Santa Fe Institute Workshop Proposal

Title: *Universal DiversityPatterns Across the Sciences*

Organizer information:

Jeffrey C. Nekola; University of New Mexico; jnekola@unm.edu

Summary

Several intensively studied ecological diversity patterns appear to have almost exact analogues across a wide range of disciplines including physics, economics, linguistics, sociology and the geosciences. As a result, the ultimate mechanisms underlying them would seem to be more general than those typically offered or considered by disciplinary researchers. The aim of this workshop is to initiate an interdisciplinary collaborative investigation which will: (1) evaluate the regularity of apparently universal diversity patterns across multiple disciplines, (2) explore their mutual connections (3) conduct tests to show which of these patterns can be considered to be non-superficially regular, (4) explore these regularities via a set of candidate explanative frameworks (e.g. statistical constraints, system self-organization, maximum entropy, random branching processes and fractality), and (5) explore possible links between these frameworks. This investigation will assist documentation of important commonalities between disciplines, and will help focus attention on those unified processes that generate diversity across the physical, biological, and socio-cultural world.

Problem Statement

Across the sciences, investigations into patterns of entity number, abundance, and distribution constitute some of the most important statistics used to describe system diversity. In ecology, evolutionary biology and biogeography the functional forms of these patterns often appear universal across taxa and spatial and/or temporal scales, including not only the frequency/probability distributions of species abundance, body size, and range size, but also the bivariate relationships between certain variables (e.g. species number vs. area or compositional similarity vs. intersample distance) or their statistical properties (e.g. mean vs. variance). The ability to generate these patterns has become vital to discriminating between various competing theoretical ecological models (Bell 2000, Hubbell 2001, Chave et al. 2002, Chave 2004). However, these patterns also appear to have almost exact analogues across a wide range of other disciplines including physics, economics, linguistics, sociology and the geosciences (Preston 1981, Gaston *et al.* 1993, Ugland *et al.* 2005, Nekola & Brown 2007). As a result, the mechanisms underlying them must not be wholly biological. For instance, the processes of birth, death, immigration, speciation and evolution invoked by Hubbell (2001), or the competitive tradeoffs invoked by Tilman (2004) or Chave *et al.* (2004) could not have generated the rarity-enriched abundance distributions found in precipitation levels, paper citations, stock volumes, drink tumbler longevity, and marriage ages, or the power-law accumulation of unique mineral, word, garden seed varieties, or comprehensive exam scores with increasing sample size. Conversely, it seems unlikely that performances of the band Cowboy Junkies mimic community ecology process, even though their setlists display rarity-enriched abundance distributions, power-law accumulation curves and non-linear distance decay (Nekola & Brown 2007).

This strongly suggests that diversity patterns must be at least partially influenced by processes more fundamental than those typically offered or considered by ecologists or other disciplinary investigators. The list of possible fundamental drivers includes, but is not limited to: simple statistical constraints of sample size, design, and the number of agents; system self-organization, including self-organized criticality; maximum entropy; random branching processes; and fractality. While at various times each of these concepts has been individually advanced as a possible explanation for a particular diversity pattern, little effort has been afforded to the joint consideration of these processes. As a result, the relative importance and interactions between these potential drivers remains unknown.

Significant progress towards general understanding of these seemingly ubiquitous patterns and the processes underlying them has been impeded by the following three issues:

1. It is not clear to what extent the apparent universality of these patterns is superficial and coincidental or real and due to nontrivial similarities in ultimate process. In particular, it is important to determine if elements of these distributions indicate the presence of specific quantitative properties which do not follow immediately from simple statistical reasoning before subjecting them to more rigorous analyses.
2. The commonality of these patterns might indicate that they are different two dimensional projections of a common multidimensional structure or process. It is thus also imperative that the potential mathematical and logical connections between these patterns be better understood. As this has been done for only some pairs (e.g. frequency distribution of abundances and the accumulation of unique events), it is thus unknown to what extent they might best be treated within a single universal framework (e.g. maximum entropy).
3. Many of these patterns have been studied and their functional forms tested only by the simple fitting of mathematical functions describing their statistical relationships. However, a much more powerful way to test their universality is to explore their invariance across various observational scales with differing sample resolutions, and/or metrics. This has been done only in a few cases, and often not by the most appropriate method.

We believe that these issues can be most readily addressed by initiating collaboration between a group of experts familiar with these patterns across a wide variety of fields. We furthermore feel that the Santa Fe Institute represents the obvious venue to commence this project, given its longstanding history of supporting such types of projects. The aim of this project will be to begin: (1) evaluating the regularity of apparently universal diversity patterns across multiple disciplines, (2) exploring their mutual connections (3) conducting tests to identify which properties of these patterns can be considered non-superficially regular, (4) exploring these regularities via a set of candidate explanative frameworks (e.g. statistical constraints, system self-organization, maximum entropy; random branching processes; fractality), and (5) exploring possible links between these frameworks. Such documentation of the ultimate causes for universal diversity patterns has the potential to uncover important commonalities between disciplines, and to help focus attention on fundamental processes that generate common patterns seen across the physical, biological, and socio-cultural world.

Proposed Activities

The workshop will begin these investigations via three meetings, staggered at approximately 6-month intervals. During these meetings we will evaluate the universality of a given pattern, explore the root causes of such regularities from multiple explanatory frameworks, and consider potential links between these frameworks and possible connections between patterns. To encourage optimal development of ideas, for part of each meeting we will break into subgroups which will each consider the utility of a given explanatory framework (statistical constraints, system self-organization, maximum entropy, random branching processes, fractality) in the generation of a given pattern found to exhibit interesting universal aspects. Participant interaction and idea exchange between and following meetings will be coordinated via a workshop Wiki.

To keep the goals for this workshop tractable, we plan to initially consider five apparently universal diversity patterns:

1. ***Rarity-enriched frequency distributions*****.** Preston (1950) noted that the Boltzmann frequency distribution of molecular kinetic energies in gasses, the Pareto frequency distribution of personal incomes in countries, and the frequency distribution of species abundances in ecological communities (SAD) were remarkably similar, and later (1981) expanded this list to also include the first marriage ages of Danish, British, and US women, the longevity of restaurant drink tumblers, and the stress required to fracture microscope slides. While Preston claimed that all these distributions were close to lognormal, his analyses were severely limited by the careless use of abundance log transformation. More careful analysis demonstrates, however, that abundance distributions across a wide variety of systems (e.g., precipitation classes, paper citations, stock volumes, garden seed offerings, concert setlists, word use and US trailer home frequencies; Nekola & Brown 2007) typically demonstrate non-linear monotonically decreasing probability functions. While some of these can be characterized by negative power-laws (e.g. word use frequency), many do not display a power law signature (Newman 2005) and appear to be modeled at least as well (or better) via exponential or log-series decay. The patterns of occurrence for these various distributions across systems remain unknown, as is the potential universe of generating mechanisms.

2. ***Mean-variance relationship***. While variation of many factors generally increases with mean, this regularity often represents a power-law relationship with a coefficient of 2. This relationship has been observed over a wide array of fields including population ecology (Taylor 1961), epidemiology (Rhodes & Anderson 1996, Keeling & Grenfell 1999, Philippe 1999), behavior (Anderson & May 1988), oncology (Kendal 2002), genomics (Kendal 2003), histology (Azavedo & Leroi 2001), body-size evolution (Boyer *et al. in preparation*), and sociology (Boyer, *unpublished data*). Although population ecologists have sought explanation for this pattern in the factors regulating species-specific spatiotemporal population fluctuations, it is clear that such specific mechanisms cannot apply across all of these systems.

3. ***Sample size and the frequency of occurrence of unique entities.*** In ecology, this relationship is exemplified by the power-law increase of unique species with increasing sample area (the Species Area Relationship; SAR) or time (Species Time Relationship; STR). Ecologists have generally sought explanations for this pattern in the rules governing the assembly of communities and metacommunities. However, similar a similar power-law accumulation of unique events with increasing spatial or temporal extent is also revealed in systems not governed by such inherently ecological mechanisms including minerals across the globe, words in texts of differing lengths, varieties of garden seeds sold over time in North America, comprehensive exam scores and stock prices over time, and unique song performances in band setlists (Ugland et al. 2005, Nekola & Brown 2007).

4. ***Distance decay of similarity***. The closer observations are either in space or time, the more similar they tend to be. Examples of this phenomenon include not only a wealth of ecological systems, spanning all taxa groups from mountain tops to the abyssal plain (Nekola &

White 1999; Nekola, *unpublished data*), but also the yearly commercial offerings of garden seeds in North America, the ingredients used in global cuisines, and concert setlists (Nekola & Brown 2007). This pattern of Distance Decay (DD) is so prevalent that it has been enshrined as the 'First Law of Geography' (Tobler 1970). While this pattern may be a simple outcome of spatiotemporal heterogeneity within systems, it seems likely that other fundamental factors may be at play given the systematic differences found in the functional form of this relationship typically observed within natural (exponential decay) and human (power law decay) systems.

5. ***Quarter-power scaling of energy use to system size.*** The power-law scaling in biological systems of body size to metabolic rate and other various life-history attributes such as fecundity tend to possess highly regular quarter power coefficients. While it has been clearly shown how this effect can arise from existence of an area-preserving, closed, branching fractal circulatory system (West *et al.* 1997; Brown *et al.* 2004: West & Brown 2005), it is less clear why this pattern also exits for invertebrates with open circulatory systems (Peters 1996) and bacteria and single-cell eukaryotes (West *et al.* 2002) that lack circulatory systems. Similar quarter-power scaling also exists in the relationship between population number and abundance (Keitt *et al.* 2002) and between fecundity (Moses & Brown 2003), gross domestic product (William Woodruff, *personal communication*) and energy use in modern human societies. These examples suggest that the West *et al.* model may function as a special case within a more general framework, perhaps related to general issues of network geometry (Banavar 2002) or greedy optimization (Gastner & Newman 2006).

Workshop Panel:

James H. Brown, University of New Mexico
John Harte; University of California – Berkeley
Fangliang He; University of Alberta
Fabrizio Lillo; University of Palermo
Brian McGill; University of Arizona
Jeff Nekola; University of New Mexico
Arnost Sizling; University of Sheffield
David Storch; Charles University

Anticipated results and beneficiaries

We anticipate that this workshop will initiate collaborative investigations regarding universal diversity patterns across the sciences. Over the course of the three scheduled sessions, we will identify what diversity patterns are truly universal, consider potential root drivers, and explore possible linkages between them. We fully recognize that full completion of these tasks are likely too ambitious to expect within only six days of meetings. As such, we feel that one of the most important outcomes from this project is for stimulating additional research by bringing together this group of experts for the first time. We will also use the last schedules meeting to set goals and priorities for future research efforts and to consider ways to secure additional funding via other private and public organizations.

We ultimately anticipate that the results of this and any future workshops would be readily communicated via a contribution to SFI's "Studies in the Sciences of Complexity" series, and also envision broad dissemination of findings to the scientific community via a series of papers submitted to top journals.

References

Anderson, R. & R. May. 1988. Epidemiological parameters of HIV transmission. *Nature* **333**, 514–519.

Azevedo, R. & A. Leroi. 2001. A power law for cells. *Proceedings of the National Academy of Sciences USA* **98**, 5699–5704.

Banavar, J.R., J. Damuth, A. Maritan & A. Rinaldo. 2002. Supply-demand balance and metabolic scaling. *Proceedings of the National Academy of Sciences USA* **99**, 10506–10509.

Bell, G. 2000. The distribution of abundance in neutral communities. *American Naturalist* **155**, 606-617.

Boyer, A.G., A. Clauset & C.R. McClain. *in review*. An evolutionary random walk predicts mean-variance scaling in body size across the animal kingdom.

Brown, J.H., J.F. Gillooly, A.P. Allen, V.M. Savage & G.B. West. 2004. Toward a metabolic theory of ecology. *Ecology* **85**, 1771-1789.

Chave, J. 2004. Neutral theory and community ecology. *Ecology Letters* **7**, 241-253.

Chave, J., H.C. Muller-Landau & S.A. Levin. 2002. Comparing classical community models: theoretical consequences for patterns of diversity. *American Naturalist* **159**, 1-23.

Gastner, M.T. & M.E.J. Newman. 2006. Shape and efficiency in spatial distribution networks. Journal of Statistical Mechanics: Theory and Experiment. doi:10.1088/1742-5468/2006/01/P01015

Gaston, K.J., Blackburn, T.M. & Lawton, J.H. 1993. Comparing animals and automobiles: a vehicle for understanding body size and abundance relationships in species assemblages? *Oikos*, **66**, 172-178.

Hubbell, S.P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology #32. Princeton University Press, Princeton, New Jersey.

Keeling, M. & B. Grenfell. 1999. Stochastic dynamics and a power law for measles variability. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **354**, 769–776.

Keitt H., L.A.N. Amaral, S.V. Buldryev & H.E. Stanley. 2002. Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **357**, 627-633.

Kendal, W. 2002. A frequency distribution for the number of hematogenous organ metastases. *Journal of Theoretical Biology* **217**, 203–218.

Kendal, W. 2003. An exponential dispersion model for the distribution of human single nucleotide polymorphisms. *Molecular Biology E* **20**, 579–590.

Moses, M.E. & J.H. Brown. 2003. Allometry of human fertility and energy use. *Ecology Letters* **6**, 295-300.

Nekola, J.C. & J.H. Brown. 2007. The wealth of species: ecological communities, complex systems and the legacy of Frank Preston. *Ecology Letters*, **10**:188-196.

Nekola, J.C. and P.S. White. 1999. Distance decay of similarity in biogeography and ecology. *Journal of Biogeography*. 26:867-878.

Newman, M.E.J. 2005. Power laws, Pareto distributions, and Zipf's law. *Contemporary Physics.* **46:**323-351.

Peters, R.H. 1986. *The ecological implications of body size*. Cambridge University Press, Cambridge.
Philippe, P. 1999. The scale-invariant spatial clustering of leukemia in San Francisco. *Journal of Theoretical Biology* **199**, 371–381.
Preston, F.W. 1950. Gas laws and wealth laws. *Scientific Monthly*, **71**, 309-311.
Preston, F.W. 1981. Pseudo-lognormal distributions. *Ecology*, **62**, 355-364.
Rhodes, C. & R. Anderson. 1996. Power laws governing epidemics in isolated populations. *Nature* **381**, 600–602.
Taylor, L. 1961. Aggregation, variance and the mean. *Nature* **189**, 732–735.
Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences USA* **101**, 10854-10861.
Tobler, W. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*. 46:234-240.
Ugland, K.I., J.S. Gray & J.D. Lambshead. 2005. Species accumulation curves analysed by a class of null models discovered by Arrhenius. *Oikos*, **108**: 263-274.
West, G.B. & J.H. Brown. 2005. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology* **208**, 1575-1592.
West, G.B., J.H. Brown & B.J. Enquist. 1997. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122-126.
West, G.B., W.H. Woodruff & J.H. Brown. 2002. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proceedings of the National Academy of Sciences USA 99*, 2473-24