 第 2.1 小节中的类型 1 到类型 4.

类型 2 不含固定效应参数.

形式参数 ϕ_{it} 不含有固定效应参数, 仅有随机效应, 例如,

$$\phi_{it} = u_{it}^{(SD)} + u_{it}^{(SI)} + u_{it}^{(SD \times SI)} + u_{it}^{(A)} + u_{it}^{(A \times SD)}.$$

因此, 模型中形式参数的个数可能多于固定效应参数的个数 ($p \geq p_0$).

类型 3 含有固定参数, 其取值依赖于其他分类变量 (固定效应参数分级).

在某些情况下, 固定效应参数 β_t ($t = 1, \dots, p_0$) 的取值可能受到某些分类变量的影响. 例如, β_1 的取值与观测点分类变量 A 有关时, 对 β_1 按变量 A 的取值进行分级, 这时分类变量 A 相当于多元回归中的哑变量. 分类变量 A 称为 β_t 的分级变量.

下面以 (1.3) 为例进一步说明考虑分级时模型中固定效应参数的构造.

假如形式参数 ϕ_1 受随机因素 SD 的影响, 并且 ϕ_1 中的固定效应参数 β_1 与分类变量 A 的取值有关, ϕ_2 受随机因素 SD 和 SI 的直接影响, 这样它们所产生的随机效应和分布是 $u^{(SD)} \sim N(0, \Psi^{(SD)})$, $u^{(SI)} \sim N(0, \Psi^{(SI)})$, 因为随机因素 SD 同时作用在形式参数 ϕ_1 和 ϕ_2 上, 则 $u^{(SD)}$ 为 2 维随机效应参数, $u^{(SD)} = (u_1^{(SD)}, u_2^{(SD)})$, $\Psi^{(SD)}$ 为它的方差, 大小为 2×2 维. 而随机因素 SI 只作用在形式参数 ϕ_2 中, 则 $u^{(SI)}$ 为 1 维随机效应参数, $\Psi^{(SI)}$ 为它的方差, 大小为 1×1 维.

分类变量 A 有两个取值 a_1 和 a_2 , β_1 在 A 不同取值的观测点上分别取值 β_{1,a_1} 和 β_{1,a_2} , 因此, β 为 3 维向量. 基础模型 (1.3) 的 NLMEM 为

$$\begin{cases} H_{(A,SD,SI)j} = 1.3 + \exp \left[(\beta_{1,A} + u_1^{(SD)}) + \frac{\beta_{2,A} + u_2^{(SD)} + u_2^{(SI)}}{D_{(A,SD,SI)j}} \right] + \varepsilon_{(A,SD,SI)j}, \\ \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}_{(A,SD,SI)} = I_2 \begin{pmatrix} \beta_{1,A} \\ \beta_2 \end{pmatrix} + I_2 \begin{pmatrix} u_1^{(SD)} \\ u_2^{(SD)} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} (u^{(SI)}), \end{cases}$$

其中 $\beta_{1,A} = (\beta_{1,a_1}^{(A)}, \beta_{1,a_2}^{(A)})$, 把 β_1 的取值分为两个级别. 这相当于回归中的数量化.

2.3 随机效应分级

与固定效应参数分级类似, 随机效应参数 u_i 同样可能受到某些分类变量的影响, 即随机效应的分级, 可以理解为随机效应参数的数量化.

假定形式参数 ϕ_1 直接受随机因素 SD 的影响, 但是对于不同观测点, 形式参数 ϕ_1 受随机因素 SD 的影响与该观测点的变量 A 取值有关, 即随机因素 SD 对 ϕ_1 的影响需要按 A 分级. 这时称 A 是形式参数 ϕ_1 的随机效应分级变量.

下面以 (1.3) 为例进一步说明考虑随机效应参数分级时模型的参数构造.

假如形式参数 ϕ_1 受随机因素 SD 的影响, 其影响与 A 的取值有关. ϕ_2 仍然受随机因素 SD 和 SI 的直接影响, 这样它们所产生的随机变量和分布是 $u^{(SD)} \sim N(0, \Psi^{(SD)})$, $u^{(SI)} \sim N(0, \Psi^{(SI)})$. 随机因素 SD 产生的 $u^{(SD)}$ 为 3 维随机效应参数向量, $u^{(SD)} = (u_{1,a_1}^{(SD)}, u_{1,a_2}^{(SD)}, u_2^{(SD)})$. 随机因素 SI 产生的 $u^{(SI)}$ 为 1 维随机效应参数向量. 基础模型 (1.3) 的 NLMEM 可以写为

$$\begin{cases} H_{(A,SD,SI)j} = 1.3 + \exp \left[(\beta_1 + u_{1,A}^{(SD)}) + \frac{\beta_{2,A} + u_2^{(SD)} + u_2^{(SI)}}{D_{(A,SD,SI)j}} \right] + \varepsilon_{(A,SD,SI)j}, \\ \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}_{(A,SD,SI)} = I_2 \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + I_2 \begin{pmatrix} u_{1,A}^{(SD)} \\ u_2^{(SD)} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} (u^{(SI)}), \end{cases}$$

其中把 ϕ_1 的随机效应分成了 2 个级别. 模型中其他随机效应参数与 $u^{(SD)}$ 一样, 都可以考虑分级变量. 值得注意的是, 分级变量只能为分类变量或若干个分类变量的乘积.

3 参数估计

本文提出线性逼近 - 逐步 2 次规划算法求解正态 NLMEM 统一标准形式 (模型 (2.1)) 中参数. 首先通过一阶 Taylor 展开^[13] 把模型 (2.1) 转化为线性混合效应模型 (linear mixed effect model, LMEM). 然后利用混合模型正规方程法^[14] 求解固定效应参数和随机效应参数. 方差和协方差参数由逐步 2 次规划算法计算得到.

3.1 符号和矩阵表达式

在模型 (2.1) 中, 随机构造变量集合 Ω 中的元素可以编号, 设 Ω 中共有 K 个元素, 每个 $k = 1, \dots, K$ 表示一个随机构造变量, 等级数为 M_k . 为书写方便, 把集合 Ω 中元素以对应的编号表示. 因此, 正态 NLMEM 的标准形式可写为

$$\begin{cases} y_{ij} = f(\phi_i, v_{ij}) + \varepsilon_{ij}, & i = 1, \dots, M, \quad j = 1, \dots, n_i, \\ \phi_i = A_i \beta + \sum_{k=1}^K B_i^{(k)} u_i^{(k)}, \\ u_i^{(k)} \sim N(0, \Psi^{(k)}), \quad \text{cov}(u_i^{(k)}, u_i^{(l)}) = 0, \quad k \neq l, \quad k = 1, \dots, K, \quad l = 1, \dots, K. \end{cases} \quad (3.1)$$

对于 $i = 1, \dots, M$, 令

$$y_i = [y_{i1}, \dots, y_{in_i}]^T, \quad v_i = [v_{i1}^T, \dots, v_{in_i}^T]^T,$$

$$\begin{aligned}\tilde{A}_i &= [A_i, \dots, A_i]_{n_i p \times p_0}, \quad \tilde{B}_i^{(k)} = [(B_i^{(k)})^T, \dots, (B_i^{(k)})^T]_{n_i p \times q^{(k)}}, \\ \tilde{\phi}_i &= [\phi_i^T, \dots, \phi_i^T]_{n_i p \times 1}, \quad f(\tilde{\phi}_i, v_i) = [f(\phi_i, v_{i1}), \dots, f(\phi_i, v_{in_i})]^T,\end{aligned}$$

则模型 (3.1) 可写成矩阵形式:

$$\begin{cases} y_i = f(\tilde{\phi}_i, v_i) + \varepsilon_i, & i = 1, \dots, M, \\ \tilde{\phi}_i = \tilde{A}_i \beta + \sum_{k=1}^K \tilde{B}_i^{(k)} u_i^{(k)}, \\ u_i^{(k)} \sim N(0, \Psi^{(k)}), \quad \text{cov}(u^{(k)}, u^{(l)}) = 0, \quad k \neq l, \quad k = 1, \dots, K, \quad l = 1, \dots, K. \end{cases} \quad (3.2)$$

模型 (3.2) 中其他变量及参数与模型 (3.1) 完全相同.

3.2 线性化步

通过一阶条件期望线性化方法 (first order conditional expansion, FOCE) [13] 把模型 (3.2) 转化为

$$\hat{w}_i \approx \hat{X}_i \beta + \sum_{k=1}^K \hat{Z}_i^{(k)} u_i^{(k)} + \varepsilon_i, \quad i = 1, \dots, M, \quad (3.3)$$

其中 \hat{w}_i 为伪数据向量, 由

$$\hat{w}_i = y_i - f(\hat{\phi}_i, v_i) + \hat{X}_i \hat{\beta} + \sum_{k=1}^K \hat{Z}_i^{(k)} \hat{u}_i^{(k)} \quad (3.4)$$

计算得到, 这里 $\hat{\beta}$ 和 $\hat{u}_i^{(k)}$ ($k = 1, \dots, K$) 分别为 β 和 u_i 的当前估值; $\hat{\phi}_i$ 是形式参数向量的估计量, 表达式为

$$\tilde{\phi}_i = \tilde{A}_i \hat{\beta} + \sum_{k=1}^K \tilde{B}_i^{(k)} \hat{u}_i^{(k)}; \quad (3.5)$$

\hat{X}_i 和 $\hat{Z}_i^{(k)}$ 分别为 $\hat{\beta}$ 和 $\hat{u}_i^{(k)}$ ($k = 1, \dots, K$) 的设计矩阵, 计算公式分别为

$$\hat{X}_i = \left. \frac{\partial f(\hat{\phi}_i, v_i)}{\partial \beta^T} \right|_{\hat{\beta}, \hat{u}_i^{(1)}, \dots, \hat{u}_i^{(K)}, v_i}, \quad (3.6)$$

$$\hat{Z}_i^{(k)} = \left. \frac{\partial f(\hat{\phi}_i, v_i)}{\partial [u_i^{(k)}]^T} \right|_{\hat{\beta}, \hat{u}_i^{(1)}, \dots, \hat{u}_i^{(K)}, v_i}, \quad k = 1, \dots, K. \quad (3.7)$$

令

$$\begin{aligned}\hat{w} &= (\hat{w}_1^T, \dots, \hat{w}_M^T)^T, \quad \hat{X} = (\hat{X}_1^T, \dots, \hat{X}_M^T)^T, \quad \hat{Z} = (\hat{Z}^{(1)}, \dots, \hat{Z}^{(K)}), \\ u &= [(u^{(1)})^T, \dots, (u^{(K)})^T]^T, \quad \varepsilon = [\varepsilon_1^T, \dots, \varepsilon_M^T]^T, \quad \Psi = \text{diag}(\Psi^{(1)}, \dots, \Psi^{(K)}),\end{aligned}$$

其中

$$\begin{aligned}\hat{Z}^{(K)} &= \text{diag}(\hat{Z}_1^{(K)}, \dots, \hat{Z}_{M_K}^{(K)}), \quad u^{(K)} = [(\hat{u}_1^{(K)})^T, \dots, (\hat{u}_{M_K}^{(K)})^T]^T, \\ \tilde{\Psi}^{(K)} &= \text{diag}(\Psi^{(K)}, \dots, \Psi^{(K)})_{(M_K q^{(K)}) \times (M_K q^{(K)})},\end{aligned}$$

$t = 1, \dots, M_K$, $\hat{Z}_t^{(K)}$ 为第 K 个构造变量第 t 个等级产生的随机效应 $\hat{u}_t^{(K)}$ 所对应的设计矩阵, 为

$$\hat{Z}_t^{(K)} = [(\hat{Z}_{(M \times (t-1)/M_K + 1)}^{(K)})^T, \dots, (\hat{Z}_{(M \times t/M_K)}^{(K)})^T]^T.$$

由以上定义可知, 模型 (3.3) 写为

$$\hat{w} \approx \hat{X}\beta + \hat{Z}u + \varepsilon, \quad (3.8)$$

模型 (3.8) 是线性混合效应模型, 因此在当前点可得到模型 (3.1) 参数估计的线性近似值. 以下几小节将详细介绍参数的计算方法.

3.3 固定效应和随机效应参数估计

当 Ψ 和 R 已知时, 由混合模型正规方程^[14] 得到 β 和 u 的估计值 $\hat{\beta}$ 和 \hat{u} 分别为

$$\hat{\beta} = (\hat{X}^T \Sigma^{-1} \hat{X})^{-1} \hat{X}^T \Sigma^{-1} \hat{w}, \quad (3.9)$$

$$\hat{u} = \Psi \hat{Z}^T \Sigma^{-1} (\hat{w} - \hat{X} \hat{\beta}), \quad (3.10)$$

其中 Σ 为 \hat{w} 的方差, 由下式计算得到:

$$\Sigma = \text{cov}(\hat{Z}u + \varepsilon) = \hat{Z}\Psi\hat{Z}^T + R, \quad (3.11)$$

因此需估计方差和协方差矩阵 Ψ 和 R .

3.4 方差和协方差参数估计

文献 [15] 给出了模型 (3.8) 参数估计的具体算法, 为叙述方便, 我们摘录主要公式如下: 由正态性, 直接得到 (3.8) 的负二倍对数极大似然函数为

$$-2\ell_F(\sigma^2, \Psi | y) = \log |\Sigma| + (\hat{w} - \hat{X}\beta)^T \Sigma^{-1} (\hat{w} - \hat{X}\beta) + N \log(2\pi), \quad (3.12)$$

负二倍对数限制极大似然函数为

$$-2\ell_R(\sigma^2, \Psi | y) = \log |\hat{X}^T \Sigma^{-1} \hat{X}| - 2\ell_F(\Psi, R, \beta) - p_0 \log(2\pi). \quad (3.13)$$

假定 Ψ 和 R 由 r 个参数 $\theta_1, \dots, \theta_r$ 所决定, 设 $\theta = (\theta_1, \dots, \theta_r)^T$, 估计参数 θ 的问题转化为, 在限制 Ψ 为半正定矩阵和 R 为正定矩阵条件下, 寻找参数 θ , 使得 $L_F(\theta) = -2\ell_F(\theta)$ 或 $L_R(\theta) = -2\ell_R(\theta)$ 达到极小, 即非线性规划问题 (3.14) 和 (3.15).

对于极大似然 (maximum likelihood, ML) 参数估计:

$$\begin{aligned} \min L_F(\theta) &= \log |\Sigma| + (\hat{w} - \hat{X}\beta)^T \Sigma^{-1} (\hat{w} - \hat{X}\beta) + N \log(2\pi) \\ \text{s.t. } \Psi(\theta) &\text{ 为半正定矩阵, } R(\theta) \text{ 为正定矩阵.} \end{aligned} \quad (3.14)$$

对于限制极大似然 (restricted maximum likelihood, REML) 参数估计:

$$\begin{aligned} \min L_R(\theta) &= \log |\hat{X}^T \Sigma^{-1} \hat{X}| - 2\ell_F(\Psi, R, \beta) - p_0 \log(2\pi) \\ \text{s.t. } \Psi(\theta) &\text{ 为半正定矩阵, } R(\theta) \text{ 为正定矩阵.} \end{aligned} \quad (3.15)$$

文献 [3,9,16] 都假定 Ψ 为无结构方差类型, 并通过 Ψ 自身的 Cholesky 因子 L (为上三角矩阵) 构造协方差矩阵 $\Psi = L^T L$, 从而保证 Ψ 在计算过程中为半正定或正定. 该方法可以推广到 $BD(k)$, 即存在上三角矩阵 $L(k)$ 使得 $BD(k) = L(k)^T L(k)$. 但是也存在其他方差结构类型, 如 $LP(k)$, 很难用相同个数的参数得到半矩阵 $L(k)$ 的表达式, 因此, 本文对可以用线性 (或非线性) 约束表示的方差矩阵类型选用带约束条件的逐步 2 次规划算法求解参数 $\hat{\theta}$, 使得 $L_F(\hat{\theta})$ 或 $L_R(\hat{\theta})$ 达到极小值. 该算法从理论上保证了任意类型结构的 Ψ 和 R 为半正定或正定矩阵. 下面将详细介绍逐步 2 次规划的计算方法和常用方差类型满足正定或半正定矩阵的约束条件等. 关于参数向量 θ 对 (3.14) 和 (3.15) 的一阶和二阶求导, 采用文献 [15] 中的计算公式.

3.4.1 逐步 2 次规划的计算方法

对于 ML 参数估计方法, 逐步 2 次规划算法的计算步骤如下.

在 (3.14) 中, 假设 $\Psi(\theta)$ 为正定或半正定矩阵的约束条件为 $g_\Psi(\theta) \leq 0$, $R(\theta)$ 为正定矩阵的约束条件为 $g_R(\theta) < 0$, 其中 $g_\Psi(\theta)$ 和 $g_R(\theta)$ 分别是参数向量 θ 的函数. 因此, 非线性规划问题 (3.14) 可写成下列形式:

$$\begin{aligned} \min L_F(\theta) &= \log |\Sigma| + (\hat{w} - \hat{X}\beta)^T \Sigma^{-1} (\hat{w} - \hat{X}\beta) + N \log(2\pi), \quad \theta \in \mathbb{R}^r \\ \text{s.t. } g_\Psi(\theta) &\leq 0, \quad g_R(\theta) < 0. \end{aligned} \quad (3.16)$$

记 θ 可行区域为 Λ (Λ 为凸集合), 在 $\theta \in \Lambda$ 处寻找增量 d , 使得 $L_F(\theta + d) < L_F(\theta)$, 且 $\theta + d \in \Lambda$. 对 $L_F(\theta)$ 在 $\hat{\theta}$ 处二阶 Taylor 展开, 得到

$$L_F(\theta) \cong L_F(\hat{\theta}) + \nabla L_F(\hat{\theta})d + \frac{1}{2}d^T \nabla^2 L_F(\hat{\theta})d. \quad (3.17)$$

除去常数 $L_F(\hat{\theta})$, 非线性规划问题 (3.16) 变成一个近似二次规划问题:

$$\begin{aligned} \min \varphi(d) &= \nabla L^{(F)}(\hat{\theta})d + \frac{1}{2}d^T H d \\ \text{s.t. } g_\Psi(\hat{\theta}) + \frac{\partial [g_\Psi(\theta)]}{\partial \theta} \Big|_{\theta=\hat{\theta}} d &\leq 0, \\ g_R(\hat{\theta}) + \frac{\partial [g_R(\theta)]}{\partial \theta} \Big|_{\theta=\hat{\theta}} d &\leq 0, \end{aligned} \quad (3.18)$$

其中

$$H = U \text{diag}(h_1, \dots, h_r) U^T,$$

$\nabla^2 L_F(\hat{\theta})$ 的谱分解为

$$\nabla^2 L_F(\hat{\theta}) = U \text{diag}(\lambda_1, \dots, \lambda_r) U^T.$$

令

$$h_i = \begin{cases} \lambda_i, & \text{若 } \lambda_i > 0, \\ \delta, & \text{其他情形,} \end{cases}$$

其中 δ 是一个适当的正实数. 由二次规划问题 (3.18) 得到最优解 \hat{d} , 把 \hat{d} 作为方向, 修改步长为 $\lambda \hat{d}$ ($0 \leq \lambda \leq 1$). 通过优选法做一维搜索:

$$\begin{aligned} \min L^{(F)}(\hat{\theta} + \lambda \hat{d}) \quad (1 \leq \lambda \leq 1) \\ \text{s.t. } \hat{\theta} + \lambda \hat{d} \in \Lambda, \end{aligned} \quad (3.19)$$

得到 $\hat{\lambda}$ 的最优值, 然后设 $\hat{\theta} = \hat{\theta} + \hat{\lambda}\hat{d}$, 重新开始寻找增量 d . 对于限制极大似然法, 把 (3.16) 中 $L_F(\theta)$ 替换成 $L_R(\theta)$, 其他计算与极大似然法完全一致.

3.4.2 线性逼近 - 逐步 2 次规划算法计算步骤

线性逼近 - 逐步 2 次规划算法计算模型 (3.2) 中参数的主程序步骤如下:

第 1 步 给固定效应参数 β 赋初始值 (通常取最小二乘估计量), 指定允许误差 e ($e = 10^{-10}$) 和一个很小的非负实数 δ .

第 2 步 设初始值 $u = 0$ 和 $\theta = (\theta_1, \dots, \theta_r)^T$, 使得 Ψ 为半正定矩阵及 $\sigma^2 > 0$. 本算法中参数 θ 初始值的给定方法为, 对于 Ψ , 每个参数初始值为 $Q/(2r_1)$; 对于 R , 每个参数初始值为 $Q/(2r_2)$, 其中 Q 是被分析模型不考虑随机效应参数时的残差平方和, r_1 和 r_2 分别为 Ψ 和 R 中参数个数.

第 3 步 由 (3.6) 和 (3.7) 分别计算 \hat{X}_i 和 $\hat{Z}_i^{(k)}$ ($i = 1, \dots, M$), 得到 X 和 Z .

第 4 步 由 (3.4) 计算 \hat{w}_i ($i = 1, \dots, M$), 得到 \hat{w} .

第 5 步 根据计算出的 \hat{w} 、 X 和 Z , 得到标准形式的线性混合效应模型 (3.8). 通过求解模型 (3.8) (见子程序 A), 得到 $\hat{\theta}$ 、 $\hat{\beta}$ 和 \hat{u} .

第 6 步 检查收敛准则. 利用 (3.12) (或 (3.13)), 计算 $L_F(\hat{\theta})$ (或 $L_R(\hat{\theta})$), 如果满足

$$\begin{aligned} |L_F(\theta) - L_F(\hat{\theta})| &\leq e \vee |L_F(\theta) - L_F(\hat{\theta})| \leq e |L_F(\hat{\theta})| \\ (\text{或 } |L_R(\theta) - L_R(\hat{\theta})| &\leq e \vee |L_R(\theta) - L_R(\hat{\theta})| \leq e |L_R(\hat{\theta})|), \end{aligned}$$

得到一个解 $\hat{\theta}$, 终止计算; 否则, 令 $\theta = \hat{\theta}$, $\beta = \hat{\beta}$, $u = \hat{u}$ 转入第 3 步.

子程序 A:

线性混合效应模型子程序 (改进文献 [15] 算法) 为

第 1 步 根据主程序得到的 X 、 Z 、 \hat{w} 、 Ψ 和 R , 利用 (3.11)、(3.9) 和 (3.10) 分别计算 $\hat{\Sigma}$ 、 $\hat{\beta}$ 和 \hat{u} .

第 2 步 计算似然函数的一阶导数 $\nabla L_F(\theta)$ ($\nabla L_R(\theta)$) 和二阶导数 $\nabla^2 L_F(\theta)$ ($\nabla^2 L_R(\theta)$), 计算公式参见文献 [15].

第 3 步 求解近似二次规划问题. 对于 ML 估计, 求解 (3.18), 如果是 REML 估计, 把 (3.18) 中的 $L_F(\theta)$ 替换成 $L_R(\theta)$.

第 4 步 做一维搜索, 极小化 (3.19) 得到修改步长 $\hat{\lambda}\hat{d}$, 计算 $\hat{\theta} = \theta + \hat{\lambda}\hat{d}$.

该算法已在 ForStat “非线性混合效应” 模块实现, 算法的详细介绍参见文献 [17].

4 含有组变量的 NLMEM

在模型 (2.1) 中, 如果随机变量 $u_i^{(E)}$ 的方差矩阵与某些分类因子 (categorical factor, CF) 有关, 则称 CF 为方差组变量 (简称组变量), 含组变量的模型表达式如下:

$$\begin{cases} y_{gij} = f(\phi_{gi}, v_{gij}) + \varepsilon_{gij}, & g = 1, \dots, G, \quad i = 1, \dots, M, \quad j = 1, \dots, n_{gi}, \\ \phi_{gi} = A_{gi}\beta + \sum_{E \in \Omega} B_{gi}^{(E)} u_{gi}^{(E)}, \\ \mathbb{E}(u_{gi}^{(E)}) = 0, \quad \text{var}(u_{gi}^{(E)}) = \Psi_g^{(E)}, \quad \text{cov}(u_{gi}^{(E)}, u_{gi}^{(F)}) = 0, \quad F \in \Omega, \quad E \neq F, \\ \varepsilon_{gi} \sim N(0, R_{gi}), \quad \varepsilon_{gi} = (\varepsilon_{gi1}, \dots, \varepsilon_{gin_{gi}})^T, \end{cases} \quad (4.1)$$

其中 G 为组变量的等级数; M 为观测对象的总个数; n_{gi} 为组变量第 g 个等级中的第 i 个对象重复观测次数; y_{gij} 和 v_{gij} 分别是组变量第 g 个等级中第 i 个对象第 j 次观测的 1 维因变量值和 s 维自变量值; ϕ_{gi} 是 p 维形式参数向量; β 为 $p_0 \times 1$ 维固定效应参数向量; $u_{gi}^{(E)}$ 是构造变量 E ($E \in \Omega$, Ω 为构造变量集合) 在组变量第 g 个等级中的第 i 个对象上产生的 $q^{(E)} \times 1$ 维随机效应参数向量; $\Psi_g^{(E)}$ 为 $u_{gi}^{(E)}$ 的正定或半正定协方差矩阵; 对于不同的 E ($E \in \Omega$), $u_{gi}^{(E)}$ 相互独立; A_{gi} 和 $B_{gi}^{(E)}$ 分别是 β 和 $u_{gi}^{(E)}$ 的设计矩阵; ε_{gi} 为误差项, 对于不同的 g 和 i , ε_{gi} 相互独立, 并且假定 ε_{gij} 与 $u_{gi}^{(k)}$ ($k = 1, \dots, K$) 相互独立.

显然, 模型 (4.1) 是模型 (2.1) 的一种推广; 当模型 (4.1) 中不考虑组变量, 即 $G = 1$ 时, 模型 (4.1) 与 (2.1) 完全等价.

同样, 把模型 (4.1) 中的随机构造变量集合 Ω 中元素以对应的编号表示. 设 Ω 中共有 K 个元素, 每个 $k = 1, \dots, K$ 表示一个随机构造变量, 它的等级数为 M_k . 并且假定模型 (4.1) 中各种随机效应参数 $u_{gi}^{(E)}$ ($E \in \Omega$) 都服从正态分布. 对于 $g = 1, \dots, G, i = 1, \dots, M$, 令

$$\begin{aligned} y_{gi} &= (y_{gi1}, \dots, y_{gin_i})^T, & v_{gi} &= (v_{gi1}^T, \dots, v_{gin_i}^T)^T, \\ \tilde{A}_{gi} &= (A_{gi}^T, \dots, A_{gi}^T)^T, & \tilde{B}_{gi}^{(k)} &= [(B_{gi}^{(k)})^T, \dots, (B_{gi}^{(k)})^T]^T, \\ \tilde{\phi}_{gi} &= (\phi_{gi}^T, \dots, \phi_{gi}^T)^T, & f(\tilde{\phi}_{gi}, v_{gi}) &= (f(\phi_{gi}, v_{gi1}), \dots, f(\phi_{gi}, v_{gin_i}))^T, & \varepsilon_{gi} &= (\varepsilon_{gi1}, \dots, \varepsilon_{gin_i})^T, \end{aligned}$$

则模型 (4.1) 可写成矩阵形式如下:

$$\begin{cases} y_{gi} = f(\tilde{\phi}_{gi}, v_{gi}) + \varepsilon_{gi}, & g = 1, \dots, G, \quad i = 1, \dots, M, \\ \tilde{\phi}_{gi} = \tilde{A}_{gi}\beta + \sum_{k=1}^K \tilde{B}_{gi}^{(k)} u_{gi}^{(k)}, \\ u_{gi}^{(k)} \sim N(0, \Psi_g^{(k)}), \quad \text{cov}(u_{gi}^{(k)}, u_{gi}^{(l)}) = 0, \quad k \neq l, \quad k = 1, \dots, K, \quad l = 1, \dots, K. \end{cases} \quad (4.2)$$

对于 $g = 1, \dots, G$, 令

$$\begin{aligned} y_g &= (y_{g1}, \dots, y_{gM})^T, & v_g &= (v_{g1}, \dots, v_{gM}), & A_g &= (\tilde{A}_{g1}^T, \dots, \tilde{A}_{gM}^T)^T, & u_g^{(k)} &= (\tilde{u}_{g1}^{(k)}, \dots, \tilde{u}_{gM}^{(k)})^T, \\ B_g^{(k)} &= \text{diag}(\tilde{B}_{g1}^{(k)}, \dots, \tilde{B}_{gM}^{(k)}), & \tilde{\Psi}_g^{(k)} &= \text{diag}(\Psi_g^{(k)}, \dots, \Psi_g^{(k)}), & \phi_g &= (\tilde{\phi}_{g1}^T, \dots, \tilde{\phi}_{gM}^T)^T, \\ f(\phi_g, v_g) &= (f(\tilde{\phi}_{g1}, v_{g1})^T, \dots, f(\tilde{\phi}_{gM}, v_{gM})^T)^T, & \varepsilon_g &= (\varepsilon_{g1}^T, \dots, \varepsilon_{gM}^T)^T, \end{aligned}$$

$\tilde{B}_{gt}^{(k)}$ ($t = 1, \dots, M_k$) 为第 g 组中第 k 个构造变量第 t 个等级产生的随机效应 $\tilde{u}_{gt}^{(k)}$ 所对应的设计矩阵, 则模型 (4.2) 可写为

$$\begin{cases} y_g = f(\phi_g, v_g) + \varepsilon_g, & g = 1, \dots, G, \\ \phi_g = A_g\beta + \sum_{k=1}^K B_g^{(k)} u_g^{(k)}, \\ u_g^{(k)} \sim N(0, \tilde{\Psi}_g^{(k)}), \quad \text{cov}(u_g^{(k)}, u_g^{(l)}) = 0, \quad k \neq l, \quad k = 1, \dots, K, \quad l = 1, \dots, K. \end{cases} \quad (4.3)$$

模型 (4.3) 只含有一个分类变量, 即组变量. 与模型 (2.1) 类似, 可以通过线性逼近 - 逐步 2 次规划算法求解参数.

5 案例分析

本文选用中国林业科学研究院设置在东北汪清林业局 5 个区组共 20 块固定样地中的 510 株落叶松数据. 分别调查样地海拔高度、落叶松密度和地位级指数以及样地中每株树木直径 D 和树高 H 等. 数据结构见表 1. 由于模型 (2.1) 常被用来分析单木树高与直径之间关系 (参见文献 [8]), 故选它作为构建混合效应模型的基础模型. 在实验过程中调查了每株样木的海拔高度, 通过海拔高度大小把所有落叶松分为两个组, 大于等于 700 m 的为第一组, 小于 700 m 的为第二组. 本实例中, 考虑构造变量 (林分密度、地位级指数、林分密度和地位级指数的交互作用) 产生的随机效应对应的方差与海拔高度有关, 对应的构造变量 $\Omega = \{SD, SI, SD \times SI\}$, 组变量为海拔高度. 由于海拔高度有两个等级, 即 $G = 2$, 因此对于 $i = 1, \dots, M_1, j = 1, \dots, M_2, (ij)$ 为一个研究对象, 树高 - 直径模型可写为

$$\begin{cases} H_{g(ij)l} = 1.3 + (\beta_1 + u_{g(ij)}^{(M)} + u_{g(ij)}^{(S)} + u_{g(ij)}^{(M \times S)})[1 - \exp(-\beta_2 D_{g(ij)l})]^{\beta_3} + \varepsilon_{g(ij)l}, \\ u_{g(ij)}^{(M)} \sim N(0, \Psi_g^{(M)}), \quad u_{g(ij)}^{(S)} \sim N(0, \Psi_g^{(S)}), \quad u_{g(ij)}^{(M \times S)} \sim N(0, \Psi_g^{(M \times S)}), \quad \varepsilon_{g(ij)l} \sim N(0, \sigma^2), \end{cases} \quad (5.1)$$

其中 $g = 1, \dots, G; l = 1, \dots, n_{g(ij)}; n_{g(ij)}$ 为林分密度第 i 等级、地位级指数第 j 等级所调查的落叶松株数, 具体取值见表 2; $H_{g(ij)l}$ 和 $D_{g(ij)l}$ 分别为第 g 组中林分密度第 i 等级、地位级指数第 j 等级对应的第 l 株落叶松的树高和直径; $u_{g(ij)}^{(M)}$ 、 $u_{g(ij)}^{(S)}$ 和 $u_{g(ij)}^{(M \times S)}$ 分别为第 g 组中由林分密度、地位级指数以及它们之间交互作用产生的随机效应, 大小都为 1 维; β_1 、 β_2 和 β_3 为固定效应参数; 假定 $\Psi_g^{(M)}$ 、 $\Psi_g^{(S)}$ 和 $\Psi_g^{(M \times S)}$ 为 1×1 维的无结构类型.

选用 ML 和 REML 2 种参数估计方法求解模型 (5.1) 中参数, 允许误差取 $e = 10^{-10}$. 把固定效应参数的最小二乘估计量 $\beta_0 = (17.924, 0.1388, 1.6533)$ 作为算法的初始化值.

在利用线性逼近 - 逐步 2 次规划计算模型 (5.1) 时, 对于 ML 参数估计, 经过 7 次迭代计算达到收敛; 对于 REML 参数估计, 需要迭代 6 次. ML 和 REML 参数估计方法对应的方差和协方差参数、对数似然函数值和固定效应参数的迭代过程分别见图 1 和 2. 线性逼近 - 逐步 2 次规划算法在 2 种参数估计方法中通过第一次迭代就把每个参数的估计值落入了最优值附近, 后面几次的迭代步长非常小. 运算所需要的时间分别为 56 s 和 51 s (硬件条件: 2.4 GHz CPU, 2 GB 内存). 计算收敛时, 各参数的 ML 和 REML 估计量见表 3. 从表中知, 对于 ML 参数估计, 由林分密度产生的主随机效应参数为 0; 而对于 REML 参数估计, 在第一组中, 由林分密度产生的主随机效应参数为 0, 而在第二组中, 主随机效应参数接近为 0. 从而说明, 单纯的林分密度对树高影响较小, 而林分密度与地位级指数之间交互作用对树高的影响较大. 根据表 3 中的参数估计量, 可以利用模型 (5.1) 对落叶松树高进行预测.

表 2 不同组中林分密度和地位级指数不同等级数中落叶松株数

林分密度	地位级指数				
	1	2	3	4	5
1	0 0	0 0	0 0	0 30	30 0
2	0 30	0 0	0 30	0 30	0 0
3	0 0	30 0	30 0	0 30	0 0
4	30 30	30 0	0 0	0 0	0 0
5	0 0	0 30	0 0	0 0	0 30
6	30 30	0 0	30 30	0 0	0 0

注: “|” 左边表示第一组中落叶松株数, “|” 右边表示第二组中落叶松株数.

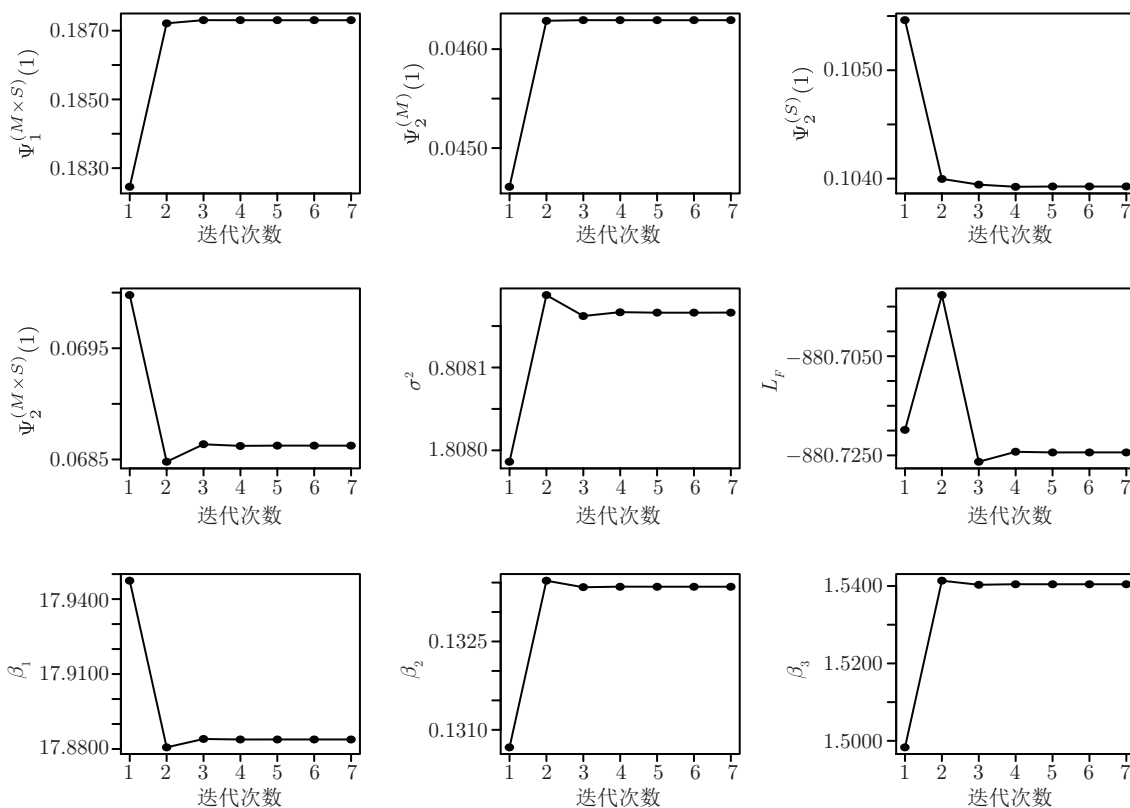


图 1 利用 ML 方法求解模型 (5.1) 时的参数迭代过程图

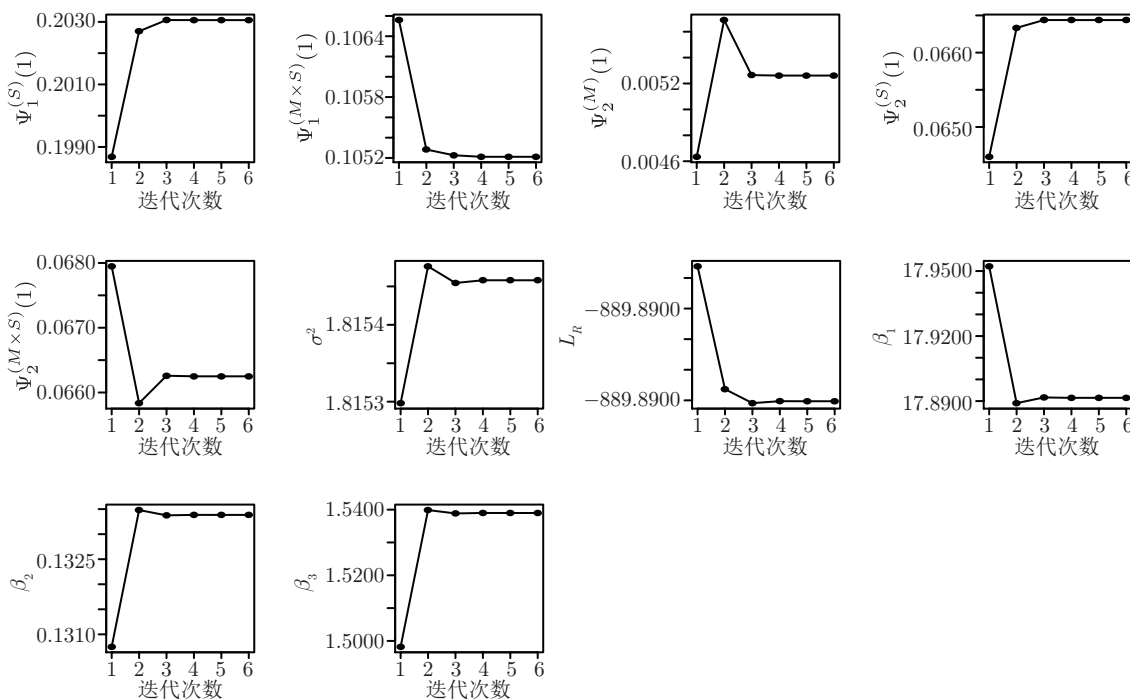


图 2 利用 REML 方法求解模型 (5.1) 时的参数迭代过程图

表 3 模型 (5.1) 中参数的 ML 和 REML 估计值

估计方法	参数	估计值	参数	估计值	参数	估计值	参数	估计值	
ML	β_1	17.8838	$u_{1(i1)}^{(S)}$	-0.2420	$u_{1(15)}^{(M \times S)}$	0.1034	$u_{2(24)}^{(M \times S)}$	0.0152	
	β_2	0.1334	$u_{1(i2)}^{(S)}$	0.3842	$u_{1(32)}^{(M \times S)}$	0.0802	$u_{2(34)}^{(M \times S)}$	-0.0002	
	β_3	1.5405	$u_{1(i3)}^{(S)}$	0.4763	$u_{1(33)}^{(M \times S)}$	-0.1524	$u_{2(41)}^{(M \times S)}$	0.0432	
	σ^2	1.8082	$u_{1(i4)}^{(S)}$	0.0000	$u_{1(41)}^{(M \times S)}$	0.0219	$u_{2(52)}^{(M \times S)}$	-0.2070	
	$\Psi_1^{(M)}(1)$	0.0000	$u_{1(i5)}^{(S)}$	-0.0615	$u_{1(42)}^{(M \times S)}$	0.1330	$u_{2(55)}^{(M \times S)}$	-0.0911	
	$\Psi_1^{(S)}(1)$	0.1873	$u_{2(i1)}^{(S)}$	-0.1219	$u_{1(61)}^{(M \times S)}$	-0.1562	$u_{2(61)}^{(M \times S)}$	-0.1892	
	$\Psi_1^{(M \times S)}(1)$	0.1039	$u_{2(i2)}^{(S)}$	-0.1396	$u_{1(63)}^{(M \times S)}$	0.4167	$u_{2(63)}^{(M \times S)}$	0.1396	
	$\Psi_2^{(M)}(1)$	0.0000	$u_{2(i3)}^{(S)}$	-0.0724	$u_{2(14)}^{(M \times S)}$	0.2761			
	$\Psi_2^{(S)}(1)$	0.0463	$u_{2(i4)}^{(S)}$	0.1964	$u_{2(21)}^{(M \times S)}$	-0.0346			
	$\Psi_2^{(M \times S)}(1)$	0.0686	$u_{2(i5)}^{(S)}$	-0.0615	$u_{2(23)}^{(M \times S)}$	-0.2469			
	REML	β_1	17.8915	$u_{2(2j)}^{(M)}$	-0.0194	$u_{2(i2)}^{(S)}$	-0.1803	$u_{2(14)}^{(M \times S)}$	0.2433
		β_2	0.1334	$u_{2(3j)}^{(M)}$	-0.0014	$u_{2(i3)}^{(S)}$	-0.0888	$u_{2(21)}^{(M \times S)}$	-0.0346
		β_3	1.5390	$u_{2(4j)}^{(M)}$	0.0038	$u_{2(i4)}^{(S)}$	0.2305	$u_{2(23)}^{(M \times S)}$	-0.2297
		σ^2	1.8154	$u_{2(5j)}^{(M)}$	-0.0204	$u_{2(i5)}^{(S)}$	-0.0769	$u_{2(24)}^{(M \times S)}$	0.0047
$\Psi_1^{(M)}(1)$		0.0000	$u_{2(ij)}^{(M)}$	-0.0028	$u_{1(15)}^{(M \times S)}$	0.1034	$u_{2(34)}^{(M \times S)}$	-0.0181	
$\Psi_1^{(S)}(1)$		0.2031	$u_{1(i1)}^{(S)}$	-0.2536	$u_{1(32)}^{(M \times S)}$	0.0802	$u_{2(41)}^{(M \times S)}$	0.0432	
$\Psi_1^{(M \times S)}(1)$		0.1052	$u_{1(i2)}^{(S)}$	0.3879	$u_{1(33)}^{(M \times S)}$	-0.1524	$u_{2(52)}^{(M \times S)}$	-0.1798	
$\Psi_2^{(M)}(1)$		0.0053	$u_{1(i3)}^{(S)}$	0.4822	$u_{1(41)}^{(M \times S)}$	0.0219	$u_{2(55)}^{(M \times S)}$	-0.0767	
$\Psi_2^{(S)}(1)$		0.0664	$u_{1(i4)}^{(S)}$	0.0000	$u_{1(42)}^{(M \times S)}$	0.1330	$u_{2(61)}^{(M \times S)}$	-0.1765	
$\Psi_2^{(M \times S)}(1)$		0.0662	$u_{1(i5)}^{(S)}$	0.1888	$u_{1(61)}^{(M \times S)}$	-0.1562	$u_{2(63)}^{(M \times S)}$	0.1411	
$u_{2(1j)}^{(M)}$		0.0193	$u_{2(i1)}^{(S)}$	-0.1480	$u_{1(63)}^{(M \times S)}$	0.4167			

注: $\Psi_g^{(k)}(1)$ 表示第 g ($g = 1, 2$) 组中构造变量 k 产生的随机效应对应的协方差参数. $u_{g(ij)}^{(M)}$ ($i = 1, \dots, M_1$) 表示第 g ($g = 1, 2$) 组中林分密度第 i 等级对应的随机效应. $u_{g(ij)}^{(S)}$ ($j = 1, \dots, M_2$) 表示第 g 组中地位级指数第 j 等级对应的随机效应. $u_{g(ij)}^{(M \times S)}$ 表示第 g 组中林分密度第 i 等级、地位级指数第 j 等级所产生的交互随机效应.

6 数值模拟验证

本节通过模拟研究评价线性逼近 - 逐步 2 次规划算法的可靠性. 选用模型 (5.1) 进行拟合研究. 在模拟研究中, 数据结构与第 5 节案例类似, 设 $G = 2$, $M_1 = 6$, $M_2 = 5$, $n_{g(ij)} = 20$ ($i = 1, \dots, M_1$, $j = 1, \dots, M_2$), 则 $N = 1200$. 用真实值

$$\beta = (18, 0.5, 2)^T, \quad \Psi_1^{(M)} = 2, \quad \Psi_2^{(M)} = 2.5, \quad \Psi_1^{(S)} = 4, \quad \Psi_2^{(S)} = 3, \quad \Psi_1^{(M \times S)} = 1, \quad \Psi_2^{(M \times S)} = 0.5$$

和 $\sigma^2 = 4$ 产生 1,000 组拟合数据 (根据给定的模型参数, 由随机效应参数服从正态分布产生, 具体方法参见文献 [18]). 选用 ML 和 REML 参数估计方法对 1,000 组拟合数据进行计算.

表 4 列出了固定效应参数、方差和协方差参数的简要统计量, 如平均值、偏差和均方误差等. 从表中可知, 2 种参数估计方法对应的参数估计量都非常接近真实值, 从而进一步说明线性逼近 - 逐步 2 次规划算法有较高的计算精度. 固定效应参数 β_1 、 β_2 和 β_3 和误差项参数 σ^2 对应的偏差比协方差参数要小, 并且在估计方差和协方差参数时, REML 对应的偏差比 ML 要小.

表 4 模型 (5.1) 中各参数估计的拟合结果

参数	ML			REML		
	平均值	偏差	均方误差	平均值	偏差	均方误差
β_1	17.9812	0.0188	4.7534	17.9859	0.0141	4.7583
β_2	0.4873	0.0127	0.1297	0.4913	0.0087	0.1308
β_3	1.9821	0.0179	0.7634	1.9867	0.0133	0.7721
σ^2	4.0032	-0.0032	1.4921	4.0014	-0.0014	1.4975
$\Psi_1^{(M)}(1)$	1.9153	0.0847	0.3452	1.9524	0.0476	0.3591
$\Psi_1^{(S)}(1)$	3.9281	0.0719	1.1635	3.95785	0.0422	1.1715
$\Psi_1^{(M \times S)}(1)$	0.9325	0.0675	0.2142	0.9475	0.0525	0.2238
$\Psi_2^{(M)}(1)$	2.5893	-0.0893	0.7836	2.5712	-0.0712	0.7914
$\Psi_2^{(S)}(1)$	2.9246	0.0754	0.8220	2.9511	0.0489	0.8732
$\Psi_2^{(M \times S)}(1)$	0.4684	0.0316	0.0847	0.4826	0.0174	0.0853

7 结论

本文给出了一种正态 NLMEM 的统一表达式, 除已有的单水平和逐级嵌套多水平 NLMEM 外, 还包含其他多种类型的 NLMEM, 例如, 只含主效应的多因素 NLMEM、包括主效应和交互效应的 NLMEM, 以及某几种类型组合的一般性 NLMEM 等. 提出了一种线性逼近 - 逐步 2 次规划算法求解正态 NLMEM 的标准形式, 即模型 (2.1). 选用 ML 和 REML 2 种参数估计方法, 给出线性逼近 - 逐步 2 次规划算法的具体计算公式, 并详细介绍用逐步 2 次规划算法求解模型中方差参数的计算方法. 推导出 10 种常用的协方差类型矩阵满足正定或半正定时的条件. 同时还列出线性逼近 - 逐步 2 次规划算法的具体计算步骤, 该算法已在 ForStat 2.2 版本“非线性混合效应”模块实现. 在对落叶松树高数据分析中发现线性逼近 - 逐步 2 次规划算法达到计算收敛时只需较少迭代次数, 并且计算速度也相对较快. 对 3 种不同模型通过数值模拟验证得出, 该算法具有较高的计算精度. 除此之外, 线性逼近 - 逐步 2 次规划算法理论上保证了随机效应参数或误差项方差为正定或半正定矩阵, 从而克服其他软件 (如 SAS 中 mixed 模块) 不能保证随机效应参数方差为非负定矩阵的缺陷.

本文在模型 (2.1) 的基础上, 进一步考虑随机效应方差与某些因素有关, 故考虑组变量的 NLMEM; 基于模型 (2.1) 给出了考虑组变量的 NLMEM 一般形式, 即模型 (4.1). 模型 (2.1) 是模型 (4.1) 的一种广义形式, 当不考虑组变量时, 两者完全等同. 同样可以利用线性逼近 - 逐步 2 次规划算法求解考虑组变量的正态 NLMEM 标准形式. 整个研究假定随机效应参数 $u_i^{(E)}$ ($E \in \Omega$) 和误差项 ε_i 服从正态分布, 对于服从其他分布 (如二项分布和 β 分布等) 的情形尚需进一步研究.

参考文献

- 1 Sheiner L B, Beal S L. Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. J Pharmacokinetics Biopharm, 1980, 8: 553-571
- 2 Bates D M, Watts D G. Nonlinear Regression Analysis and Its Applications. New York: John Wiley & Sons, 1988
- 3 Pinheiro J C, Bates D M. Mixed-Effects Models in S and S-Plus. New York: Springer-Verlag, 2000
- 4 Hall D B, Clutter M. Multivariate multilevel nonlinear mixed effects models for timber yield predictions. Biometrics, 2004, 60: 16-24
- 5 Calegario N, Daniels R F, Maestri R, et al. Modeling dominant height growth based on nonlinear mixed-effects model:

- A clonal *Eucalyptus* plantation case study. For *Ecol Manage*, 2005, 204: 11–21
- 6 Sharma M, Parton J. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. For *Ecol Manage*, 2007, 249: 187–198
 - 7 Panhard X, Samson A. Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. *Biostatistics*, 2009, 10: 121–135
 - 8 Yang Y, Huang S. Comparison of different methods for fitting nonlinear mixed forest models and for making predictions. *Can J For Res*, 2011, 41: 1671–1686
 - 9 Pinheiro J C, Bates D M. Approximations to the log-likelihood function in the nonlinear mixed effects model. *J Comput Graph Statist*, 1995, 4: 12–35
 - 10 Davidian M, Giltinan D M. *Nonlinear Models for Repeated Measurement Data*. New York: Chapman & Hall, 1995
 - 11 Littell R C, Milliken G A, Stroup W W, et al. *SAS System for Mixed Models*. Cary: SAS Institute, 1996
 - 12 Littell R C, Milliken G A, Stroup W W, et al. *SAS for Mixed Models*, 2nd ed. Cary: SAS Institute, 2006
 - 13 Lindstrom M J, Bates D M. Nonlinear mixed effects models for repeated measures data. *Biometrics*, 1990, 46: 673–687
 - 14 Henderson C R. *Applications of Linear Models in Animal Breeding*. Guelph: University of Guelph, 1984
 - 15 Wolfinger R, Tobias R, Sall J. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM J Sci Comput*, 1994, 15: 1294–1310
 - 16 Lindstrom M J, Bates D M. Newton-raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Amer Statist Assoc*, 1988, 83: 1014–1022
 - 17 Fu L, Wang M, Wang Z, et al. Maximum likelihood estimation of nonlinear mixed-effects models with crossed random effects by combining first-order conditional linearization and sequential quadratic programming. *Int J Biomath*, 2019, 12: 1950040
 - 18 Browne W J, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal*, 2006, 1: 473–514

A general formulation of nonlinear mixed effect models and its application

Liyong Fu & Shouzheng Tang

Abstract Nonlinear mixed effect models (NLMEMs) have become popular in various disciplines over the past several decades. However, the existing methods for parameter estimation implemented in standard statistical packages such as SAS and R/S-Plus are generally limited to single- or multi-level NLMEMs that only allow nested random effects and are unable to cope with crossed random effects within the framework of nonlinear mixed effect modeling. In this paper, we propose a general formulation of NLMEMs that can accommodate both nested and crossed random effects, and then develop a computational algorithm (linearization approximation-sequential quadratic programming) for parameter estimation based on normal assumptions. The general formulation of NLMEMs and linearization approximation-sequential quadratic programming algorithm is evaluated using the height and diameter data measured on trees from Korean larch (*L. olgensis var. changpaiensis*) experimental plots as well as simulation studies. We show that the linearization approximation-sequential quadratic programming method converges fast with high accuracy and also were implemented in ForStat software. In addition, we also spread the normal NLMEMs standard expression to those NLMEMs where the variances of random effects are related to some factors (also called the group variable).

Keywords crossed random effects, nested random effects, first order conditional expansion, nonlinear mixed effect models, sequential quadratic programming

MSC(2010) 60J60

doi: 10.1360/N012018-00173