

The Stochastic Community

THE STOCHASTIC COMMUNITY
Toward a Statistically Exact Theory of Biodiversity

A. K. Dewdney, Professor Emeritus
Departments of Computer Science and Biology
Western University*, London, Ontario
CANADA N6A 5B7

*formerly The University of Western Ontario

CONTENTS

Section	Page number
Foreword	(to come)
1. The J-curve and the J distribution	1/1
1.1 The stochastic species hypothesis and the J distribution	1/1
1.2 Appearance of the J-distribution	1/3
1.3 The community and the sample	1/5
1.4 The need for an appropriate theory	1/7
1.5 A positive test using multi-sample data	1/10
1.6 Fluctuating populations	1/12
2. The J-distribution and its variations	2/1
2.1 Properties of the probability density function	2/1
2.1.1 The capacity constant	2/1
2.1.2 The mean and variance of the J distribution	2/3
2.2 The J distribution: species and individuals	2/3
2.2.1 The discrete J distribution	2/4
2.2.2 The canonical sequence	2/5
2.2.3 An implicit formula for rank abundance	2/8
2.2.4 Effect of a log transformation on the J distribution	2/9
3. Sampling in practice and in theory	3/1
3.1 Randomness and random numbers	3/1
3.1.1 Generating random numbers	3/1
3.2 Communities and samples	3/3
3.3 The sampling process	3/4
3.3.1 The variety of sampling activity	3/4
3.3.2 Estimating sample intensity	3/5
3.3.3 How samples have been used: calculating "biodiversity"	3/6
3.4 Computer simulation of sampling	3/8
3.4.1 A sample simulation algorithm	3/9
3.5 The general theory of sampling	3/10
4. Compiling and analysing field data	4/1
4.1 Histograms and distributions	4/1
4.2 Other representations: rank abundance	4/3
4.3 Other representations: logarithmic abundance	4/4
4.4 Estimating parameters of the J distribution	4/5
4.4.1 The chi square test	4/5
4.4.2 The Kolmogoroff-Smirnov (K-S) test	4/7

4.5 Application example: sample overlap & similarity	4/8
5. Predictions from data	5/1
5.1 Predicting maximum abundance	5/1
5.2 Predicting species richness	5/2
5.2.1 Inadequacies in current methods of estimation	5/4
5.3 Exact estimation methods	5/8
5.3.1 The two-step method with an example	5/10
5.4 Experimental evaluation of methods	5/12
5.4.1 The two-step method	5/13
5.4.2 The one-step method	5/14
5.5 The behaviour of error terms	5/15
5.6 Analysis of the two sources of variance	5/16
6. Extending the sample	6/1
6.1 The effect of sampling on parameters	6/1
6.2 Accumulation curves	6/3
6.2.1 Accumulation with replacement	6/4
6.2.2 Accumulation without replacement	6/6
6.3 The species area relationship	6/9
6.3.1 Other species-area laws	6/11
7. Stochastic systems and the stochastic community	7/1
7.1 The stochastic community and stochastic systems	7/1
7.1.1 Probability and time	7/2
7.1.2 Generalizations of the multi species logistic system (MLS)	7/3
7.1.3 Stochastic abundances	7/5
7.2 Stochastic communities in nature	7/6
7.2.1 Stochastic species hypotheses for communities	7/6
7.2.2 Detecting the J distribution in natural communities	7/7
7.2.3 The stochastic orbit and its variations	7/9
7.2.3 Long runs in stochastic behaviour	7/10
7.2.4 Cyclic changes in abundance	7/11
8. The metastudy: A review	8/1
8.1 Background: the Chi square theorem and test	8/1
8.1.1 An illustration of the central methodology	8/2
8.2 Applying chi square theory to multiple source histograms	8/4
8.2.1 The score conversion process	8/6
8.3 The metastudy	8/7
8.3.1 The data collection procedure	8/7
8.3.2 Testing the data against two distributions	8/8
8.3.3 Converting and compiling the scores	8/9

8.4 Experimental results of the metastudy	8/9
9. Fossil J-curves	9/1
9.1 Background of the problem	9/1
9.2 Presence of the J distribution in taxonomic data	9/3
9.2.1 The test method	9/5
9.2.2 Results of the study	9/6
9.3 Evolutionary origin of the J distribution	9/7
9.3.1 Stochastic genera	9/8
9.3.2 Episodic speciation	9/9
9.4 Extinction and speciation in natural and artificial communities	9/11
9.4.1 Extinction rates in the MSL system	9/11
9.4.2 Speciation in stochastic communities	9/12
10. Summary of theory and open problems	10/1
10.1 Summary of research	10/1
10.1.1 Examples and counter-examples	10/3
10.2 A guide to field methods and theory development	10/3
10.3 Open problems and prospects	10/4
10.3.1 The multispecies logistical system and its variations	10/5
Appendix A: Mathematical Notes and Computer Tools	A/1
A.1 Derivation of the J distribution for stochastic systems	A/1
A.2 The discrete form and emergence of the second parameter	A/4
A.3 Useful formulas and equations	A/5
A.3.1 Integration by substitution	A/5
A.3.2 Summation of a power series	A/6
A.3.3 derivation of the formula for variance	A/6
A.3.4 Derivation of the canonical formula	A/7
A.3.5 Proof of the random hierarchy	A/9
A.3.6 A random time series follows the exponential distribution	A/9
A.3.7 The log-series distribution	A/10
A.3.8 The transfer equations	A/11
A.3.9 Sampling the univoltine distribution	A/13
A.4 Computer research tools	A/15
A.4.1 Systems emulators	A/16
A.4.2 Sampling and estimation programs	A/16
A.4.3 Statistical programs	A/17
A.4.4 Special utility programs	A/17
Appendix B: Results of the metastudy for the J distribution	B/1

Appendix C: Results of the test for the J distribuion in taxonomic data	C/1
Back matter	
References	
Text references	M/1
Web reference	M/9
Metastudy references	M/10
Index and Glossary (to come)	M/19
End paper	M/25
Global map of metastudy biosurvey sites	

PREFACE

In an age when the natural environment is under threat from habitat loss, pollution, and climate change, it is more important than ever that ecologists develop and employ uniform procedures for assessing the condition of all the canaries in the mine, so to speak. Unless results can be compared directly, confusion is likely to result. The lack of a universally accepted (theoretical) species abundance distribution, along with appropriate methodology for assessment, hamper the field greatly. A widely accepted and well established distribution would be greatly preferable to the confusion of tongues that now characterizes the field.

The natural focus of the population biology described in the following pages is not the species, but the community of which it forms a part. The behavior of populations may forever defy our efforts to quantify them beyond recognizing their fundamentally stochastic (i.e., random) nature. But a *community* of populations produces a collective behavior, population-wise, that almost always follows the same general pattern, a J-curve. It would seem that unpredictability, far from being a cause for despair, lies at the heart of the exact methods employed here and provides new theoretical horizons for a science that has suffered from a lack of a) adequate contact with real data and b) appropriate statistical methods. It would appear that a great deal more data is required for studies of abundance patterns than was formerly thought.

The J distribution described in this monograph is intended to capture the ubiquitous shape of the of the J-curve. Mathematically it could hardly be simpler yet it is rather odd statistically, a hyperbola truncated by its axes. This form is implied mathematically by a single hypothesis. Ecologists, field biologists and other readers of this monograph should have no trouble following the theorems and derivations that appear here, provided they have the minimum expected background of college calculus, probability and statistics, along with high school algebra. Equipped with this background the reader should be able to follow all of the mathematical developments, including theorems and their applications. The most involved derivations are left to an appendix, so as to pose minimum interruptions for the reader.

The order in which ideas are presented in this book may call for some patience from the reader. The definitions and applications of the hyperbolic theory are presented in advance of the theoretical arguments and empirical tests that establish the presence of the J distribution in natural communities. For those who wish to reassure themselves by reading ahead, the mathematical derivation of the J distribution from the stochastic species hypothesis will be found in Appendix A, while the empirical metastudy that establishes the presence of the J distribution in field samples of species abundances is the subject of Chapter 8. The stochastic species hypothesis, which leads directly and logically to the (hyperbolic) J distribution, is introduced in Section 1.6 of Chapter 1.

I have been fortunate in my relations with field biologists in assisting with the analysis of many of their datasets, not to mention microbial surveys conducted by me. (Dewdney 1996, 2010) My feet, I should hope, are firmly on biological ground. Crucial to such analyses has been the

development of a variety of computer programs, as the reader will discover in the pages to come.

I have also been fortunate to enjoy a profession as mathematician along with some training in microbiology and an abiding interest in nature from a holistic point of view. It therefore seem natural to me to be managing an ATBI project in southern Ontario, even as I work on how populations behave within communities.

Note: In previous publications I have called the J distribution by another name, the logistic-J distribution. The latter name has been dropped owing to confusion with the “logistic distribution” of Balakrishnan (1991) with which it has no particular relationship. The word “logistic” derived in this case from the limiting effect of finite biomass on the abundances to be found in any given community of organisms.

ACKNOWLEDGMENTS

Special thanks go to those who read an earlier draft manuscript of this book and offered many useful suggestions: Arthur Benke, University of Alabama; Emilio Carral Vilarino, Universidad de Santiago de Compostela; Hsin Chi, National Chung Hsing University, Taiwan, Eungchun Cho, Kentucky State University; Maurice Jansen, Netherlands Ministry of Agriculture, Marcos Mendez Iglesias, Universidad Rey Juan Carlos, Spain; Alfonso Pelli, Universidad Federal do Triangulo Mineiro, Brazil, William D. Taylor, University of Waterloo

I wish to thank the following scientists who made early contributions through discussions and consultations: Robert Bailey, Western University; David R. Barton, University of Waterloo; Daniel Botkin, University of California at Santa Barbara, Stanley Caveney, Western University; Ryan A. Chisholm, Princeton University; Albert T. Finnamore, Royal Alberta Museum; Kevin Gaston, University of Sheffield; Duncan Golicher, Universidad Autonoma de Mexico; Paul Handford, Western University, Rob Hughes, Queen Mary University of London; Tigga Kingston, Boston University; Michael Lynch, University of Waterloo; Lynn Margulis (deceased), University of Massachusetts, David Schneider, Memorial University of Newfoundland; Gregory Thorn, Western University, and Liana Zanette, Western University.

I also thank the many sources of support and assistance with (a) the collection of biosurvey literature: graduate students Monica Havelka, Michelle Bowman, Kellie White, Dave Muzia; (b) diagrams and figures: Pablo Jaramillo and Jonathan Dewdney; (c) locating literature: Linda Dunn of the Western University library System; (d) navigating the ITIS website: David Nicolson of ITIS.

Finally, I am grateful to the Western University Biology Department for the Adjunct Appointment and available facilities from 1996 to 2010. During this period the bulk of the research reported here was completed.

INTRODUCTION

“Truth proceedeth more readily from error than from confusion.”

Sir Francis Bacon

The research described in this book has been aimed at settling the question of how the same general shape of distribution arises when one takes a sample of living communities, whether fish, fowl or fungi. The shape, variously called the “J-curve” or the “hollow curve”, appears in some 15 histograms shown at the end of this introduction, all actual samples drawn *at random* from the literature.

In each histogram the species are divided into abundance categories. The first or leftmost column counts the number of species having the lowest abundance in the sample. The next column counts the number of species having the next lower abundance in the sample, and so on. Like much biological data the histograms are neither particularly smooth nor continuous. I have yet to see a field sample where the J-shape does not appear when plotted with appropriate categories. Over the years some ten formulae have been proposed in the literature as the underlying shape. In the author's opinion none of them have been properly tested and none appear to be correct.

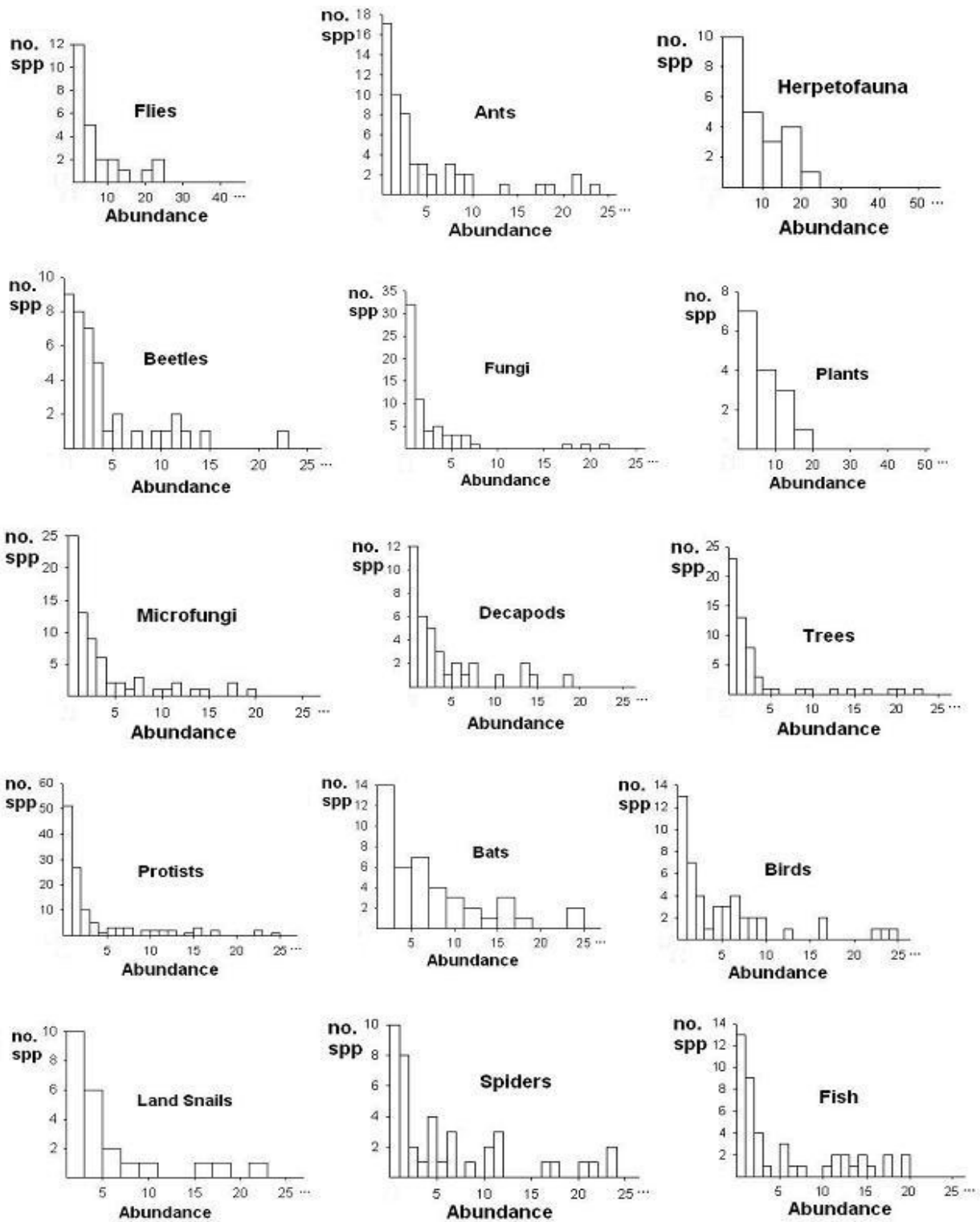
Thanks to an accidental discovery made early on, the author now finds himself nearing the end of a 20-year research program. The discovery involved a stochastic simulation of protist predation (Gause 1937) in which the J-curve appeared in the author's computer screen. It led by degrees to a very general scenario called the stochastic species hypothesis wherein the abundances of species in a community fluctuate randomly. The fluctuations could be described as stochastic vibrations based on equal (or nearly equal) birth and death probabilities. A mathematical analysis of the process next revealed a very simple underlying formula called the J distribution. An extensive metastudy involving 125 randomly selected samples from the literature compared the empirical data with the J distribution, revealing a fit (via the chi square test) that had an overall score that was optimal in the sense that it could hardly be better. The second best proposal had a score that was statistically separable from the optimum and the remaining proposals were out of the running entirely when it came to closeness of overall fit to the data.

Along the way a number of useful mathematical theorems have emerged. The most important of these states that the statistical shape of a well-taken sample of a community of organisms must be the same as the shape of the community itself. In other words, the J-curve that appears in the sample merely reflects a J-curve followed by the community. Given that the sample histogram reflects the J distribution, so must the histogram of the community, albeit with different parameters. This result led directly to a method for estimating with a certain statistical accuracy the species richness of a community based on the richness of the sample. Although several authors have made such attempts, none of the proposed methods carried an error term that could be applied to the estimate.

Other results arose in the course of developing what the author has called “exact methods”.

Given two field samples from the same general area for example, the underlying species overlap of the samples may be calculated directly with the same degree of statistical precision. In addition an appropriate species/area formula has been developed, along with other applications.

In 1948 the British ecologist C. B. Williams, working with nocturnal moths flying to a light trap, compiled a great many such histograms and was struck by the fact that even when different communities of moths were involved, the same pattern inevitably appeared. He called the general pattern the "hollow curve." His opinion about hollow curves lies at the foundation of the research reported here. He thought the hollow curves were essentially hyperbolas. In the author's opinion, Williams happened to be correct.



A random sample of random samples hints at the universal presence of the J curve