贝叶斯推断在土壤微生物生物地理学中的应用

柳旭[1,2],马玉颖[1],高贵锋[1],范坤坤[1],杨腾[1],褚海燕[1,2*]

1. 中国科学院南京土壤研究所,土壤与农业可持续发展国家重点实验室,南京 210008
2. 中国科学院大学,北京 100049

摘要 微生物生物地理学主要研究微生物的分布格局及其驱动机制,经典频率数理统计方法是当前该研究领域中广泛使用的统计方法。近年来,贝叶斯推断作为重要的随机模拟数理统计方法正不断地应用于土壤微生物生物地理学的研究中。介绍了贝叶斯推断与经典频率数理统计的区别;描述了贝叶斯推断在土壤微生物生物地理学研究中的基本分析流程,包括模型构建、模型拟合和模型优化;评价了贝叶斯推断在该研究领域中的自身优势、应用潜力和发展方向;并以岛屿生物地理学理论为研究框架,利用模拟数据,进行了贝叶斯推断流程演示。提出贝叶斯推断未来有望成为研究土壤微生物生物地理学中复杂数据和开展模型模拟的重要工具之一,在土壤微生物生物地理学中具有广阔的应用前景。

关键词 土壤微生物生物地理学;贝叶斯推断;经典频率统计;马尔科夫链蒙特卡洛抽样

土壤微生物驱动着陆地生态系统中几乎所有已知的生态过程,是联系大气圈、水圈、岩石圈和生物圈等不同圈层物质循环与能量流动的重要纽带[1]。土壤微生物生物地理学旨在研究土壤微生物多样性、群落组成和功能属性的时空分布格局,其研究有助于更好地理解微生物群落构建机制,认识微生物在调节关键生态系统过程中的重要作用[2-3]。21世纪以来,随着高通量测序技术的重大突破和生物信息学的高速发展,土壤微生物生物地理学得到了空前发展[4]。当前,土壤微生物生物地理学研究已从传统的土壤微生物的空间分布特征及其驱动机制,拓展到土壤微生物的种间互作关系,群落构建过程,生态系统功能耦联机制,以及全球变化背景下微生物分布的模型预测等方面[5]。尽

收稿日期:2021-07-19;修回日期:2021-11-22
基金项目:国家自然科学基金项目(31870480)
作者简介:柳旭,博士研究生,研究方向为土壤微生物空间分布,电子信箱:xliu@issas.ac.cn;褚海燕(通信作者),研究员,研究方向为土壤微生物学,电子信箱:hychu@issas.ac.cn
引用格式:柳旭, 马玉颖, 高贵锋, 等. 贝叶斯推断在土壤微生物生物地理学中的应用[J]. 科技导报, 2022, 40(3): 112-120; doi: 10.3981/j.issn.1000-7857.2022.03.010

管相关研究已取得了较大进展,但仍面临诸多困难与挑战。例如,复杂的生态数据难以得到全面、有效的利用;多层级的理论模型难以得到深入、合理地解读等。本文比较了经典频率学派和贝叶斯学派的数理统计方法,探讨了贝叶斯推断在土壤微生物生物地理学研究中的优势,按照贝叶斯推断的研究思路提出了整体性的操作流程和框架,着眼当前土壤微生物生物地理学中贝叶斯推断的应用,对其未来的方向进行了展望。

1　经典频率学派和贝叶斯学派

当前,土壤微生物生物地理学的研究主要采用基于频率的经典数理统计方法。然而,近年来出现了不少反对经典频率统计的声音[6-8]。例如,2019年3月20日《Nature》在线报道了一篇关于《科学家们反对统计显著性》的文章[9],800多位科学家联合抵制统计显著性检验的应用,指出经典频率统计在广泛应用的同时暴露出了一些先天的缺陷。例如,假定随机抽样样本为$\chi=(\chi_1,\chi_2,\cdots\chi_n)$,对样本进行$t$检验。如果$P$值小于$\alpha$(例如,$\alpha$=0.05),则统计显著,拒绝原假设$H_0$,接受备择假设$H_1$。实际上,$t$检验有多个假设前提,$P$值小于$\alpha$仅导出一个矛盾,并意味着每个假设前提都值得怀疑,而非仅选择拒绝原假设H_0并接受备择假设H_1这一条,因此经典频率统计中显著性检验得不到有意义的决策[10]。此外,经典频率统计更加关注假定总体分布的参数而非总体分布本身,其视假定的总体和参数为一成不变,然而真实生态学和生物学数据一些微小的变化就会导致原先主观的假定分布和参数出现偏差[11]。随着观测数据逐渐复杂化和研究深度的提高,研究者发现经典频率统计已不能满足所有的科学统计的需要。

贝叶斯推断是由英国学者Bayes在其发表的论文《论有关机会学说的求解》中提出的,并且在与经典频率学派争论中逐步发展起来[12]。贝叶斯思维提供了一种根据最新信息更新数据分布模式的机器学习过程,即它不把假定的分布和参数看成固定的,每当获得新数据时,原先的分布假定就会更新,这与经典频率学派学说不同。经典频率学派对未知参数进行统计推断前,就已经确定了总体分布和参数,即在假定模型中的总体参数是固定的,而观测数据是在总体分布中随机抽样所得。贝叶斯理论将我们更关注的参数看成随机变量,其分布也随着新信息的获得而更新。基于以上事实,我们总结了经典频率学派和贝叶斯学派在基本理论上的3大差异(表1)。

1)对于概率问题的解释不同。经典频率学派

表1　经典频率学派和贝叶斯学派的对比

学派	概念	基本特征	关于概率的解释	关于统计信息的选择	关于随机变量的认识	局限性
经典频率学派	利用抽样样本信息对总体信息或者总体特征进行未知参数θ的统计推断	未知参数θ固定分布,抽样样本信息随机分布	主张客观概率,用频率性对“概率”的进行解释。在大样本空间下定义的,通过大量重复试验的频率来近似于它的概率	经典学派在统计推断中主要运用总体信息和样本信息	样本信息是随机变量。总体信息是具有一定的概率分布,样本来源于总体,运用样本的信息可以推断出总体的性质	显著性检验得不到有意义的决策
贝叶斯学派	通过贝叶斯公式将先验信息、抽样样本信息、总体信息结合,对未知参数θ进行统计推断	未知参数θ随机分布,抽样样本信息固定分布	以主观为基础,引入了先验信息,把长久以来形成的看法、经验或者信念等当作先验信息引入统计推断中。主张对概率的双重解释:“频率”的客观解释与经验的“主观”理解	贝叶斯学派运用先验信息、总体信息和样本信息,认为先验信息具有同等重要的地位	未知参数θ是随机变量,且具有一定的概率分布	先验信息的准确性和数据集的大小依赖于主观经验的判断

认为“概率”是客观的,可以使用大量的重复试验所得的频率进行解释。而贝叶斯学派认为“概率”是主观的,大量重复试验在一些研究情境下是不现实的,如微生物的系统发育关系。因此,贝叶斯推断在一些生态和进化情景下具有根据对事件的了解和经验判断其发生可能性的优势。

2)对于统计推断利用的信息不同。经典频率学派的统计推断是根据总体信息和样本信息。而贝叶斯学派在此基础上还增加了先验信息来提高统计推断的质量,这使得贝叶斯推断具有可以综合考虑先前研究基础并对复杂模型进行推断的优势。

3)对于随机变量的认识不同。经典频率学派把样本看作是具有一定概率分布的总体。而贝叶斯学派把任何一个研究参数都看作随机变量,具有不确定性的特点,通过随机抽样概率分布来描述未知参数。因此,贝叶斯推断具有利用未知参数分布来准确预测研究对象的优势。

2 贝叶斯推断的分析流程:模型构建、拟合和优化

贝叶斯推断的分析流程主要包括模型的构建,模型对数据的拟合,以及通过检查模型的拟合或与其他模型进行比较来优化及拓展模型(图1)[13]。第一步,在构建模型时,使用主观经验和先验知识,有时也使用已经建立的理论。这个步骤在本质上通常是构建数学模型,因为它包含了贝叶斯定理中的似然性函数、先验分布和后验分布的显式定义。这些概念囊括了有关数据和参数的背景信息。模型的数据拟合是贝叶斯数据分析流程的核心。在当前方法中,它通常使用马尔科夫链蒙特卡洛算法进行独立抽样获得未知参数的随机概率分布[14]。贝叶斯推断是通过分布函数来描述未知参数,故其结果可以对研究对象准确地进行预测。第二步的目的是计算研究对象未知参数的后验概率分布和置信区间,从其先验分布开始计算,并根据新数据的补充进一步调整后验分布和参数。模型总是现实的近似值,因此评估它们的拟合优度并改进它们是贝叶斯分析的第三步。在这个步骤中,经常需要面对多组模型结构和模型参数[15],例如,基于不同生态理论构建的数学模型和不同的未知参数假定分布。以上3个步骤相互依存,共同形成了贝叶斯推断的基本逻辑架构。一个具有实际价值的贝叶斯模型通常需要3个步骤往复循环不断调整得到。

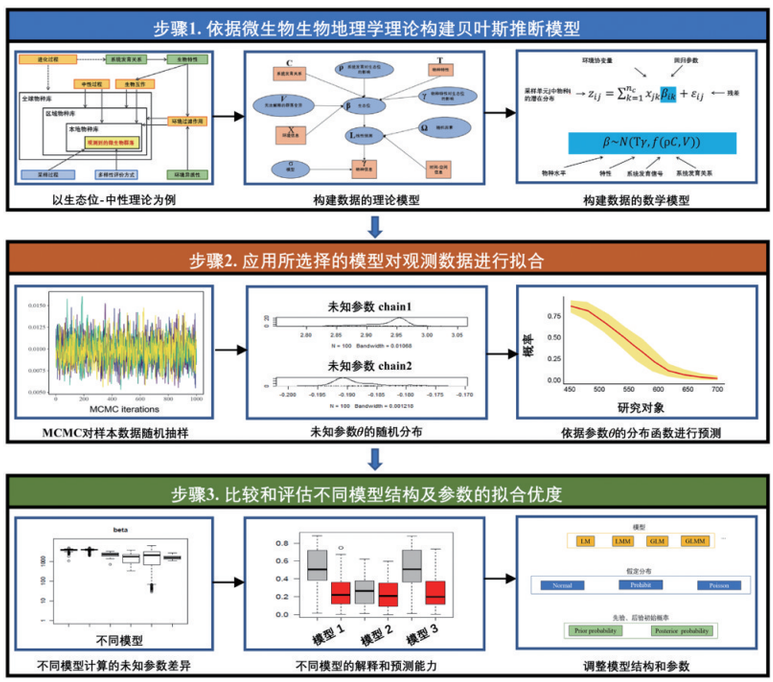


图1 土壤微生物生物地理学应用贝叶斯推断的数据分析流程

2.1　模型构建

在构建或改进模型时,会出现2个密切相关的问题。第一,为什么需要模型?构建模型可以用于2个不同的目的:解释变异和预测未知。以生态位-中性理论为例,构建观测数据的理论模型,即土壤微生物生物地理格局受控于由系统发育关系、物种特性和环境异质性组成的生态位固定因素和采样位置及时间背景的中性随机因素[16]。模型可以把生态理论落地为数学表达式,将关注的研究对象定量。第二,所构建的模型应该有多复杂?在本研究中,由于观测数据包含固定因素和随机因素,且生态和生物数据格式以矩阵形式出现,所以应用广义线性混合模型构建观测数据的数学表达式。但是,所构建的模型应依据观测数据格式和研究目的灵活调整[17]。例如,探索环境异质性对土壤微生物生物地理分布的影响时,由于不考虑随机因素的作用,可将广义线性混合模型修改为广义线性模型。

2.2　模型拟合

贝叶斯推断数据拟合算法主要包括2大类。第一类算法依赖于传统拒绝算法,即通过使用局部线性或非线性回归技术,对所构建的数学模型进行了技术改进,修正了模拟数据和观测数据之间的差异[13]。第二类算法应用马尔科夫链蒙特卡洛(Markov Chain Monte Carlo,MCMC)独立抽样,利用模拟和观察汇总统计数据间的距离迭代探索参数分布空间,即从先验分布开始独立重复抽样形成未知参数的后验分布,更新当前参数值[14,18]。在MCMC方法中,要求马尔可夫链收敛于它的稳态,由于需要靠主观判断,因此这是一个很难验证的条件。本研究以第二类MCMC算法为例,展示了MCMC对未知参数独立随机抽样,以获得其随机概率密度分布。进而,根据未知参数的概率分布函数可对研究对象进行预测。

2.3　模型优化

比较模型和评估它们的拟合优度是构建模型和统计推理过程中的重要步骤。模型比较通常基于一个决策理论框架,其目标是选择得到更高后验分布或后验支持的模型。在贝叶斯数据分析流程的研究中,比较2个模型通常使用贝叶斯因子(Bayes Factor)[19]。作为比较检验指标,贝叶斯因子对于贝叶斯学派的作用就像P值对于经典频率学派的作用一样。在显著性检验中,P值用于评估若原假设H_0为真,观察数据接受H_0可能性有多大,而在贝叶斯学派选择模型中,贝叶斯因子比较和评估不同构建模型时每个模型都对应一个特定的假设。在这个过程中,模型选择并不意味着选择一个单一的最佳模型。土壤微生物生物地理分布格局也可能对应着不同的模型,从而从不同的角度解释分布机制[20]。例如,存在-缺失(presence-absence)物种分布数据服从Probit分布或Logit分布,而丰度物种分布数据服从Poisson分布。我们不应该只关注一个模型,而应该考虑每个备选模型的合理性,并最终加权多个模型的参数估计。此外,模型的检查同样有必要,其目的是理解模型与观测数据匹配程度,进而调整模型的结构和模型的具体参数。

3　贝叶斯推断在土壤微生物生物地理学中的应用范围

贝叶斯推断具有极强的灵活性,只要有统计汇总需求和允许随机抽样,就可以对许多复杂模型和参数进行推理和研究。在当前的土壤微生物生物地理学研究中贝叶斯推断已多有涉及(图2)。

贝叶斯推断在获取土壤微生物生物地理学研究相关的基础数据时被大量使用。例如,QIIME2中的默认注释方法scikit learn应用的朴素贝叶斯(Naïve Bayes)算法,其在贝叶斯定理的基础上,利用"朴素"(Naïve)假定的每一组特征之间的条件独立性来给定注释变量的一套监督学习算法[21]。利用的贝叶斯推断的注释方法从注释准确性和注释物种范围均优于传统物种注释方法,如QIIME1中默认算法UCLUST[22]。此外,由于采样数量和范围的限制,近年来贝叶斯推断模型预测环境因子作为研究背景数据也被广泛应用。例如,Kaye等[23]建立了城市、农业和荒漠生态系统下的多水平贝叶斯土壤养分模型以获取不同生态系统更为准确的土壤性质;Majumdar等[24]通过贝叶斯层级模型建立了土

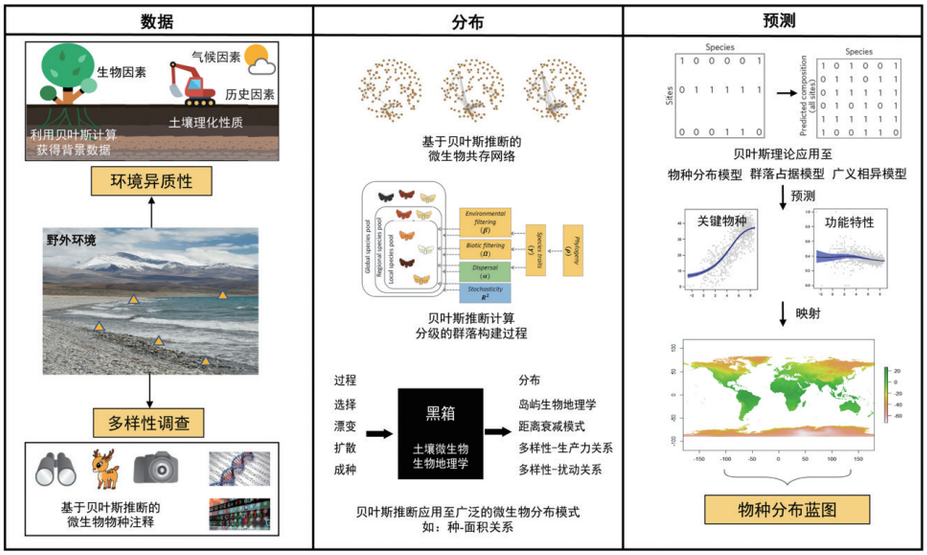


图2 贝叶斯推断在土壤微生物生物地理学中的应用

壤碳密度空间分布模型以获得不同尺度土壤碳密度数据。

在研究土壤微生物生物地理分布时,贝叶斯推断展示出对于复杂模型推断的优势。例如,物种群落层级模型(HMSC)根据群落生态学理论把群落变异划分为环境过滤作用、种间相互作用、以及随机因素的作用,通过贝叶斯推断计算不同过程对群落变异的效应值[25]。此外,贝叶斯推断在传统生物地理理论中也有所应用。例如,Jabot等[26]使用贝叶斯推断生物多样性的中性理论的参数。

在预测土壤微生物关键物种分布和功能特性时,贝叶斯推断同样是提高预测精度和准度的重要潜在方法。例如,群落占据模型(community occupancy model)[27],其应用贝叶斯推断可计算微生物的不同环境生态位宽度,并可预测不同环境中潜在微生物多样性。此外,贝叶斯推断在其他预测模型,如物种分布模型、广义相异模型等中也大有用武之地[28]。

4 案例剖析:贝叶斯推断与岛屿生物地理学理论

在本案例中,我们利用模拟岛屿生物地理学的土壤微生物生物地理分布的背景数据[29],从模型构建、拟合、优化3个方面,进行了简单的贝叶斯推断分析(图3)。

4.1 模型优化

以MacArthur 和Wilson于1967年提出的岛屿生物地理学理论作为模型生态框架[30]。其以种群生态学和遗传学的基本原理为基础,阐述隔离程度和岛屿面积调节岛屿种群迁移和灭绝之间的动态平衡,并且假定存在一个稳定的、永久性的大陆物种库,岛屿物种丰富度由岛屿的几何特征决定,即岛屿面积决定了物种灭绝率[31],其数学表达式为:

$$S = C \times A^{z} \quad (1)$$

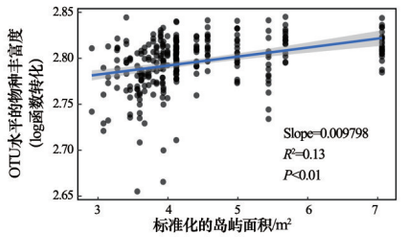
式中,S代表群落层次物种丰富度,A代表岛屿面积,C为与生物地理区域有关的拟合参数,Z为与岛屿面积有关的拟合参数(岛屿面积效应值)。为适配贝叶斯推断模型计算,对传统岛屿生物地理学的数学表达式(1)进行log数学转化,使得物种丰富度与岛屿面积之间符合线性关系,作为模型的数学框架。其表达式变形为:

$$\lg S = \lg C + Z\lg A \quad (2)$$

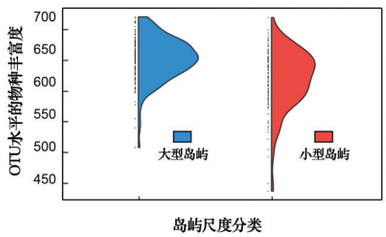
也可进一步简化为:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (3)$$

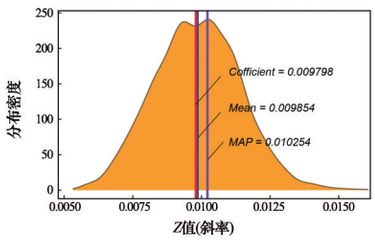
式中,y代表式(2)中的$\lg S$,α代表$\lg C$,β代表与岛屿面积有关的拟合参数Z值,ε代表该模型的随机误差。



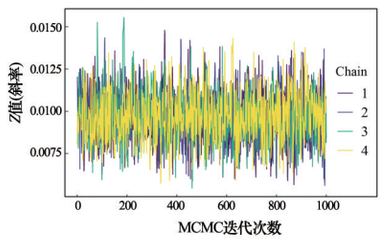
(a) 岛屿面积和微生物丰富度之间的经典线性回归



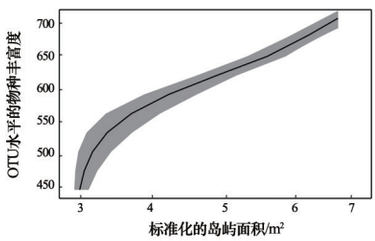
(b) 按照岛屿面积大小分类的微生物OTU物种丰富度



(c) Z值(斜率)的后验概率密度分布



(d) 马尔科夫链蒙特卡洛(MCMC)抽样计算Z值的后验分布



(e) 岛屿尺度预测土壤微生物多样性

Slope代表线性回归斜率;OTU为微生物可操作分类单元;Mean代表数据分布的平均值;MAP(highest maximum a posteriori)代表Z值密度分布峰值,系数与经典回归的斜率一致

图3　贝叶斯推断分析案例

该简化步骤进一步明晰了具有生态学意义的生物地理学模型与数学统计模型的直接联系。

4.2　模型拟合

首先,计算了基于经典频率统计的岛屿效应对土壤微生物物种丰富度的影响(Slope=0.009798,P<0.01,图3(a))。按照岛屿面积大小,展示了土壤微生物物种丰富度的分布和差异(t=3.82,P<0.01,图3(b))。Slope对应表达式(2)中的Z,表征该区域尺度下的岛屿面积的效应值。

其次,按照贝叶斯模型拟合流程,以岛屿生物地理学为构建模型的生态理论基础,根据构建数学模型表达式(2)和(3),对未知参数Z进行统计推断(图3(c))。观测数据对构建的模型拟合方法采用MCMC独立抽样算法(图3(d))。设置4条独立运行的马尔科夫链,以保证数据的精度和可靠性;设置1000个MCMC迭代为预热抽样(warm-up),目的是将先验信息控制在合理的置信区间。参数Z的随机抽样概率密度分布均值为0.009854,与经典频率统计计算的岛屿面积效应值相近。该结果表明贝叶斯方法在传统土壤微生物生物地理学研究中也能得到有意义的结果。值得注意的是,若不设置随机因子(seed),不同模拟抽样计算的结果可能出现差异。

进而,依据抽样所得Z值的随机概率分布对研究对象土壤微生物多样性进行预测(图3(e))。结果发现,当研究对象的面积尺度为大型岛屿时(按照本例分类),有较大把握预测岛屿土壤微生物群落丰富度超过600。

4.3　模型优化

群落水平的特征数据,例如,物种丰富度或者生物量,应用基于经典频率统计的线性模型所得到的结果与基于贝叶斯推断的结果相似,仍然能够满足研究的需求。然而,当关注物种个体水平对环境变量的响应或者环境对物种个体的效应值时,基于经典频率统计的研究假设将不再适应当前的生态大数据,甚至可能得到错误的结果。为探索从群落水平到个体水平的环境效应值,对模型进行了优化拓展[32]。其数学表达式可以推广为:

$$m_{ij} \sim D(L_{ij}^{F}, \sigma_{ij}^{2}) \quad (4)$$

式中,m代表物种丰度数据矩阵,D代表物种数据服从的数学分布模型,L代表环境预测值,σ^2代表预测方差,L由固定因素和随机误差组成,其数学表达式为:

$$L_{ij}^{F} = \sum X_{i\cdot}\beta_{\cdot j} + \varepsilon \quad (5)$$

式中,X为不同样点的岛屿面积,β为岛屿面积的效应值,假定其服从正态分布:

$$\beta_{\cdot j} \sim N\left(\mu_{\cdot j}, \nu_{\cdot j}\right) \quad (6)$$

β计算过程与本例中的模型拟合类似,由马尔科夫链蒙特卡洛独立抽样计算,通过重复抽样的策略,将计算效应值的问题转化成了计算未知参数随机概率分布的问题,对应物种个体水平的岛屿面积效应值即可计算得出。

5 展望与思考

土壤微生物生物地理学是一门极为复杂的科学。因此,野外和实验室环境下的观测数据将不可避免地促使建立复杂的模型进而解释现象背后的机制。近年来,贝叶斯推断已成为土壤微生物生物地理学研究中的重要统计工具,并得到一定应用和推广(图2)。然而,基于土壤微生物生物地理学的复杂性和学科特性,还有很多方面的问题值得深入研讨。在此,对贝叶斯推断在土壤微生物生物地理学中的应用做出如下展望。

1)土壤微生物生物地理学研究中的观测数据所服从的数学分布类型需要进一步明确。模型参数的设置决定着研究结果的准确与可靠。如我们通常假设生态环境数据需服从正态分布。然而,现实中的环境及微生物数据可能与假定分布相左,设置不同的数学分布作为假设前提将显著地影响最终的结果和结论。

2)土壤微生物随时间变化规律的研究存在很大空白。尽管很多研究已经观测到不同尺度的土壤微生物随时间变化的情况,如季节、昼夜等。然而,受限于当前所用的方法,时间对于土壤微生物分布的影响仍难以准确评估。基于贝叶斯推断的多层级模型或许可以很好的解决这一问题,帮助我们更好地理解土壤微生物的时空分布规律。

3)预测不同尺度土壤微生物对人类扰动和气候变化的响应成为重点研究内容。贝叶斯推断一个很重要的特点就是通过对随机抽样的概率密度分布来描述未知参数。这为贝叶斯模型预测提供了可靠的数据保障,可以帮助我们进一步评估土壤生物多样性和生态系统功能随环境变化而改变的阈值和速率等关键问题。

此外,贝叶斯推断在大数据模拟计算中具有比较明显的优势,但不能忽略推断复杂模型背后土壤微生物生物地理学机制的难点。推理机制的过程将受到模型构建和模型拟合的阻碍,因为这些关键步骤存在高度的主观性。目前,贝叶斯推断统计远远未达到与经典频率统计一样成熟的程度。因此,研究者进行科学研究时应合理选取适合数据结构和研究目的的方法。贝叶斯统计和经典统计不是相互对立的,而是相辅相成的。因此,归纳使用贝叶斯推断时的注意事项。

1)由于贝叶斯推断不可避免地使用部分数据,造成隐含信息缺失问题,从贝叶斯推断数据分析中获得的置信区间结果有可能被夸大。研究者需对贝叶斯推断得到的置信结果解读持谨慎态度。

2)所有模型都是研究者主观构建的。因此,模型检验和优化将会改善数据与模型之间的拟合优度,是探索和理解模型及数据之间差异的一种重要的方式。

3)多层级建模技术和模型选择的信息理论将有助于降低模型的主观性。层级模型有助于理解参数间依赖关系,并减少建模无法预见的隐式前提。随着高度结构化层级模型的发展,基于信息理论度量的模型复杂性的方法也有助于评估模型准确性,如赤池信息准则(AIC)或偏差信息准则(DIC)。

4)观测数据对某一模型的支持并不意味着该模型一定真实。如果不同模型有共同的参数,对不同的、可信度高的模型的参数加权,可以产生比单一模型更稳健的推断。

6 结论

土壤微生物生物地理学已成为土壤生物学和微生物生态学等研究领域的热点。随着研究深度和广度的不断增加,相信贝叶斯推断将成为土壤微生物生物地理学研究的重要统计工具之一。同时,贝叶斯推断算法也在不断发展,如处理更高维的生态数据集,评估不同模型的复杂性、准确性,以及建立更为有效的模型检验、优化等方法。贝叶斯推断算法的进步将不断推进土壤微生物生物地理学的发展。

参考文献(References)

[1] Fierer N. Embracing the unknown: Disentangling the complexities of the soil microbiome[J]. Nature Reviews Microbiology, 2017, 15(10): 579-590.

[2] Tedersoo L, Bahram M, Polme S, et al. Fungal biogeography. Global diversity and geography of soil fungi[J]. Science, 2014, 346(6213): 1256688.

[3] Fierer N, Jackson R B. The diversity and biogeography of soil bacterial communities[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(3): 626-631.

[4] 褚海燕, 王艳芬, 时玉, 等. 土壤微生物生物地理学研究现状与发展态势[J]. 中国科学院院刊, 2017, 32(6): 45-52.

[5] Chu H, Gao G, Ma Y, et al. Soil microbial biogeography in a changing world: Recent advances and future perspectives[J]. Msystems, 2020, 5(2): e00803-19.

[6] Wasserstein R L, Lazar N A. The ASA′s statement on p-values: Context, process, and purpose[J]. American Statistician, 2016, 70(2): 129-131.

[7] Wasserstein R L, Schirm A L, Lazar N A. Moving to a world beyond "*P*<0.05"[J]. American Statistician, 2019, 73: 1-19.

[8] Benjamin D J, Berger J O, Johannesson M, et al. Redefine statistical significance[J]. Nature Human Behaviour, 2018, 2(1): 6-10.

[9] Amrhein V, Greenland S, McShane B. Retire statistical significance[J]. Nature, 2019, 567(7748): 305-307.

[10] Jain A K, Duin R P W, Mao J C. Statistical pattern recognition: A review[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4-37.

[11] 谢俊. 贝叶斯统计方法与传统统计方法的比较分析与展望[J]. 中国商界, 2009(4): 115-117.

[12] Gelman A. Bayesian statistics then and now[J]. Statistical Science, 2010, 25(2): 162-165.

[13] Csillery K, Blum M G, Gaggiotti O E, et al. Approximate bayesian computation (ABC) in practice[J]. Trends in Ecology and Evolution, 2010, 25(7): 410-418.

[14] Jackman S. Estimation and inference via bayesian simulation: An introduction to markov chain monte carlo[J]. American Journal of Political Science, 2000, 44(2): 375-404.

[15] Cornell S J, Suprunenko Y F, Finkelshtein D, et al. A unified framework for analysis of individual-based models in ecology and beyond[J]. Nature Communications, 2019, 10: 14.

[16] 曹鹏, 贺纪正. 微生物生态学理论框架[J]. 生态学报, 2015, 35(22): 6-16.

[17] Ma Kowski, Ben-Shachar M S, Lüdecke D. BayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework[J]. The Journal of Open Source Software, 2019, 4(40): 1541.

[18] Hadfield J D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package [J]. Journal of Statistical Software, 2010, 33(2): 1-22.

[19] Makowski D. Indices of effect existence in the bayesian framework[J]. Frontiers in Psychology, 2018, 10: 2067.

[20] Vellend M. Conceptual synthesis in community ecology [J]. Quarterly Review of Biology, 2010, 85(2): 183-206.

[21] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python[J]. Journal of Machine Learning Research, 2011, 12: 2825-2830.

[22] Edgar R C. Search and clustering orders of magnitude faster than BLAST[J]. Bioinformatics, 2010, 26(19): 2460-2461.

[23] Kaye J P, Majumdar A, Gries C, et al. Hierarchical bayesian scaling of soil properties across urban, agricultural, and desert ecosystems[J]. Ecological Applications, 2008, 18(1): 132-145.

[24] Majumdar A, Kaye J, Gries C, et al. Hierarchical spatial modeling and prediction of multiple soil nutrients and carbon concentrations[J]. Communications in Statistics B Simulation and Computation, 2008, 37(1/2): 434-453.

[25] Tikhonov G, Opedal O H, Abrego N, et al. Joint species distribution modelling with the r-package Hmsc[J]. Methods in Ecology and Evolution, 2020, 11(3): 442-

447.

[26] Jabot F, Chave J. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests[J]. Ecology Letters, 2010, 12(3): 239-248.

[27] Ribeiro J W, Siqueira Jr T, Brejao G L, et al. Effects of agriculture and topography on tropical amphibian species and communities[J]. Ecological Applications, 2018, 28(6): 1554-1564.

[28] Bush A, Sollmann R, Wilting A, et al. Connecting Earth observation to high-throughput biodiversity data[J]. Nature Ecology and Evolution, 2017, 1(7): 9.

[29] Li S P, Wang P, Chen Y, et al. Island biogeography of soil bacteria and fungi: Similar patterns, but different mechanisms[J]. The ISME Journal, 2020, 14(7): 1886-1896.

[30] MacArthur R H, Wilson E O. The theory of island biogeography[M]. Princeton: Princeton University Press, 2001.

[31] Whittaker R J, Maria Fernandez-Palacios J, Matthews T J, et al. Island biogeography: Taking the long view of nature′s laboratories[J]. Science, 2017, 357(6354): 885.

[32] Ovaskainen O, Rybicki J, Abrego N. What can observational data reveal about metacommunity processes? [J]. Ecography, 2019, 42(11): 1877-1886.

Application of Bayesian inference in soil microbial biogeography

LIU Xu[1,2], MA Yuying[1], GAO Guifeng[1], FAN Kunkun[1], YANG Teng[1], CHU Haiyan[1,2*]

1. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

2. University of the Chinese Academy of Sciences, Beijing 100049, China

Abstract Microbial biogeography mainly focuses on the distribution patterns of microorganisms and their driving mechanisms, and currently classical frequency mathematical and statistical methods are widely used in this field. In recent years, Bayesian inference has been continuously applied as an important random simulation mathematical statistics method for soil microbial biogeographical study. In this paper, we briefly introduce the differences between Bayesian inference and classical frequency mathematical statistics, and highlight the basic analytical process of Bayesian inference in soil microbial biogeographical study in terms of model construction, model fitting, and model improvement. We also evaluate the advantages, application potentials and development directions of Bayesian inference in this field and demonstrate the Bayesian inference process using simulated data with island biogeography theory as the research framework. Finally, we give an outlook on the application prospects of Bayesian inference in soil microbial biogeography. We believe that Bayesian inference will become one of the important tools for studying complex data and conducting model simulations in soil microbial biogeography.

Keywords soil microbial biogeography; Bayesian inference; classical frequency statistics; Markov chain Monte Carlo sampling

（责任编辑　徐丽娇）