

物种间不确定性相互关系分析: 一种基于非参数估计的变系数模型

石磊^{①②*}, 王瑞武^{②*†}, 朱力行^{③*}, 曾卫民^④, 许王莉^⑤, 郑琪^②

① 云南财经大学统计与数学学院, 昆明 650221;

② 中国科学院昆明动物研究所, 遗传资源与进化国家重点实验室, 生态学与环境保护中心, 昆明 650223;

③ 香港浸会大学数学系, 香港;

④ 云南民族大学科学技术系, 昆明 650031;

⑤ 中国人民大学统计学院, 北京 100080

* 同等贡献

† 联系人, E-mail: ruiwukiz@hotmail.com

2011-02-25 收稿, 2011-05-13 接受

国家重点基础研究发展计划(2007CB411600)、国家自然科学基金(30670270, 30770500, 10761010)、云南省自然科学基金、中国科学院“西部之光”人才培养计划、中国科学院优秀青年科技专项(KSCX2-EW-Q-9)、遗传资源和进化国家重点实验室开放课题基金、国家哲学社会科学基金(08XTJ001)、香港研究资助局项目(HKBU2030, 07P)资助

摘要 物种间相互作用的研究通常假定物种间相互作用具有某种固定的模式, 因此线性或者非线性参数回归模型(如指数分布或者 logistic 模型)被广泛用来描述物种间的相互作用。然而, 这些模型假定了相互作用关系物种之间具有特定响应函数, 而这个假设可能并不适用于真正的生物群落, 如物种间相互作用关系随不同环境变化的混沌系统。为了用一个更精确的方法来描述这种关系, 我们以榕树-榕小蜂为模式系统, 建立了一个变系数分析模型进行物种间相互关系的分析。在这个模型中, 用非参数估计方法分析物种间相关系数如何随其他变量(函数)的变化而变化。当其他因素对相关系数的影响效应可以用一个指定的参数模型来描述时, 新方法就会转化为传统的参数相关分析。对于经验数据的分析, 新方法更具有普遍性和灵活性, 其与传统方法不同之处在于我们可以研究物种间的相互作用是否会随着某些因素变化而变化, 或者找出维持或改变物种间的相互作用其他关联因素。这种方法在理论研究和应用研究(如流行病学和群落管理)领域都有着重要的应用价值。

关键词

物种相互作用
非参数估计
相关系数
变系数分析
混沌震荡
密度依赖性
不确定相互关系

物种间相互作用的研究是了解群落生态学的基础。通常假定物种间有一个固定的相互作用, 这种关系可以通过物种数量特征的相关分析来描述。物种的数量特征可能是物种的适合度或种群大小、个体增长率、个体的免疫性或化学反应等。在这个假定的基础上, 一个变量对另一个变量的影响就可以用线性回归或者转换为线性回归的非线性参数回归来描述, 其相关系数最终可以描述为一个固定值或常数^[1,2]。在这些分析中, 相互作用的物种间的数量关系被认

为是确定的, 换句话说, 它可以用一个特定值来描述^[1,3-9]。而其他因素被简单假定是通过线性方式影响有关物种的生物学特征^[7,10-13]。在统计分析中应用一般线性模型时, 这些线性影响变量被当作协变量^[1,2,14]; 然而, 这种方法并不适用于很多生物系统中物种相互作用的相关分析^[1-18]。

在一个真实的生态群落中, 物种间的相互作用关系也许不能用固定的相关系数来描述。例如, 物种的关系可能是非线性的^[16], 其关系系数甚至可能随

着其他因素变化而出现混沌性震荡,也就是物种间相互关系可能存在不确定性相互作用^[13,19-22].物种关系可能会根据环境、生态条件、相关种群的种群密度或者一些其他因素而变化^[3,4,20,22-27].也就是说,这些相互作用的物种间可能并不存在某种特定的关系.物种间的这种不确定性相互作用在互利共生、共存或者对抗的生态系统中都有记录(如捕食者和被捕食者或者寄生者和宿主)^[19,22,24,27-31].遗憾的是,至今几乎没有有效的分析方法来定量地描述这种现象.

为解决这个问题,一种很恰当的方法是用一个模型去拟合潜在关系.通过这个模型,我们可以研究影响因素不同取值范围内物种间相关系数大小的变化,这就使得我们可以运用一种比传统方法更加综合、全面的分析.为了达到这个目的,本文提供了一种用于研究物种间相互作用变系数分析的新模型,并且将其应用于榕树-榕小蜂系统.这种方法将适用于任何物种间相互作用的分析,而传统方法将转化为本分析的一种特例.

1 线性回归模型简介

令 X 和 Y 表示物种的数量特征(变量).如果物种间的相互作用可以用一个反映数量特征的常量来表示,就可以用如下的线性回归模型来描述他们的相互作用

$$Y = \alpha_0 + \alpha_1 X + \varepsilon, \quad (1)$$

其中,斜率 α_1 相当于 Y 和 X 的相关系数,它用来描述 2 个物种的相互作用形式.当对 X 和 Y 进行适当的变化后,像指数分布或者 logistic 模型一样更为普遍的模型也可以转换为上面的结构.这个模型中,如果相互作用的物种间存在一种固定或者不变的关系,那么用一个显著的反映物种数量特征的正斜率就可以描述合作的物种,而用一个显著的负值来描述对抗性的物种,此外,用一个不显著的斜率来描述无相互作用物种或者有着相同的栖息地但具有不同生态位的共存物种^[3,5-8,11,12,32].

1.1 变系数模型

研究表明,物种间的相互作用可以随温度、季节和包括资源在内的其他非生物因素的变化而变化^[22,27,33].物种间的相互作用也可以是密度依赖的,并将随着其他物种的种群密度变化而变化^[16,34,35].物种间相互作用的变化存在数量化的差异^[16],但其性

质也可能发生根本的转变.在一些生态系统中,物种间的相互作用甚至可以由正相关转化为负相关,如 *Greya politella* 蛾及其宿主植物 *Lithophragma parviflorum* (虎耳草科)之间的互利共生、传粉榕小蜂及其宿主植物榕树之间的互利共生^[27,29,36]、聚果榕 *Ficus racemosa* 的寄生蜂 *Platyneura agraeensis* 及其宿主传粉蜂 *Ceratosolen fusciceps* 间的寄生-宿主系统^[33]等.

用变量 Z 表示可能导致物种数量特征的相关系数变化的因素,则含有泛函参数的回归模型如下所示

$$Y = \alpha_0(Z) + \alpha_1(Z)X + \varepsilon, \quad (2)$$

其中, ε 是和 Z 相互独立且均值为 0 的随机误差项, $\alpha_0(Z)$ 和 $\alpha_1(Z)$ 分别为截距项和斜率项,都是关于 Z 的函数.在模型(2)中,如果 $\alpha_1(Z)$ 是和第 3 个变量 Z 相互独立的(如斜率为常量),那么它就可以转化为传统线性回归模型(1).如果 Z 对相互作用物种的数量特征的影响不是固定的,这个模型就叫做变系数回归模型^[37].令 $\rho_{XY|Z}$ 表示当 Z 给定时 X 和 Y 的条件相关系数,则表达式为

$$\alpha_1(Z) = \rho_{XY|Z} (\sigma_{Y|Z} / \sigma_{X|Z}), \quad (3)$$

其中, $\sigma_{X|Z}^2$ 和 $\sigma_{Y|Z}^2$ 分别是给定 Z 时, X 和 Y 的条件方差.如果 $\sigma_{Y|Z} / \sigma_{X|Z}$ 不显著地依赖于 Z ,那么相关系数 $\rho_{XY|Z}$ 就作为 Z 的一个函数等价于 $\alpha_1(Z)$.

在模型(2)中, $\alpha_0(Z)$ 和 $\alpha_1(Z)$ 的函数形式没有预先指定,通用的方法是采用非参数平滑技术来估计这些函数.这种非参数估计方法的优势在于无需预先规定函数形式,仅仅需要借助数据信息获得估计曲线/函数.由于其他参数方法利用假定的函数结构具有做出错误假设或得到错误的结论等风险,因此这种非参数模型比其他带参数方法更具有灵活性和可靠性.定义 $\alpha(Z) = (\alpha_0(Z), \alpha_1(Z))'$, $U = (1, X)'$, 此处上标号表示向量的转置.当矩阵 $E(UU'|Z)$ 可逆时, $\alpha(Z)$ 的最小二乘估计就可以由下式唯一的确定:

$$\hat{\alpha}(Z) = E(UU'|Z)^{-1} E(UY|Z). \quad (4)$$

条件期望可以用诸如核估计^[38]等非参数平滑方法来估计.假设 $\{(X_i, Y_i, Z_i), i = 1, 2, \dots, n\}$ 是一观测样本,则

$$E(UU'|Z) = \frac{1}{n} \sum_{i=1}^n w_{ni}(Z) U_i U_i',$$

$$E(UY|Z) = \frac{1}{n} \sum_{i=1}^n w_{ni}(Z) U_i Y_i, \quad (5)$$

这里的 $U_i = (1, X_i)'$, $w_{ni}(Z) = K_h(Z - Z_i) / f_h(Z)$, $f_h(Z) = n^{-1} \sum_{i=1}^n K_h(Z - Z_i)$, $K_h(u) = h^{-1} K(u/h)$.核函数 $K(u)$ 是

一个连续有界对称的实函数且积分等于 1 (e.f. $\int K(u)du = 1$). h 是一个标量, 表示带宽. 核函数具有多种形式, 常用的一些核函数包括 Epanechnikov 核函数^[38] $K(u) = 0.75(1-u^2)I(|u| \leq 1)$, 二次方核函数 $K(u) = 15/16(1-u^2)^2 I(|u| \leq 1)$ ^[39], 这里的 $I(|u| \leq 1)$ 是一个示性函数, 当 $|u| \leq 1$ 时为 1, 否则为零. 权函数 $w_m(Z)$ 在区间 $|Z-Z_i| \leq h$ 外为零. 已经证明, 随着 h 的增大, 会导致估计量的偏差变大, 方差变小, 这种现象被称为过度平滑; h 值的减小会导致估计量的偏差变小, 方差变大, 这种现象被称为不充分平滑. 文献[39]中提出了利用交叉验证的方法得到合适的带宽 h 值. 在实际应用中, 可以根据具体的分析目标选择带宽 h 值^[39].

尽管我们可以从拟合曲线确定回归系数和 Z 的差别, 但是对曲线的函数形式仍然未知. 文献[40]中提出了一种验证 $\alpha_1(Z)$ 是否是一个特定函数形式的方法——非参数蒙特卡洛检验(NMCT).

在可变参数回归模型中, 当 $\alpha_{Y|Z}/\alpha_{X|Z}$ 的比值是定值时, 可以利用回归系数 $\alpha_1(Z)$ 来描述相关系数 $\rho_{XY|Z}$, 否则, $\alpha_1(Z)$ 不等于 $\rho_{XY|Z}$, 此时需要对 $\rho_{XY|Z}$ 作非参数估计, 如下式所示

$$\rho_{XY|Z} = \frac{\text{Cov}(X, Y | Z)}{\sqrt{\text{Var}(X | Z)\text{Var}(Y | Z)}} = \frac{\sigma_{XY|Z}}{\sigma_{X|Z}\sigma_{Y|Z}}. \quad (6)$$

当给定 Z 值时, 利用核函数可以计算出 X 和 Y 的条件协方差和方差^[39]. 例如, 使用同样的记法, 我们可以得到

$$\begin{aligned} \sigma_{XY|Z} &= \frac{1}{n} \sum_{i=1}^n w_m(Z)(X_i - \hat{E}(X | Z))(Y_i - E(Y | Z)), \\ \sigma_{X|Z}^2 &= \frac{1}{n} \sum_{i=1}^n w_m(Z)(X_i - \hat{E}(X | Z))^2, \\ \sigma_{Y|Z}^2 &= \frac{1}{n} \sum_{i=1}^n w_m(Z)(Y_i - \hat{E}(Y | Z))^2, \end{aligned}$$

$$\text{这里, } \hat{E}(X | Z) = \frac{1}{n} \sum_{i=1}^n w_m(Z)X_i, \hat{E}(Y | Z) = \frac{1}{n} \sum_{i=1}^n w_m(Z)Y_i.$$

即便对一个复杂的结构, 类似于线性模型, 我们可以使用相关系数 $|\rho_{XY|Z}|$ (虽然它是 Z 的函数) 来研究变量间的相关性.

1.2 模拟研究

本文提出的可变参数模型非参数估计的计算通过 Matlab 编程来实现, 可变参数 $\alpha_i(Z)$, ($i = 0, 1$) 的估计通过 Matlab 文件 `cvm.m` 计算, 相关系数的估计通

过 `npcc.m` 计算. 我们对本文所提的可变参数分析新方法的性能做了模拟研究.

在模拟中, 考虑 $\alpha_1(Z)$ 具有不同的函数形式, 且 $\alpha_0(Z)$ 为一常数. 我们比较了当 $\alpha_1(Z)$ 为常数, 线性函数及其他两种特定函数时, 绘制了当 z 从 0 变化到 8, x 和 ε 为标准常态随机变量, 样本大小为 300 时, 模型 $y = \alpha_0(z) + \alpha_1(z)x + \varepsilon$ 的性能曲线. 对于每一个 $\alpha_1(z)$ 函数形式, 我们获得产生的数据集 $\{(x_i, y_i, z_i), i = 1, 2, \dots, n\}$, 然后利用前文介绍的方法计算 $\hat{\alpha}_1(z)$ 的估计值, 利用 ad-hoc 方法选取带宽 $h = 0.5$, 图 1 和图 2 中分别绘制了 $x-y$, $z-\hat{\alpha}_1(z)$ 和 $z-\alpha_1(z)$ 的散点图, 从图上可以看出, 当 $\alpha_1(z)$ 为常数时, y 和 x 是线性相关, 如图 1(a)所示, $\hat{\alpha}_1(z)$ 的估计值也近似为一常数, 如图 1(b)所示. 然而当 $\alpha_1(z)$ 不是一个常数时, 不管是线性或其他具体形式时, y 与 x 不再服从线性关系(图 1(c), 2(a)和 2(c)). 可任意看出, 本文所提的方法较好地给出 $\alpha_1(z)$ 的估计值 $\hat{\alpha}_1(z)$, 并充分拟合给定函数 $\alpha_1(z)$ (图 1(d), 2(b)和 2(d)).

这些模拟结果表明, 当真实的回归系数是某些其他因素的函数时, 使用常系数的线性回归模型并不合适. 而此时使用变系数模型进行分析却能较好地揭示潜在的相关关系.

2 经验数据分析示例

我们使用榕树-榕小蜂互利共生作用的例子来说明我们所推荐的方法. 榕树-榕小蜂的互利共生是被广为人知的种间合作系统之一. 榕树提供部分雌花资源以供榕小蜂幼虫的生长, 榕小蜂则为榕树的花传粉, 因而双方都从这种互惠关系中获得利益^[41,42]. 在以前的研究中, 榕树种子被用来衡量榕树的适合度, 而榕小蜂幼虫的数量(榕小蜂造成的瘿花)则被用来衡量榕小蜂的适合度^[43-45].

在合作演化的传统理论, 如亲选择理论(Hamilton定理)、互惠选择理论(多次重复囚徒困境博弈原理)或群选择理论中, 纳什均衡或进化稳定策略都存在于合作者之间^[46]. 相互合作者的适合度的相关性应该仅只存在正相关性^[11,12,32], 并且这种适合度的正相关性可以通过合作者的空间异质性或者自我抑制而得以维持^[47,48]. 如果在合作系统中存在一个稳定的均衡, 那么合作者的适合度间的相关系数应该是一个常数. 但是在榕树-榕小蜂共生关系中, 榕树与榕小蜂之间的适合度在是否是正相关这一问题上存在

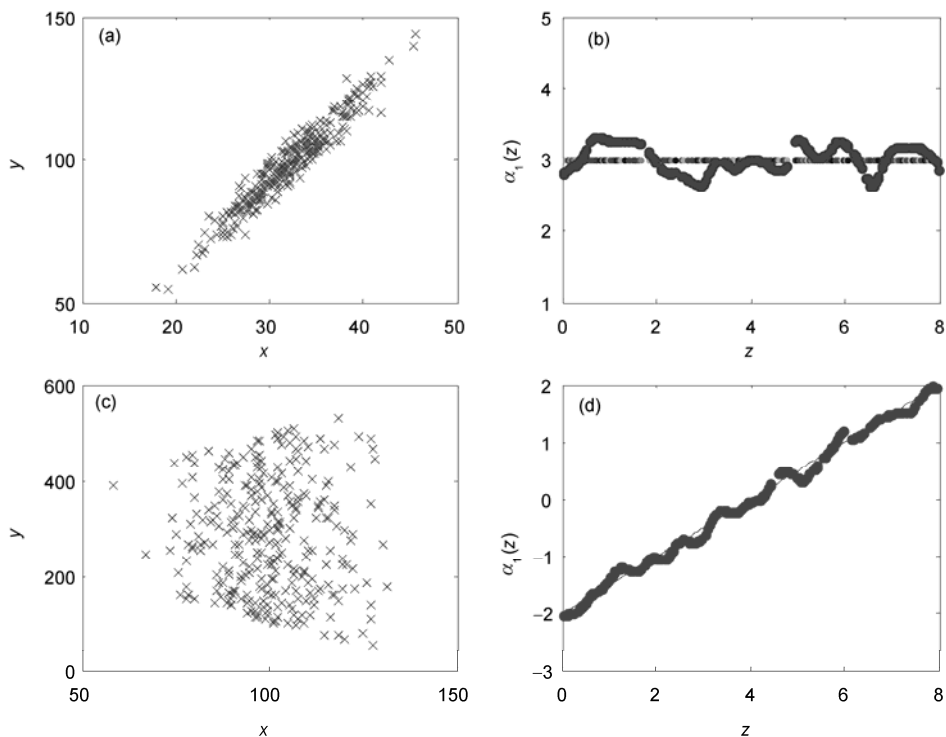


图1 变系数模型模拟示例图(常系数和直线系数模型)

(a) 利用模型 $y = 2 + \alpha_1(z)x + \varepsilon$ 模拟得到 y 和 x 的散点图. 这里的 $\alpha_1(z) = 3$, $n = 300$, x 是一个均值为 32、标准差为 4 的正态随机变量, z 是一个 0~8 的随机数, ε 是一个均值为 0、标准差为 1 的正态随机误差, 这个模型具有一个不变斜率. (b) 基于(a)中数据, $\hat{\alpha}_1(z)$ 与 Z 和 $\alpha_1(z)$ 与 Z 的散点图, 从图上可以看出 $\hat{\alpha}_1(z)$ 近似为一个常数. (c) 当模型为 $y = 300 + \alpha_1(z)x + \varepsilon$ 时 x 和 y 的关系散点图, 这里 $\alpha_1(z) = 0.5z - 2$, $n = 300$, x 是一个均值为 100、标准差为 12.5 的正态随机变量, z 是一个 0~8 的随机数, ε 是一个均值为 0、标准差为 0.2 的正态随机误差, 这个模型包含 $\alpha_1(z)$ 的线性函数. (d) 基于(b)中数据, $\hat{\alpha}_1(z)$ 与 Z 和 $\alpha_1(z)$ 与 Z 的散点图, 从图上可以看出估计值 $\hat{\alpha}_1(z)$ 充分接近于实函数 $\alpha_1(z)$

巨大的分歧^[43,45,49,50]. 我们的分析与实验数据表明, 这些适合度相关性在文献中相互矛盾的结果可能是由于不合理甚至是错误使用简单或多元线性回归所致, 因为即使在同种榕树-榕小蜂系统中, 其适合度相关系数在不同环境下也有较大变化^[27,31,44,45].

我们使用最常用的方法来分析种子与小蜂幼虫之间的相关系数, 以此描述榕树与榕小蜂之间的关系. 我们在不同年份从 4 个样地随机收集数据, 共收集了 225 个样本进行观测. 种子数与小蜂幼虫数的分布信息中均有很多扰动. 但是这些扰动可能是由于混沌效应造成的, 因为在这样的环境中很难区分随机扰动与混沌^[51,52].

我们首先利用与 Sugihara 和 May^[51]相似的图形化的评价方法来检验线性回归分析是否能够描述种子数与小蜂幼虫数之间的关系. 图 3(a)和(b)清楚地显示了图 1(a)中模拟数据的一般线性关系. 图 3(c)和

(d)展示了榕树-榕小蜂共生系统(聚果榕)的真实数据, 与图 3(a)与(b)的结果相比较可以看出: 在此使用常系数线性回归分析并不合适. 使用真实数据拟合的线性回归模型的 F 检验的 P 值为 0.501, 很明显这是不显著的.

使用本文中给出的变系数分析方法, 结果显示, 榕树适合度与榕小蜂适合度的相关系数随着榕小蜂的适合度与资源(空雌花)的变化而不断变化(图 4). 这表明当有大量空雌花存在或者被产卵的花(瘦花)数量很少时, 产卵小蜂更多的产卵能增加榕树种子数量. 这可能是因为母蜂产卵的同时也是在帮榕树传粉^[27,43]. 那么当有很多未被产卵的花时, 种子数与小蜂幼虫数能够同时增加. 在这种情况下, 榕树种子形成与小蜂幼虫的存在之间就呈现出正相关关系. 但是, 当大多数雌花都完全被传粉或者被产卵时, 如果花不能阻止产卵小蜂在被授粉雌花中产卵或者产

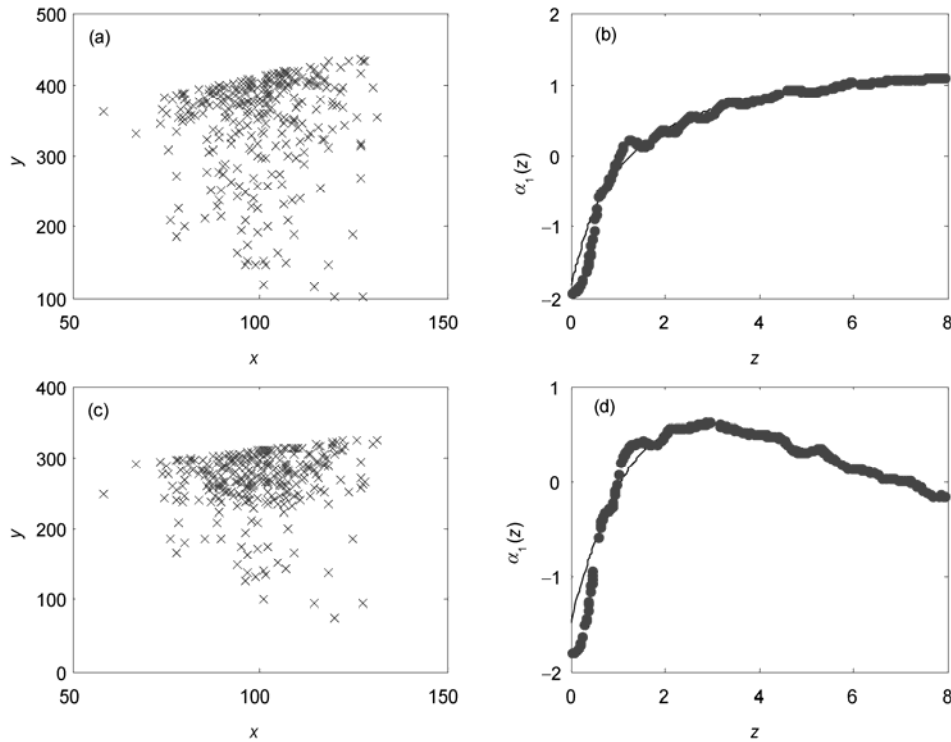


图 2 变系数模型模拟示例图(曲线系数模型)

(a) 将数据按模型 $y = 300 + \alpha_1(z)x + \varepsilon$ 模拟所得的 y 关于 x 的散点图, 其中 $\alpha_1(z) = (1.5z - 1.9)/(1+z)$, $n = 300$, x 是均值为 100、标准差为 12.5 的正态随机变量, z 是一个 0~8 的随机数, ε 是均值为 0、标准差为 0.2 的正态随机误差. (b) (a) 中的 $\hat{\alpha}_1(z)$ 与 Z 和 $\alpha_1(z)$ 与 Z 的散点图. 可以看到估计值 $\hat{\alpha}_1(z)$ 较好地近似 $\alpha_1(z)$. (c) 按模型 $y = 250 + \alpha_1(z)x + \varepsilon$ 模拟所得的 y 关于 x 的散点图. 其中, $\alpha_1(z) = (0.5 + 1.2z + 0.1z^2) - 2$, $n = 300$, x , z 以及 ε 的含义与 (a) 中相同. (d) 给出了 (c) 中的 $\hat{\alpha}_1(z)$ 与 Z 和 $\alpha_1(z)$ 与 Z 的散点图. 通过 (b) 可以看出估计值 $\hat{\alpha}_1(z)$ 较好地近似真实函数值 $\alpha_1(z)$

卵小蜂不能自我抑制对雌花这一资源的过度利用, 那么更多的产卵就是在消耗种子. 由于以前的定量测量表明花的构造无法有效阻止产卵小蜂在花中产卵^[27,52,53], 那么当所有雌花都被使用或者被产卵的雌花的数量很高时, 榕树种子的形成与小蜂幼虫的存在之间就会呈现出负相关关系.

榕树种子和榕小蜂的后代数受传粉小蜂的产卵和授粉效率、产卵小蜂数目以及雌花的丰度等多种因素影响, 而这些因素又受栖息地、季节, 甚至月份的影响. 因而种子和榕小蜂的后代数之间的相关系数随着很多因素的变化而变化^[52], 这就能完整地解释为何适合度相关性由于不同雌花丰度^[27]、不同的资源分配^[44]、不同的季节^[31]、不同的收成月份^[45,52]而表现出极大的不同. 我们的榕树-榕小蜂系统实验数据展示了非对称性相互作用将可导致其关系系数的不确定性. 比如主要存在于榕小蜂之间干扰竞争^[36], 榕果对不合作榕小蜂的非对称惩罚^[54-56], 这

些非对称性相互作用导致了空间受限异质性, 因而导致一个不确定性的相互作用. 这些结果显示, 合作双方的相互关系将随着其适合度或与公共资源的变化而产生混沌性扰动. 其关系系数并没有出现亲缘选择、互惠选择(囚徒困境)和群选择的均衡理论中所预期的常数关系. 这样一个相关变化在之前的研究中没有被发现, 原因可能是过去研究人员对线性回归或广义线性回归的的误用或滥用所致^[27].

Nowak^[57]指出, 严格均衡或者稳定进化策略在生物系统中可能难以实现, 这是因为如果系统存在均衡状态, 任何一个新的突变都难以侵入原有均衡系统, 进化将会因此而难以发生. 理论上说, 空间混沌更有可能^[47,52,58,59], 而混沌是否真正存在于生命系统一直是种间合作、寄主-猎食关系^[24,28]以及种群生物学^[60]极富争论的问题. 使用变系数方法, 物种相关随其他因素变化而变化的方式能够充分地我们的数据中被描述出来(图 4, 附录 A), 提供了适合

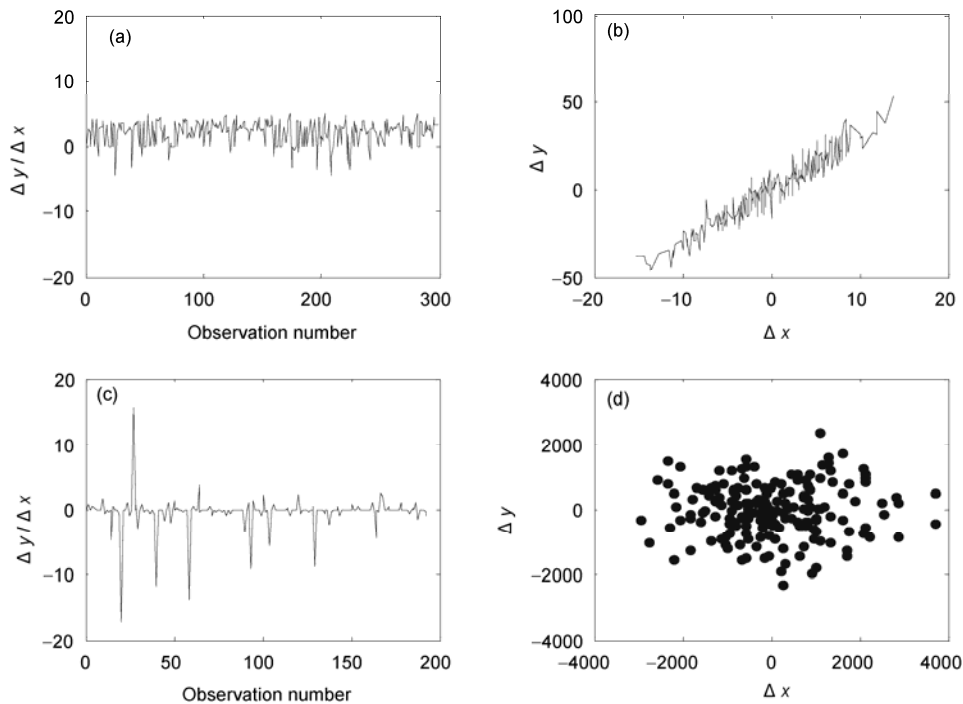


图3 模拟数据与榕树-榕小蜂共生系统真实数据的比较

(a) 展示了图1(a)中模拟数据的各观测的 $\Delta y/\Delta x$ 值. 通过图(a)可以看出构建 y 对 x 的常系数线性回归模型是合适的, 因为 $\Delta y/\Delta x$ 近似于一个常数. (b) 是模拟数据的 Δy 对 Δx 的线图, 图中表现出清晰的线性关系. (c) 展示了榕树-榕小蜂数据的各观测的 $\Delta y/\Delta x$ 值. $\Delta y/\Delta x$ 的值不稳定, 这表明构建 y 对 x 的常系数线性回归模型是不合适的. (d) 是榕树-榕小蜂数据中 Δy 对 Δx 的散点图, 从中无法看出明显的线性关系. 在(c)与(d)中, y 表示榕树种子数, x 表示小蜂幼虫数

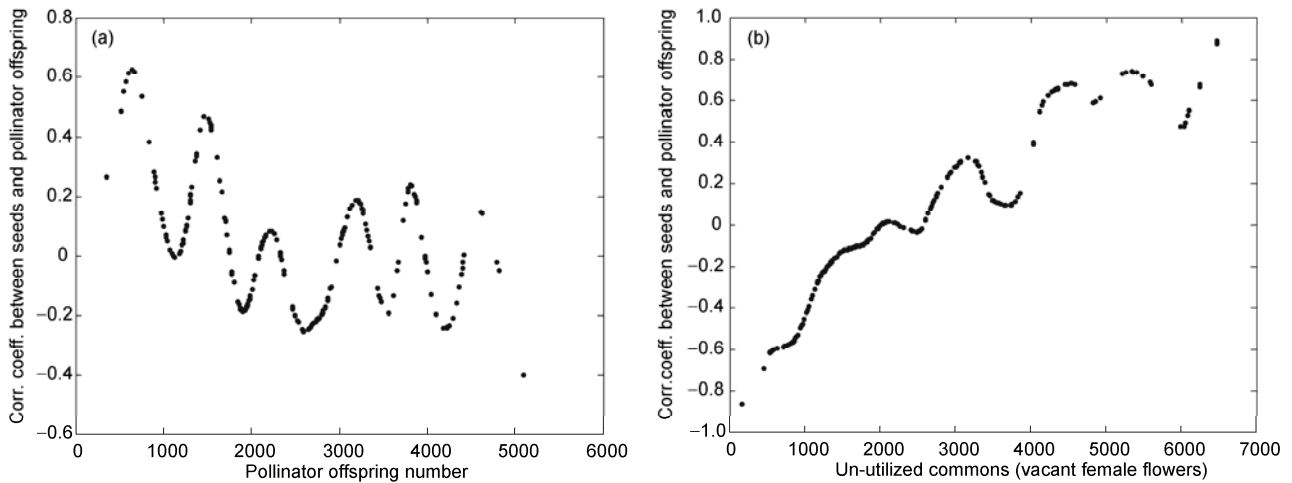


图4 用变系数方法对榕树与传粉小蜂相互关系的分析

图中显示了相关系数随小蜂后代数量的变化震荡. 当共同可用资源充足时, 呈现正相关关系, 而当共同可用资源有限时, 则呈现负相关状态. 实验显示, 互惠共生关系取决于可用资源水平, 参看 Wang 等人^[27]. $\alpha_0(z)$ 的平滑曲线核函数为二次函数, 带宽为 500, 样本量为 225, 并且样本是从 4 个产地、不同年份采集的

度相关变化的直接证据。

附录 A 给出了应用该方法对共存和敌对物种的经验数据分析结果。

3 讨论

理论预测和经验数据显示在合作(互惠)、敌对和共存物种系统中, 相同物种的正向、负向和中性相关性都可能被观测到。物种相关可能在不同环境、不同生态条件, 或者不同的适合度或种群密度下会有很大不同。在物种间没有稳定关系的情况下, 线性或者非线性回归分析是不能用来分析这样的物种相关性的。而变系数分析方法, 没有预先指定物种定量特征的响应函数, 因此是恰当的。通过使用这种方法, 生物学特征的相关系数随着其他因素变化的方式能够在我们的经验数据分析中被描述出来。

当相关系数是一个固定值时, 变系数线性回归模型在物种相互作用的分析中就转化为一个参数分析方法(图 1(a)和(b))。这意味当物种间相互作用的基本模型结构无任何先验信息的情况下, 这一新的方法比任何现存的参数分析更普遍, 更可靠。模拟数据的结果(图 1(c)和(d), 图 2(a)和(b))显示, 指定的响应函数是难以估计的, 而代之以非参数估计是必要的。

在模型中, 第 3 个变量 Z 的影响有时可以是物种相互作用特有的参数变量, 因此我们可以用 X 或者 Y 替代 Z 描述物种相互作用的相关系数, 物种相互作用相关系数是根据物种相互作用的特性决定的。用 X 或者 Y 代替 Z 在下面两点实践中能使得事情简化: (1) 决定物种相互合作相关系数变化的关键因素无法容易地识别, 只有变量 X 和 Y 在实践中是可用的; (2) 如果变量 Z 的系数 α_1 是繁杂的, 或者不能很好被描述, 使用 X 或者 Y 变量替代 Z 作为 X 轴去整合 Z 因素的影响。用 X 或者 Y 替代 Z 的技术细节在附录 B 中显示是可行的。

在传统的统计分析中, 其他因素的效应通常被看作是一个协变量, 此时偏回归系数分析只能被用于控制因素是物种相互作用的线性函数^[1,7,12,32], 这种方法可以帮助我们估计物种相互作用的相关系数。然而, 如果这些变量的影响不是线性的, 回归系数就可能是其他影响变量的函数, 这些变量就不能作为参数回归分析中的协变量。当这些变量是可控的时候, 参数回归分析中简单地将影响变量作为协变量处理只能得到系数的均值估计, 这个均值是不太可能准确地描述物种间相互作用。如果这些变量显著地影响相关系数, 这种方法可能导致一个不正确的结论^[27]。

致谢 感谢朱力平博士和匿名审稿人对本文提出的评论和修订建议。

参考文献

- 1 李春喜, 王志和, 王文林, 著. 生物统计学. 北京: 科学出版社, 2000
- 2 McCullagh P, Nelder J A. Generalized Linear Models. 2nd ed. London: Chapman and Hall, 1989
- 3 Krebs J K. Ecology: The Experimental Analysis of Distribution and Abundance. 2nd ed. New York: Harper and Row Press, 1978
- 4 May R M. Simple mathematical models with very complicated dynamics. Nature, 1976, 261: 459-467
- 5 孙儒泳. 动物生态学原理. 北京: 北京师范大学出版社, 1992
- 6 马知恩. 种群生态学的数学建模与研究. 合肥: 安徽教育出版社, 1996
- 7 Frank S A. The price equation. Fisher's fundamental theorem, kin selection and causal analysis. Evolution, 1997, 51: 1712-1729
- 8 Molles M C. Ecology: Concepts and Applications. New York: McGraw-Hill Companies, 1999
- 9 Whittingham M J, Stephen P A, Bradbury R B, et al. Why do we still use stepwise modeling in ecology and behaviour? J Anim Ecol, 2006, 75: 1182-1189
- 10 Fisher R A. The Genetical Theory of Natural Selection. Clarendon: Oxford, 1930
- 11 Price G R. Selection and covariance. Nature, 1970, 227: 520-521
- 12 Queller D C. A general model for kin selection. Evolution, 1992, 46: 376-380
- 13 Stiling P D. Ecology: Theories and Applications. New Jersey: Prentice Hall, Upper Saddle River, 1996
- 14 Draper N R, Smith H. Applied Regression Analysis. New York: John Wiley and Sons, 1981
- 15 Raubenheimer D. Problems with ratio analysis in nutritional studies. Functional Ecol, 1995, 9: 21-29
- 16 Inouye B D. Response of surface experiment design for investigating interspecific competition. Ecology, 2001, 82: 2696-2706

- 17 Travis J, Trexler J C. Interactions among factors affecting growth, development and survival in experimental populations of *Bufo terrestris* (Anura: Bufonidae). *Oecologia* (Berlin), 1986, 69: 110–116
- 18 Trexler J C, McCulloch C E, Travis J. How can the functional response best be determined? *Oecologia*, 1988, 76: 206–214
- 19 Holling C S. Resilience and stability of ecological systems. *Ann Rev Ecol Syst*, 1973, 4: 1–23
- 20 May R M. *Theoretical Ecology: Principles and Applications*. Oxford: Blackwell Scientific Publishers, 1981
- 21 Nowak A M, Bonhoeffer S, May R. Spatial games and the maintenance of cooperation. *Proc Natl Acad Sci USA*, 1994, 91: 4877–4881
- 22 Jiang L, Morin P J. Temperature fluctuation facilitates coexistence of competing species in experimental microbial communities. *J Anim Ecol*, 2007, 76: 660–668
- 23 Hofbauer J, Sigmund K. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press, 1998
- 24 Hanski I, Turchin P, Korpimäki E, et al. Population oscillations of boreal rodents: Regulation by mustelid predators lead to chaos. *Nature*, 1993, 364: 232–235
- 25 Ives A R. Predicting the response of populations to environmental change. *Ecology*, 1995, 76: 926–941
- 26 Gonzalez A, Descamps-Julien B. Population and community variability in randomly fluctuating environments. *Oikos*, 2004, 106: 105–116
- 27 Wang R W, Shi L, Ai S M, et al. Trade-off between the reciprocal mutualists: Local resource availability oriented interaction in fig/fig wasp mutualism. *J Anim Ecol*, 2008, 77: 616–623
- 28 Hassell M P, Comins H N, May R M. Spatial structure and chaos in insect population dynamics. *Nature*, 1991, 353: 255–258
- 29 Thompson J N, Fernandez C C. Temporal dynamics of antagonism and mutualism in a geographically variable plant-insect interaction. *Ecology*, 2006, 87: 103–112
- 30 Smith J M, White A, Sherratt A J, et al. Disease effects on reproduction can cause population cycles in seasonal environments. *J Anim Ecol*, 2008, 77: 378–389
- 31 Wang R W, Yang J X, Yang D R. Seasonal changes in the trade-off among the fig-supported wasps and viable seeds in figs and their evolutionary implications. *J Integr Plant Biol*, 2005, 47: 144–155
- 32 Frank S A. Genetics of mutualism: The evolution of altruism between species. *J Theor Biol*, 1994, 170: 393–400
- 33 Sun B F, Wang R W, Hu Z, et al. The relation between two non-pollinating wasps oviposition and the fruit abscission on *Ficus racemosa*. *Acta Ecol Sin*, 2009, 29: 1–6
- 34 Rudolf V H W. Consequences of size structure in the prey for predator-prey dynamics: The composite functional response. *J Anim Ecol*, 2008, 77: 520–528
- 35 Wang R W, Sun B F. The seasonal change in the structure of fig-wasp community of figs and its implication for conservation. *Symbiosis*, 2009, 47: 77–83
- 36 Wang R W, Ridley J, Sun B F, et al. Interference competition and high temperatures reduce the virulence of fig wasps and stabilize a fig-wasp mutualism. *PLoS ONE*, 2009, 4: e7802
- 37 Hastie T, Tibshirani R. Varying-coefficient models (with discussion). *J Roy Stat Soc B*, 1993, 55: 757–796
- 38 Epanechnikov V. Nonparametric estimates of a multivariate probability density. *Theor Prob Appl*, 1969, 14: 153–158
- 39 Härdle W. *Applied Nonparametric Regression*. Berlin: Springer, 1994
- 40 Zhu L X. *Nonparametric Monte Carlo Tests and Their Applications*. New York: Springer, 2005
- 41 Janzen D H. How to be a figure. *Ann Rev Ecol Syst*, 1979, 10: 13–51
- 42 Wiebes J T. Co-evolution of figs and their insect pollinators. *Ann Rev Ecol Syst*, 1979, 10: 1–12
- 43 Bronstein J L, Hossaert-McKey M. Variation in reproductive success within a subtropical fig-pollinator mutualism. *J Evol Biol*, 1996, 23: 433–446
- 44 Anstett M C, Bronstein J L, Hossaert-McKey M. Resource allocation: A conflict in the fig-fig wasp mutualism. *J Evol Biol*, 1996, 9: 417–428
- 45 Herre E A, West S A. Conflict of interest in a mutualism: Documenting the elusive fig wasp-seed trade-off. *Proc Roy Soc B*, 1997, 264: 1501–1507
- 46 Maynard S J. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982
- 47 Nowak A M, May M R. Evolutionary games and spatial chaos. *Nature*, 1992, 359: 826–829
- 48 Doebeli M, Knowlton N. The evolution of interspecific mutualisms. *Proc Natl Acad Sci USA*, 1998, 95: 8676–8680
- 49 Bronstein J L. Seed predator as mutualists: Ecology and evolution of the fig-pollinator interaction. In: Bernays E, ed. *Insect-Plant Interaction*. Boca Raton: CRC Press, 1992
- 50 Herre E A, Knowlton N, Mueller U G, et al. The evolution of mutualisms: Exploring the paths between conflict and cooperation. *Trend Ecol Evol*, 1999, 14: 49–53

-
- 51 Sugihara G, May R M. Non-linear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 1990, 344: 734–741
- 52 Wang R W, Sun B F, Zheng Q, et al. Asymmetric interaction and indeterminate fitness correlation between cooperative partners in the fig/fig wasp mutualism. *J Roy Soc Interface*, 2011, doi: 10.1098/rsif.2011.0063
- 53 Nefdt R J C, Compton S G. Regulation of seed and pollinator production in the fig-fig wasp mutualism. *J Anim Ecol*, 1996, 65: 170–182
- 54 Wang R W, Sun B F, Zheng Q. Diffusive co-evolution and mutualism maintenance mechanisms in a fig-fig wasp system. *Ecology*, 2010, 91: 1308–1316
- 55 Wang R W, Zheng Q. Structure of a figs wasp community: Temporal segregation of oviposition and larval diets. *Symbiosis*, 2008, 45: 113–116
- 56 Wang R W, Sun B F. The seasonal change in the structure of fig-wasp community of figs and its implication for conservation. *Symbiosis*, 2009, 47: 77–83
- 57 Nowak A M. An evolutionary stable strategy may be inaccessible. *J Theor Biol*, 1990, 142: 237–241
- 58 Wang R W, Shi L. The evolution of cooperation in asymmetric systems. *Sci China Life Sci*, 2010, 53: 139–149
- 59 Wang R W, He J Z, Wang Y Q, et al. Asymmetric interaction will facilitate the evolution of cooperation. *Sci China Life Sci*, 2010, 52: 1041–1046
- 60 Tilman D, Wedin D. Oscillation and chaos in dynamics of a perennial grass. *Nature*, 1991, 353: 653–655
-

补充材料

附录 A 对共存和敌对物种的经验数据分析

附录 B 关于因素 Z 的例子

本文的以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

附录 A 对共存和敌对物种的经验数据分析

分析有相同的食性但是产卵顺序分离的共存物种的数据[58], 结果同时也显示共存物种间的相关系数也常常是密度依赖性的, 依赖于自己种群的种群密度(图 5a). 而分析敌对种群的数据, 结果显示寄主和宿主间的相关系数同样有密度依赖性, 通常依赖于宿主种群的种群密度(图 5b). 结果也指出在共存和敌对物种间没有稳定的平衡(固定的相关系数), 与 Hassell 等(1991)提出的纯理论预测相一致. 他们指出寄生-宿主种群间的相互作用应该是波动的混沌状. 我们的分析结果表明: 过去有关共存或者竞争物种具有固定的相关系数分析结果可能源于不合理的分析方法所致.

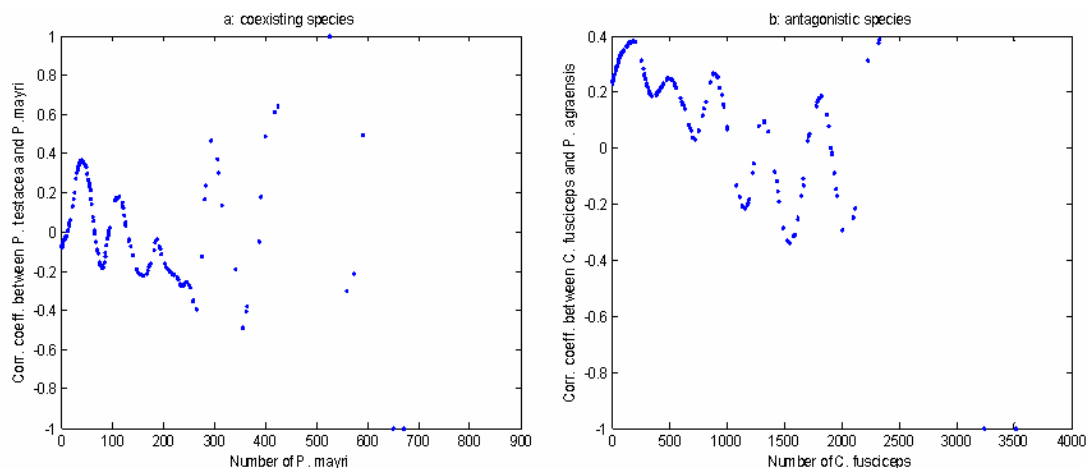


图 1 共存和敌对物种间相互作用的相关系数. 结果显示共存和敌对物种的相关系数并不是一个固定值, 而是作为与其相关物种密度的一个混沌震荡

(a) 造瘿蜂 *platyneura testacea* 和 *P.mayri* 间的相关系数是关于 *P.mayri* 数量的一个函数. 两种小蜂都是榕树造瘿者, 他们利用相同的雌花资源^[59]. 其产卵顺序是分离的但有重叠^[58]; (b) 寄主 *P.agraensis* 和宿主 *Ceratosolen fusciceps* 的相关系数是关于 *Ceratosolen fusciceps* 数量的函数. $\alpha_0(z)$ 的光滑曲线是一个二次函数, 宽窗 h 是 300. 样本大小为 178, 所有的样本收集花了一年时间

附录 B 关于因素 Z 的例子

正如我们在引言中提到的那样, 我们想获得关于因素 z 的连续相关系数函数, 以便于得到更全面的分析. 当 Z 作为 X 和 Y 的独立观察量时, 在我们的数据分析中, 我们能够用传统的统计方法对数据进行准确的分析. 但是, 有时 Z 取决于 X . 下面我们就用一个关于 Z 的例子说明我们的分析方法. 对于任意的数据 X_1, \dots, X_n , 我们能够按照从最小到最大升序排列, 记为 $X_{(1)}, \dots, X_{(n)}$. 对于任意的 l , 我们考虑 n 个段, 每个段都像如下所述的那样有 l 个数值. 对于任何的 i , 使得该段有 l 个以 $X_{(i)}$ 为中心点, 也就是说, 如果 l 是偶数那么数据点是 $\{X_{(i-l/2)}, \dots, X_{(i+l/2)}\}$, 如果 l 是奇数则为 $\{X_{(i-(l-1)/2)}, \dots, X_{(i+(l-1)/2)}\}$. 当 i 等于或小于 $l/2$ (or $(l-1)/2$), 我们不能有下标是 $i-l/2$ 的数据点, 因此, 我们根据 i 定义集合 $\{X_{(1)}, \dots, X_{(i+l/2)}\}$; 当 i 等于或大于 $n-l/2$ (或者是 $n - (l-1)/2$), 我们定义集合为 $\{X_{(i)}, \dots, X_{(n)}\}$. 因此, 我们定义 n 个段, 我们能够看到这些段是连续的, 而且只要 j 小于 $i+j$ (当 $j > i$) $\{X_{(i)}, \dots, X_{(i+l)}\}$ 和 $\{X_{(j)}, \dots, X_{(j+l)}\}$ 就有重叠. 然后我们定义 Z_k 作为段 $\{X_{(k)}, \dots, X_{(k+l)}\}$ 的示性值, $K=1 \dots n$. 例如, 我们令 Z_k 等于段 $\{X_{(k)}, \dots, X_{(k+l)}\}$ 的所有值的平均值. 这意味着, 对于任意 X 的段, Z 是常数, 但是对于不同的段取不同的值. 因此, 对于给定的 Z_i , 我们有 $\{X_i, \dots, X_{i+l}\}$ 和相应的 $\{Y_i, \dots, Y_{i+l}\}$, 然后我们能够计算各段的相关系数. 我们也能够依次计算给定 Z 时的条件相关系数.