



Power Laws in the Information Production Process: Lotkaian Informetrics

Leo Egghe

TABLE OF CONTENTS

Preface	vii
Table of contents	xi
Introduction	1
Chapter I Lotkaian Informetrics: An Introduction	7
I.1 Informetrics	7
I.2 What is Lotkaian informetrics ?	14
I.2.1 The law of Lotka	14
I.2.2 Other laws that are valid in Lotkaian informetrics	19
I.3 Why Lotkaian informetrics ?	25
I.3.1 Elementary general observations	26
I.3.2 The scale-free property of the size-frequency function f	27
I.3.3 Power functions versus exponential functions for the size-frequency function f	32
I.3.4 Proof of Lotka's law based on exponential growth or based on exponential obsolescence	34
I.3.4.1 Proof of Lotka's law based on exponential growth: the Naranan model	34
I.3.4.2 Proof of Lotka's law based on exponential obsolescence: solution of a problem of Buckland	40
I.3.5 Derivation of Mandelbrot's law for random texts	42
I.3.6 "Success Breeds Success"	45
I.3.6.1 The urn model	46
I.3.6.2 General definition of SBS in general IPPs	49
I.3.6.3 Approximate solutions of the general SBS	52
I.3.6.4 Exact results on the general SBS and explanation of its real nature	55

I.3.7	Entropy aspects	65
I.3.7.1	Entropy: definition and properties	66
I.3.7.2	The Principle of Least Effort (PLE) and its relation with the law of Lotka	70
I.3.7.3	The Maximum Entropy Principle (MEP)	76
I.3.7.4	The exact relation between (PLE) and (MEP)	78
I.4	Practical examples of Lotkaian informetrics	85
I.4.1	Important remark	85
I.4.2	Lotka's law in the informetrics and linguistics literature	86
I.4.3	Lotka's law in networks	87
I.4.4	Lotka's law and the number of authors per paper	90
I.4.5	Time dependence and Lotka's law	92
I.4.6	Miscellaneous examples of Lotkaian informetrics	94
I.4.7	Observations of the scale-free property of the size-frequency function f	98
Chapter II	Basic Theory of Lotkaian Informetrics	101
II.1	General informetrics theory	101
II.1.1	Generalized bibliographies: Information Production Processes (IPPs)	101
II.1.2	General informetric functions in an IPP	104
II.1.3	General existence theory of the size-frequency function	110
II.2	Theory of Lotkaian informetrics	114
II.2.1	Lotkaian function existence theory	114
II.2.1.1	The case $\rho_m = \infty$	114
II.2.1.2	The general case $\rho_m < \infty$	116
II.2.2	The informetric functions that are equivalent with a Lotkaian size-frequency function f	121

II.3 Extension of the general informetrics theory: the dual size-frequency function h	144
II.4 The place of the law of Zipf in Lotkaian informetrics	150
II.4.1 Definition and existence	150
II.4.2 Functions that are equivalent with Zipf's law	152
Chapter III Three-dimensional Lotkaian Informetrics	157
III.1 Three-dimensional informetrics	157
III.1.1 The case of two source sets and one item set	158
III.1.2 The case of one source set and two item sets	159
III.1.3 The third case: linear three-dimensional informetrics	161
III.1.3.1 Positive reinforcement	163
III.1.3.2 Type/Token-Taken informetrics	168
III.1.4 General notes	172
III.2 Linear three-dimensional Lotkaian informetrics	175
III.2.1 Positive reinforcement in Lotkaian informetrics	175
III.2.2 Lotkaian Type/Token-Taken informetrics	177
Chapter IV Lotkaian Concentration Theory	187
IV.1 Introduction	187
IV.2 Discrete concentration theory	192
IV.3 Continuous concentration theory	196
IV.3.1 General theory	196
IV.3.2 Lotkaian continuous concentration theory	199
IV.3.2.1 Lorenz curves for power laws	199
IV.3.2.2 Concentration measures for power laws	205
IV.3.3 A characterization of Price's law of concentration in terms of Lotka's law and of Zipf's law	214

IV.4 Concentration theory of linear three-dimensional informetrics	218
IV.4.1 The concentration of positively reinforced IPPs	219
IV.4.2 Concentration properties of Type/Token-Taken informetrics	226
Chapter V Lotkaian Fractal Complexity Theory	231
V.1 Introduction	231
V.2 Elements of fractal theory	232
V.2.1 Fractal aspects of a line segment, a rectangle and a parallelepiped	233
V.2.2 The triadic von Koch curve and its fractal properties. Extension to general self-similar fractals	234
V.2.3 Two general ways of expressing fractal dimensions	236
V.2.3.1 The Hausdorff-Besicovitch dimension	236
V.2.3.2 The box-counting dimension	239
V.3 Interpretation of Lotkaian IPPs as self-similar fractals	242
Chapter VI Lotkaian Informetrics of Systems in which Items can have Multiple Sources	247
VI.1 Introduction	247
VI.2 Crediting systems and counting procedures for sources and "super sources" in IPPs where items can have multiple sources	253
VI.2.1 Overview of crediting systems for sources	254
VI.2.1.1 First or senior author count	254
VI.2.1.2 Total author count	254
VI.2.1.3 Fractional author count	255
VI.2.1.4 Proportional author count	255
VI.2.1.5 Pure geometric author count	255
VI.2.1.6 Noblesse Oblige	256
VI.2.2 Crediting systems for super sources	256

VI.2.3 Counting procedures for super sources in an IPP	256
VI.2.3.1 Total counting	257
VI.2.3.2 Fractional counting	258
VI.2.3.3 Proportional counting	258
VI.2.4 Inequalities between $Q_T(c)$ and $Q_F(c)$ and consequences for the comparison of $Q_T(c)$, $Q_F(c)$ and $Q_P(c)$	261
VI.2.5 Solutions to the anomalies	266
VI.2.5.1 Partial solutions	267
VI.2.5.2 Complete solution to the encountered anomalies	269
VI.2.6 Conditional expectation results on $Q_T(c)$, $Q_F(c)$ and $Q_P(c)$	270
 VI.3 Construction of fractional size-frequency functions based on two dual Lotka laws	276
VI.3.1 Introduction	276
VI.3.2 A continuous attempt: $z \in \mathbb{R}^+$	278
VI.3.3 A rational attempt: $q \in \mathbb{Q}^+$	282
 Chapter VII Further Applications in Lotkaian Informetrics	295
 VII.1 Introduction	295
 VII.2 Explaining "regularities"	297
VII.2.1 The arcs at the end of a Leimkuhler curve	297
VII.2.2 A "type/token-identity" of Chen and Leimkuhler	298
 VII.3 Probabilistic explanation of the relationship between citation age and journal productivity	300

VII.4 General and Lotkaian theory of the distribution of author ranks in multi-authored papers	304
VII.4.1 General theory	304
VII.4.2 Modelling the author rank distribution using seeds	308
VII.4.3 Finding a seed based on alphabetical ranking of authors	310
VII.5 The first-citation distribution in Lotkaian informetrics	313
VII.5.1 Introduction	313
VII.5.2 Derivation of the model	317
VII.5.3 Testing of the model	320
VII.5.3.1 First example: Motylev (1981) data	320
VII.5.3.2 Second example: JACS to JACS data of Rousseau	322
VII.5.4 Extensions of the first-citation model	323
VII.6 Zipfian theory of N-grams and of N-word phrases: the Cartesian product of IPPs	326
VII.6.1 N-grams and N-word phrases	326
VII.6.2 Extension of the argument of Mandelbrot to 2-word phrases	329
VII.6.3 The rank-frequency function of N-grams and N-word phrases based on Zipf's law for $N = 1$	333
VII.6.4 The size-frequency function of N-grams and N-word phrases derived from Subsection VII.6.3	347
VII.6.5 Type/Token averages μ_N and Type/Token-Taken averages μ_N^* for N-grams and N-word phrases	352
Appendix	365
Appendix I	365
Appendix II	370

Appendix III Statistical determination of the parameters in the	
law of Lotka	372
A.III.1 Statement of the problem	372
A.III.2 The problem of incomplete data (samples) and Lotkaian informetrics	373
A.III.3 The difference between the continuous Lotka function and the discrete Lotka function	378
A.III.4 Statistical determination of the parameters K , a , n_{\max} in the discrete Lotka function K/n^a , $n = 1, \dots, n_{\max}$	386
A.III.4.1 Quick and Dirty methods	387
A.III.4.2 Linear Least Squares method	388
A.III.4.3 Maximum Likelihood Estimating method	390
A.III.5 General remarks	393
A.III.5.1 Fitting Zipf's function	393
A.III.5.2 The estimation of ρ_m and n_{\max}	394
A.III.5.3 Fitting derived functions such as Price's law	394
A.III.5.4 Goodness-of-fit tests	395
Bibliography	397
Subject Index	423