

中性场理论 第六章： 物种的空间分布谱



集智热力学小组读书会

报告人：东方和尚

二〇一〇年春

今天贫僧向大家报告的内容是 Hubbell 《中性场理论》的第六章，也就是物种的空间分布情况。

这一章的内容基本上独立成篇，除了一处以外，没有用到之前王钺公讲到的数学推导，其结论也是相对比较简单而符合直觉的，用到的数学也不多。因此诸位即使没有从头开始听，也能够很好的理解本章所述的道理，进而掌握整个中性场理论的核心意义。一路上遇到的所有符号、公式，贫僧也会略作讲解，以便大家能够理解。

生物学上的一个现象，可以有两种不同的解释方法：一是用一个数学模型来解释，得到了同实际情况吻合的分布曲线，我们就认为现象得到了解释；二是不用数学模型，而是用白话来解释（如“羊多了，把草吃完了”）；这两种解释都是必要的，因为数学模型并不是客观实际，生物学也远没有到可以数学化的阶段。一个数学模型也可以有许多种不同的方法来解读；这一章用到的数学虽然不多，但是对于同一个三段曲线，却存在许多种解释方法，各有各的道理，也各有适用范围，希望数学好的朋友，也不要因此忽视对于实际的分析。

根据这个考虑，同王钺公上次只讲数学的做法不同，贫僧会基本按照书上的顺序来讲。这样可以让我们对生态学有比较直观的认识，同时也可以学习 Hubbell 的谋篇布局。看似简单的一个道理，要把很多事实以及研究背景串在一起，形成有分量的综述，举轻若重，其实也是不容易的！

以上是开场白，下面进入正题：

一、物种数量和区域面积的关系

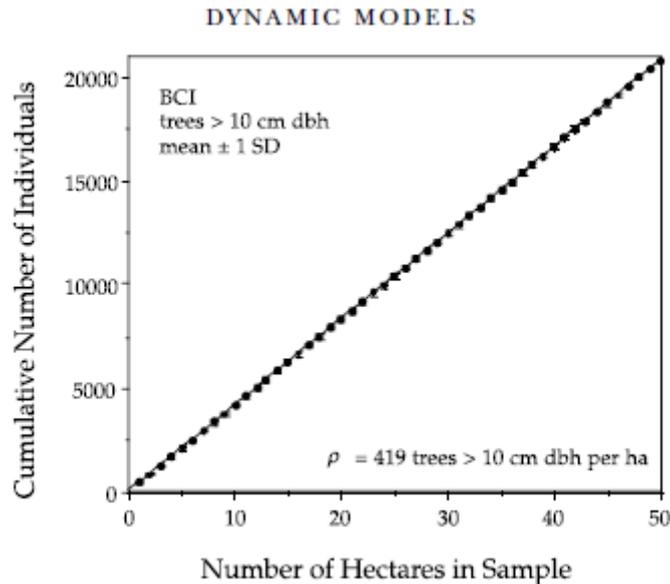
这个问题看上去很简单，就是说你在地上画一个圈，然后开始数圈里面的生物物种数。我们今天要讨论的问题，无非就是物种数和圈子大小，有什么样的关系。

这里有几个事情需要注意的，先说清楚：

- 1、我们这里讨论的是物种数，而不是生物总数。所谓物种数，并不是包罗万象什么都算的，而只是指某一个营养级别下的数量，就是相互有竞争关系的物种。你研究的树，那么就只算树的数量，其他的比如草、动物、细菌等等就不管。
- 2、本章有三个基本参数：面积（A）、物种数（S）和个体总数（I）。比如某一公园占地 1 平方公里，有 100 颗树，由 30 颗柳树、20 颗杨树和 50 颗桂花树这三种不同的树组成。那么 $A=1$ ， $S=3$ ， $I=100$ 。这点请大家牢记。
- 3、在田野调查的时候，总是对一个区域进行调查，而得到的最基本结论，一定也是我发现了多少种物种（很长一段时间，生物学家的工作就是去寻找新的生物），那么调查的面积和物种数量之间的关系，必然也成为了生态学最基本的数学关系之一。但是到目前为止，我们还远没有弄清此问题。

这个问题看上去很简单，就是考察的面积越大，遇到的生物个体越多，生物物种数也越多。

生物的个体数量 I 是和面积 A 成正比关系的（如图一所示，你可以认为 $A=I$ ），因为任何一片土地，平均来说，能够承载的生物量都类似。另一方面，由于同一个物种会有很多个体，所以物种数 S 就一定小于 I ，即随着测量面积 A 的增加， S 增加的速度会小于 A 的增加的速度。问题就是小多少？



图一、生物数量和地理面积呈线性关系

两个变量的缩放比例关系，称为 **Scaling**，比如长度增加到 2 倍，面积增加到 4 倍，体积增加到 8 倍，这就是一个典型 **Scaling** 的关系，现在我们的目的，进一步的细化到了： S 和 A （或者说和 I ，因为 $A=I$ ）遵循何种比例关系？

在历史上，这个问题的研究比《物种起源》更早。不同生物学家根据自己的调查结果，提出了花样繁多的猜想，也充满着争论。原因就是物种数和区域大小之间的关系有点复杂，依赖于取样的范围，并非遵循简单的函数关系。

值得注意的是，如果对于多个点进行取样，那么就可以平均掉区域上的差异，而得到 S 和 I 的关系，这时候问题就回到了前几章的讨论（前几章实则就是研究 S 和 I 的关系），发现 S 和 A 遵循生物多度关系，也就是对数序列关系，即对于给定的相对物种频率 x ，一个基本的物种相对多度分布模型可以用下式来表示：

$$\Phi(x) = \frac{\theta(1-x)^{\theta-1}}{x}$$

式中 θ 为 Hubbell 的基本生物多样性参数。

二、幂律关系与幂律分布

所谓**幂律关系**，最简单的解释就是长度、面积、体积的关系：一个物体，就说一个立方体吧，如果长度增加到 2 倍，则面积增加到 4 倍、体积增加到 8 倍。这就是幂律关系，它的维度就是幂律的指数。一般地说，如果我们遇到两个参数 A、B 具有某种关系，当 B 增加到某个倍数 b 的时候，A 也增加到某个倍数 a，那么我们就说 A、B 之间有幂律关系，而幂指数就是 $\log_b a$ 。

举例，假设我们发现公园里树的种类和公园面积呈以下关系：当公园面积增加到原来的 10 倍的时候，树的种类会正好多 1 倍，那么这种情况下，物种数同面积就遵循幂指数为 $\log_{10} 2 = 0.3$ 的幂律关系。

幂律关系是本书的基本关系，也是复杂性科学的最基本现象之一，请大家务必彻底搞懂。如果写出来的话就是：

$$\text{物种数 } S = cA^z \quad (\text{A 是面积, } c \text{ 是比例常数, } z \text{ 就是幂律})$$

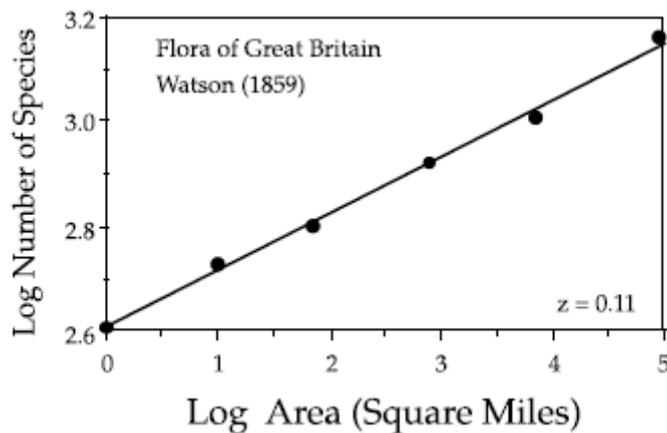
这个关系式子也常常写成 $S = \alpha A^\beta$ 用以表示三种不同的多样性：

- **Alpha 多样性**就是等式前的乘数，它表示物种的绝对多样性，是指在特定地区、族群、生态系内的丰富度，通常以种为单位，刻画生态系内种类的绝对数量。
- **Beta 多样性**就是幂律大小所表示的相对多样性，也即不同生态系统之间多样性的比较，它可以代表物种组成沿环境梯度或者在群落间的变化率，也就是生物种类对环境异质性的反应程度。
- **Gamma 多样性**指大区域（地理区域）上总的多样性，包括一个地区内不同生态系的全部多样性的量测值。

请读者注意，生物学和社会科学上面往往用到的数学很简单，但不要因此觉得这个现象的内涵也会简单。所谓重剑无锋，大巧不工。单单一个 **Beta 多样性**，就可以写很多本书，都不足以形容大自然的丰富多彩，而用一个幂指数，就可以把如此变幻多姿的现象尽皆收揽无遗。你觉得自己真的懂得这个公式了吗？幂律究竟是什么意思，现在人类所知的还很少很少。

另外有一个常常会与幂律关系混淆的概念是幂律分布，这里也稍微提一点，所谓幂律分布，就是有这么一堆个体，用一个变量刻画，然后你按照这个变量，把它的分布画出来。两者的最大差别就是，**幂律关系是两个变量，幂律分布是一个变量！**比如我们现在说的物种数和面积、生物个数成幂律关系，但是物种数本身是什么分布呢？那就不是幂律分布，而是比较复杂的 Z 形曲线。前几章讨论的就是这个问题。这个问题虽然很简单，却很搞脑子，希望大家画画图，仔细吃透。

幂律现象是非常常见，也是非常有趣的现象。到目前为止，人类已经发现了数以千万计的幂律分布和幂律关系。很多很多现象，只要你统计一下，在双对数坐标（x、y 轴的单位都是 1、10、100、1000 这样的坐标纸叫做双对数坐标）上面一画，就会是很漂亮的一条直线，（如图二）。但是为什么幂律这么常见呢？是否在幂律分布的背后存在着类似于大数定律和中心极限定律这样的本质原因呢？目前人类还没有答案，希望本俱乐部的成员能对此问题作出贡献。（Jake 大人和计算士都是幂律方面的专家，在集智网站上也有大量的幂律科普和专门文献，大家可以自己去挖掘）



图二、生物种数和地理面积呈幂律关系

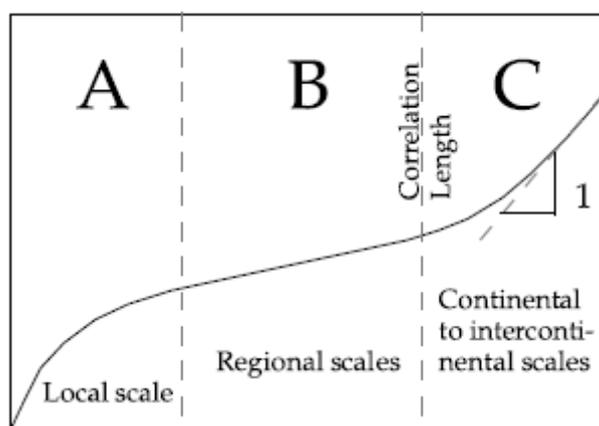
幂律分布与幂律关系存在着深刻的联系。具体来说，给定一条 rank 曲线（参见[第二章总结的补充部分](#)），那么我们知道这条曲线与横坐标的交点横坐标我们记作 A ，这条曲线与横轴围成的面积记为 X ，在很多实际的系统中 X 会与 A 呈现出幂律关系。比如，我们考察一个岛屿上物种的多度分布曲线，把它画成 Rank 形式，这条曲线的横坐标表示对所有物种的排序，那么曲线与横轴的交点坐标就是岛屿上所有的物种总数 S 。另外，曲线的每一个纵坐标表示的是给定某一个物种的个体数，所以你把这条曲线求积分得到的就是所有物种个体总数的和 I 。当 Rank 曲线满足一定条件的时候 S 就会与 I 呈现幂律关系。这种分布曲线和幂律关系曲线的关系在姓氏分布研究（参考此文：[姓氏分布](#)）、语言中的词语分布研究都有所讨论（参考此文：[Zipf 律导致 Heap 律](#)）。

三、物种的三段式分布

假如你在北大校园里开始调查，碰到一个人就把他归为一类，遇到不能归类的则另立一类，那么随着你调查面积的不断扩大，这个类别数会经历什么样的变化呢？

很容易理解：一开始的时候，你遇到的都是不一样的人，包括不同年龄，各个学科的师生，因此类别是增加的很快的；过了一段时间，各个专业都调查的差不多了，很难遇到新鲜的情况了，类别的增长就开始放缓；最后，你的调查越过了北大围墙，到了中关村和圆明园，那么又会遇到大量新的类别，工人、白领、掮客等等，因此类别的增长又会加速。总之，这里有一个快-慢-快的三段式增长关系。

物种数量同地理面积也是遵循这样的关系：



图三：物种数和地理面积的三段式幂律关系

下面的工作就是要把这样的直觉推理尽可能地数学化，得到上述的分布曲线。

我们知道，根据中性场理论的假设，生物是随机繁殖、死亡、变异的。新物种会不断地取代老物种的位置。另一方面，如果不考虑地理上的扩散因素，物种的分布并不是幂律分布，而是对数序列分布。那么根据前面讨论的分布曲线与幂律关系之间的关系，我们就能够得到物种数与个体数之间的关系：

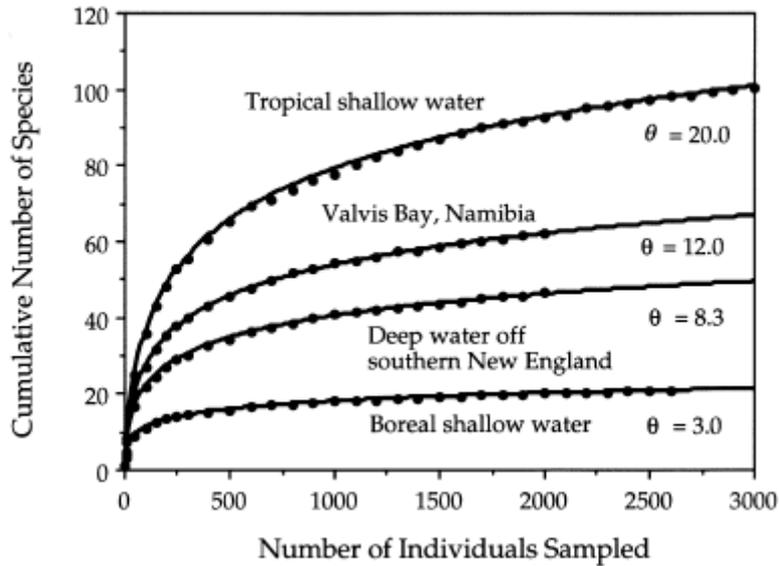
$$S(\alpha) = \alpha \cdot \ln \left(1 + \frac{J_M}{\alpha} \right).$$

有意思的是，这条曲线正好可以用来近似幂律折线的第一个折点（也就是图三种的 A 和 B 区域的曲线）。其中 α 就是 θ ，这是全书最重要的参数：生物多样性指数， $\theta = 2\lambda p v$ 粗略地说，你可以认为 θ 就是单位时间中受到变异影响的全局生物数量：比如整个环境中单位时间有 2 个生物发生变异，产生了 2 个新生物，那么 $\theta = 2+2=4$ 。

Hubbell 证明了一件事，就是用他的中性模型 θ ，可以推出前面提到的对数序列分布。而他的 θ 同费雪的 α 正好相等。

$$S(\theta) \approx 1 + \theta \cdot \ln \left(1 + \frac{J_M - 1}{\theta} \right).$$

（这个等式同上一个等式基本一致）



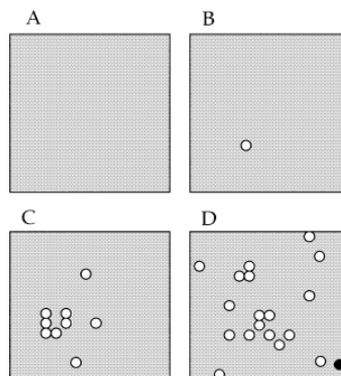
图四：实测物种分布，被对数曲线很好近似

Hubbell 的工作很类似于一道排列组合习题：就是有一堆生物在变异，然后我从中任取 10 个，问平均 10 个中间不同的生物共有几种？这个期望值算出来也就是上面的那个对数曲线。但是，要老老实实推出 Hubbell 的结论，是很困难的，需要用到相当麻烦的数学，贫僧根本不会。而且上次王钺公已经把能讲清楚的都讲清楚了，这里不再赘言。

四、物种的扩散因素

如果不考虑物种的扩散，即把所有物种放在一起，当作一个质点考虑，那只能得到对数曲线，而不能得到三段折线。要得到第二个折点，必须考虑实际的物种是由于基因突变而出现的，基因突变只能在一个地点发生。因此，物种的分布必然是空间不匀质的，在靠近突变位置看到这种生物的机会要比其他位置要高。另一方面，突变本身又是均匀散布在整个空间上的，大量突变平均起来，就导致了一种“驻波”效应。就是当你的观察范围大到某一个程度以后，突然能看到大量新物种。这也就是第二个折点的成因。

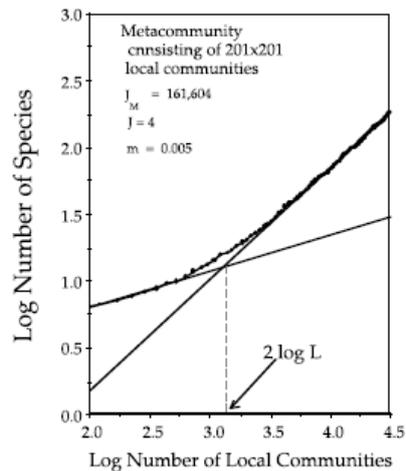
要研究这种现象，不能通过数学的解析推导，只能通过做模拟来研究：



图五：Hubbell 的生物演化模拟

简单说，这个模拟是在第四和第五章的结合，也就是模拟中的每一个点都看作是一个局域群落（第四章的模型），所有这些格点构成一个大的集合群落，并且每一个局域群落之和它的上下左右四个邻居之间发生物种迁移的作用。

这个模拟之下，一个新物种出现后不能一下子占据全局，只能逐渐扩散开去，B 中的那个圈代表新物种，从 B 到 C 到 D 可见它从发生位置逐渐扩散的过程。用这样一种模拟，Hubbell 成功得到了种-面积的三段分布和之前提到的种-个数的 Z 形分布（即图 3 中的 C 曲线）。



图六：通过模拟，得到在大尺度下的幂律改变

要说明的是，个体迁移的过程是很复杂的，涉及到生物和非生物因素的影响。如果按照我们的假设，认为每个物种在空间上均呈均匀随机分布，种-面积关系将出现以上所说的三段式折线。但由于各种因素的影响，每个物种扩散常常是有限的，每个物种的空间分布呈现斑块状。这些斑块具有不同的规模，加以异质地理环境的存在，可以对物种迁移产生物理障碍。这种景观格局常常导致跳跃式的种-面积函数关系，并不见得严格遵循某个幂律关系。因此以上所说的，只是一种较好的近似方式，或解释方式。学习中性场理论，主要是追求一种学术方法上的启发。不应胶柱鼓瑟，把中性场理论的结论当作客观真理看待。

后记：

这次活动来的人很少，几乎到无法正常举行的程度。因此首先想对坚持来的朋友表示感谢，同时也对因故无法来的朋友表示理解。这个讲义主要是把基本的概念讲的比较透彻，很多具体的结论因为篇幅限制，都没有提，大家感兴趣的可以自己去看书或者单独在邮件组里讨论。

中性场理论的各章基本上是独立成篇而有机结合的，前后章节的连接不是特别明显，所以即使前面没有来，从后面听起也不会听不懂，而且一样也可以对整本书获得整体的理解。贫僧接下来又要回上海，参禅闭关，与大家要暂别几个月，但希望大家还是要坚持到底，尽量来参加接下来的活动。

认真参加读书会学习科学的好方法，但有一个问题是，学习科学到底有没有用，有没有意义呢？贫僧乃和尚，出家人不打诳语，我向大家保证，学习科学是有用的，是有意义的。因此希望大家相信我，好好学习，天天向上。

好了，春天到了，有天皆丽日，无地不春风，祝大家春日早发！