

島の生物地理学と Zipf の法則

時田 恵一郎, 入江 治行[†]

大阪大学サイバーメディアセンター・院理・院生命機能

[†] 広島大学情報メディア教育研究センター・院理

The island biogeography and the Zipf's law

Kei Tokita and Haruyuki Irie[†]

Cybermedia Center, Grad. Sci. Sci. & Grad. Sci. Frontier Biosci., Osaka University

[†] Information Media Center & Grad. Sci. Sci., Hiroshima University

MacArthur と Wilson が創始した「島の生物地理学 [1]」の概念は、生態学以外のより一般的な群集に適用可能である。島の面積とそこに生息する種数についての種数面積関係と個体数分布を結び解析を通じて、一般的な群集において普遍的に観察される分布 (Zipf の法則 [2]) の起源を示唆する結果が得られた。単語のベキ分布を説明する古典的な Simon モデル [3] を拡張して、大陸から島への確率的移入のダイナミクスを考えることにより、分布の指数が移入率や死亡率などにより表されることなども示す。

1 島の生物地理学

ダーウィンがガラパゴス諸島を訪れて以来、島は進化生態学研究の最も重要なプラットフォームであり続けてきた。島は比較的閉じた生態系の有り様を教えてくれるので、生物多様性を駆動するメカニズムに興味をもつ生態学者は、古くから島に生息する種の数が何によって決められているかを調べてきた。MacArthur と Wilson は、その不朽の名著「島の生物地理学の理論 [1]」の中で、さまざまな島のいろいろな生物について、島にすむ特定グループの種数 S が、その島の面積 A に

$$S = CA^z$$

というベキ関数で表されることを指摘している。ただし C, z は正定数である。

この関係は、島に限らず、ある生態系において、調査面積と観測される種数の間にも成り立つことが知られており、「種数面積関係」とよばれている。大陸など総種数が多い大きな生態系においては、指数 z は比較的小さな値になり、島などのように比較的種数の少ない生態系では大きな値をとることが多いとされる。

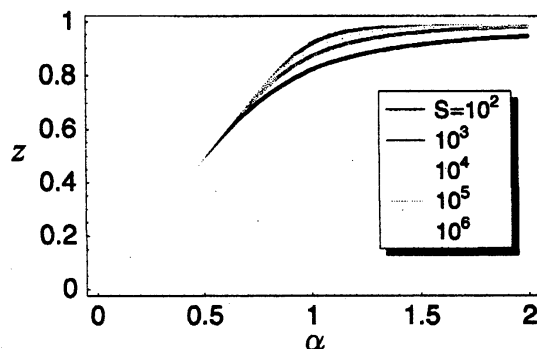


図 1: 種数面積関係のベキ指数 z と個体数分布のベキ指数 α の関係。

島概念はより一般的な生息域へと拡張できる場合がある。例えば、湖は「陸地に浮かぶ島」とみなすことができるし、宿主の体は寄生者からすれば動く島である。さらに、本稿で示すように、非生物的な群集の存在場所も島とみなせる場合がある。さらに、島の理論とパッチ動態やメタ個体群動態の理論との間には類似点が多いので、共通の枠組みで議論が可能な場合がある。

種数面積関係は、個体数分布 $p(x)$ (個体数 x をもつ種の割合) と関係がある。総個体数 N が面積 A に比例するという仮定のもとで、個体数分布から種数面積関係 $S(A)$ を導くことができる。例えば、Preston は実際の観測データによくあてはまる個体数分布として対数正規分布を提案し、経験的に成り立つ「正準仮説」を用いて、(1) 式の種数面積関係を導いた [4]。May は、対数正規分布以外にさまざまな分布を検討して、対応する種数面積関係を導いた [5]。一方、我々は、これまで検討されなかったベキ型の個体数分布 $p(x) \propto x^{-(1+\alpha)}$ が干潟の群集で成り立つことを見だし、正準仮説を仮定することなく (1) 式の種数面積関係が導かれることを示した [6]。さらに、最近我々は有限の種数 S をパラメータとする、種数面積関係のベキ指数 z とベキ型個体数分布の指数 α との関係

$$z = \frac{dy}{d\mu} = \frac{e^{\gamma y} - 1}{(1 + \gamma)e^{\gamma y} - 1} \quad (1)$$

$$y \equiv \ln(S + 1), \quad \mu \equiv \ln A, \quad \gamma \equiv \frac{1 - \alpha}{\alpha} \quad (2)$$

を導いた (図 1) [7]。種数 S が大きい極限では、

$$z = \begin{cases} \alpha & (\text{for } \alpha < 1) \\ 1 & (\text{for } \alpha \geq 1) \end{cases} \quad (3)$$

というシンプルな関係が得られる。図 1 から予想されるように、 $\alpha = 1$ で折れ曲がり、以下で示す Zipf の法則 ($\alpha = 1$) の特異性を表している。

2 Zipf の法則

ベキ分布については、生態学以外に、社会学や言語学などの文脈でも数多くの研究がなされてきた [8]。19 世紀末から、様々な都市の人口分布などがベキ分布に従うことが指摘されていたが、Zipf [2] が、書籍の中の単語の分布など、ベキ分布の中でも特にその指数 α が 1 に近い場合が多いことを指摘して以来、様々な分野の研究者がその理由の説明を試みてきた。しかし、ベキ分布を生成するメカニズムに関する研究は多いが、その指数がなぜ 1 に近いのかを説明する理論はあまりない。後述する Simon のモデルはその数少ないものの一つである。Zipf の法則はベキ分布と同一視される場合もあるが、ここでは、一般のベキ分布と区別するために、ベキ分布の中でも特に $\alpha \simeq 1$ のものを Zipf の法則と定義する。

図 2 に、最も有名な Zipf の法則の例として、書物の中の単語の出現頻度の分布を示す。図 2 は個体数分布ではなく、等価なアバンダンス順位関係を示している。横軸は、単語を

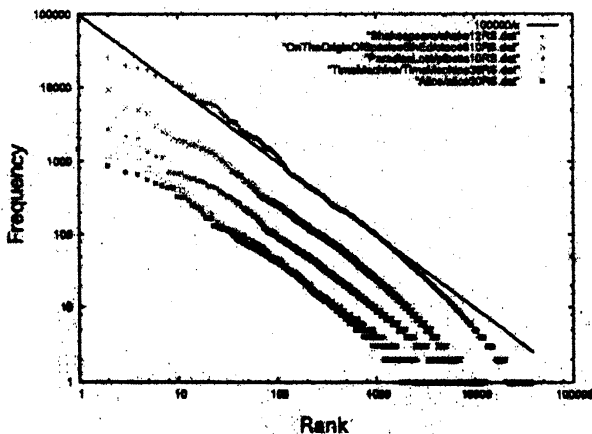


図 2: 書物の単語数に現れる Zipf の法則. 上から, シェークスピア戯曲集, ダーウィン/種の起原第6版, ミルトン/失楽園, ウェルズ/タイムマシン, キャロル/不思議の国のアリス. 直線は $\alpha = 1$ のフィッティング. 電子書籍データは, Project Gutenberg[9] による.

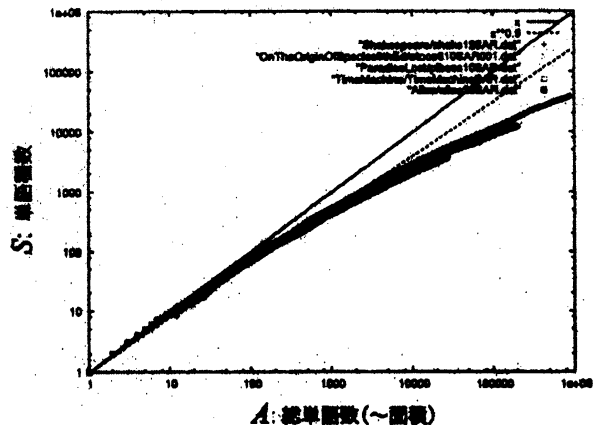


図 3: 書物における, 1 ページ目からの単語数 A と単語種数 S の間に現れる「種数面積関係」. 上の直線は $\alpha = 1$, 下は $\alpha = 0.9$ のフィッティング.

出現する頻度順に並べたときの順位, 縦軸は頻度である. 総単語数が異なる 5 種類の小説におけるベキ型の「個体数分布」が全て同じ指数 $\alpha \simeq 1$ をもつことが見てとれる.

Zipf の法則は, ミクロな分子レベルでも観測されている. さまざまな転写因子の量, ヒトの肝臓, 腎臓, 直腸がんなどの細胞, マウスの胚肝細胞, 線虫, 酵母などにおける様々な遺伝子の発現量が Zipf の法則に従うことが報告されている [10, 11]. このように, ミクロなスケールからマクロなスケールにまで渡って成り立つ Zipf の法則の普遍性は, 個々のシステムの詳細によらないメカニズムの存在を示唆している.

3 一般化された種数面積関係

個体数分布と種数面積関係の間の関係が, 書物における単語の分布においても成り立つことをチェックした. ここでは, 本の長さ, すなわち単語数 (単語の個体数) を本という「生態系」における面積であると考え, 本の最初の一文から数えた単語数とそれまでに出現した単語の「種数」の関係性を調べた. 図 3 は, 「書物における種数面積関係」を図 2 と同じデータを用いて描いたものである. 図 2 で $\alpha \simeq 1$ であるのに対して, z の値は 1 よりやや小さいが, これは S の有限性によるものであり, 式 (1), (2) で予測される値 $z \simeq 0.9$ に近い.

式 (1) から, 種数面積関係の指数 z を, 対数面積 μ を時間とみなしたときの, 対数種数 y の変化「速度」と解釈することができる. よって, z をもう一度 μ で微分した

$$a \equiv \frac{dz}{d\mu} = z \frac{\gamma^2 e^{\gamma y}}{[(1 + \gamma) e^{\gamma y} - 1]^2}$$

は, y の「加速度」と見ることができる. 加速度 a の y と α に対する依存性を図 4 に示す. 任意の対数種数 y に対して, $\alpha \simeq 1$ が加速度の最大値を与えていることがわかる. 厳密に

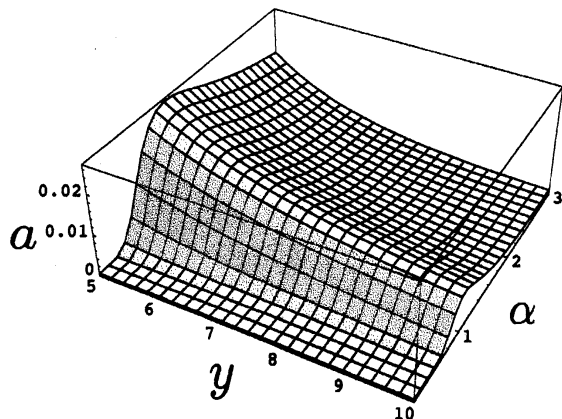


図 4: 対数種数の「増加加速度」 a と、個体数分布のべき指数 α および種数 S との関係。

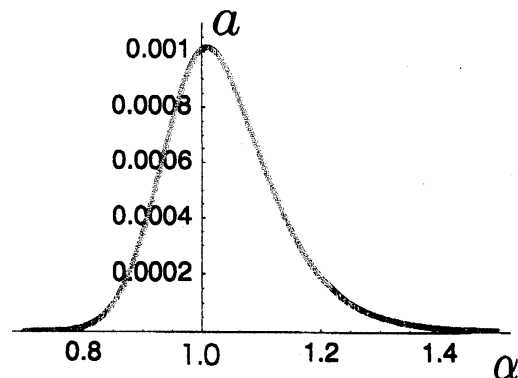


図 5: $y = 20$ のときの、対数種数の「増加加速度」と α の関係。

は、 y の小さい範囲では、 a の最大値を与える α は、 $\alpha_{max} > 1$ となるが、 y が十分大きければ、 $\alpha_{max} \rightarrow 1$ となる。 $y = 20$ のときの加速度 a と α の関係を図 5 に示す。これらの結果から、もしも、対数面積が一定速度で変化したときに、対数種数の加速度を最大にするようなメカニズムが働くと Zipf の法則 ($\alpha = 1$) が成り立つことがわかる。例えば、書物の場合には、著者が文章を書き進めるときに、使用する単語の対数種数の加速度を最大にするよう（できるだけ使用する単語の種類が多くなるように）努力すると Zipf の法則に従う単語の分布になることが予測される。

また、すでにみたように、(3) 式から、 $S \rightarrow \infty$ の極限では、指数 z (対数種数「速度」) の α 微分は特異的となるから、Zipf の法則 $\alpha \simeq 1$ は、対数種数の二次相転移点ということもできる。このように、種数面積関係を一般化することにより、Zipf の法則の特殊性が明らかになり、普遍的な指数 $\alpha \simeq 1$ が、多様性を最大化する原理のもとで一般的に現れうることが示された。

4 Simon モデル

前説で Zipf の法則を与える $\alpha = 1$ が種数増大に関して特別な値であることがわかった。ここでは、そのような特別なべき指数を導く確率過程モデルを考える。本質的には Simon モデル [3] を拡張したものである [12]。Simon は単語数、時間ともに離散変数で定式化・解析したが、ここでは簡単のため連続変数として取り扱うことにする。単語分布における Zipf の法則を導くモデルとして、Simon は以下のようなモデルを考えた。本の著者は、単位時間あたりに、無数の種類の単語からなる「単語プール」から一個の単語を確率 $0 < \nu < 1$ で取り出して文章を書いていくものとする。また、単位時間あたり $1 - \nu$ の確率で、すでに取り出した文章中に使用した単語 i の中からその出現回数 x_i に比例する確率 x_i/N で次の単語を書くとする。単位時間あたり文章の総単語数 N は 1 増えるから、 $N = N(t) = t$

である。よって種数 $S(t)$ と単語の個体数 $x_i(t)$ の時間発展を

$$\frac{dS(t)}{dt} = \nu, \quad \frac{dx_i(t)}{dt} = \frac{1-\nu}{t} x_i(t)$$

と書くことができる。これは遷移確率が x_i に比例する 1 ステップ・プロセス [13] であり、時間に依存する個体数分布 $\sigma(x, t)$ が、

$$\frac{\partial \sigma(x, t)}{\partial t} + \frac{\partial}{\partial x} \left(\frac{1-\nu}{t} x \sigma(x, t) \right) = 0$$

に従うことが示される。これを解くと、規格化された個体数分布 $p(x)$ を用いて、

$$\begin{aligned} \sigma(x, t) &= S(t)p(x) \\ S(t) &= c \exp \left(\alpha \int \frac{1-\nu}{t} dt \right) \\ p(x) &= \alpha x^{-(1+\alpha)} \end{aligned}$$

とベキ型の個体数分布が得られる。ただし、 c, α は定数である。「移入率」 ν を定数と仮定すると、 $\alpha = 1/(1-\nu)$ と、ベキ指数 α が移入率 ν で表される。これより、移入率が小さな極限 $\nu \rightarrow 0$ で Zipf の法則 $\alpha = 1$ が導かれる。

最近 Zanette と Montemurro は、移入率が $\nu(t) \sim t^{-\eta}$ のように時間のベキ関数で減少する場合には、 $\alpha = \eta (< 1)$ となることを示した [12]。また、単語には「死亡」がないので、我々は最近、Simon モデルに個体の死亡の効果を取り入れて、島の生物地理学にも適用可能なモデルへと拡張し、死亡率を $0 < \delta < 1$ として、移入率 ν が定数の場合に、

$$\alpha = \frac{1-\delta}{(1-\nu)(1-2\delta)} (\leq 1)$$

という関係を得ている [14]。さらに、移入率が一般的な時間の減少関数 $\nu(t) = dS(t)/dt$ の場合にも、簡単な仮定から Zanette と Montemurro と同様の結果が得られることもわかっている。

5 まとめと議論

島の生物地理学において知られるベキ型の種数面積関係とベキ型の個体数分布との間の関係を解析することにより、Zipf の法則がある特別な場合に成り立っていること、およびある種の相転移点に対応していることがわかった。これにより、生態学だけでなく、単語や姓名の分布、都市人口分布など、より一般的な文脈で成り立つ Zipf の法則と多様性の関数の最大化が関係していることが示された。さらに、このような一般的なベキ型の個体数分布を導く一般的な確率過程 (Simon モデル) を紹介し、最近の発展についても概説した。

島の生物地理学モデルへの Simon モデルの拡張は、島の中での種間相互作用を考慮しない、最も単純な群集集合モデル (Community assembly model) とみなすこともできる。

実際の生態系においては、 $\alpha \simeq 0.5$ というデータがあり [6], Zipf の法則はむしろ成り立っていないが、種間競争による移入率の時間的減少の効果である可能性がある。より複雑な群集集合モデルがどのような種数面積関係を導くかといった問題が今後の課題である。

参考文献

- [1] R. H. MacArthur and E. O. Wilson. *Island Biogeography*. Princeton University Press, Princeton, 1967.
- [2] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.
- [3] H. A. Simon. On a class of skew distribution functions. *Biometrika*, Vol. 42, pp. 425–440, 1955.
- [4] F. W. Preston. The commonness and rarity of species. *Ecology*, Vol. 29, pp. 254–283, 1948.
- [5] R. M. May. *Patterns of species abundance and diversity*, pp. 81–120. Belknap, Cambridge, 1975.
- [6] 入江治行, 時田恵一郎, 羽原浩史. ベントスの種個体数分布と種数面積関係. 京都大学数理解析研究所講究録 No.1432 「生物数学の理論とその応用」, pp. 116–120, 2005.
- [7] H. Irie and K. Tokita. In preparation. 2006.
- [8] <http://www.nslij-genetics.org/wli/zipf/>.
- [9] <http://www.gutenberg.org/>.
- [10] O. Ogasawara, S. Kawamoto, and K. Okubo. Zipf's law and human transcriptomes: an explanation with an evolutionary model. *C. R. Biologies*, Vol. 326, pp. 1097–1101, 2003.
- [11] C. Furusawa and K. Kaneko. Zipf's law in gene expression. *Phys. Rev. Lett.*, Vol. 90, p. 088102, 2003.
- [12] D. H. Zanette and M. A. Montemurro. Dynamics of text generation with realistic Zipf's distribution. *Journal of Quantitative Linguistics*, Vol. 12, pp. 29–40, 2005.
- [13] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1981.
- [14] K. Tokita and H. Irie. In preparation. 2006.