Computational Statistics & Data Analysis 19 (1995) 461–473
North-Holland

Unified representation of Zipf distributions

Peter Zörnig and Gabriel Altmann
Department of Linguistics, University of Bochum, Bochum, Germany

Received February 1993
Revised October 1993

Abstract: Certain discrete probability distributions, used independently from each other in linguistics and other sciences, can be considered as special cases of the distribution based on the Lerch zeta function. We will list the probability functions for some of the most important cases. Moments and estimators are derived for the general Lerch distribution.

Keywords: Lerch zeta function, Zipf distributions, Estimators, Nonlinear equation systems

1. Introduction

Certain generalizations of the Estoup model, often called "Zipf distributions", have proved to be useful in linguistics. Numerous publications on distributions of this type have appeared in different fields of linguistics. Zipf distributions arose from very heterogeneous approaches using linguistic as well as nonlinguistic argumentations (cf. the references).

In the present paper we state a unified representation of the above discrete distributions using the Lerch zeta function [12, p. 27–30], defined by

$$\Phi(p, a, c) = \sum_{x=1}^{\infty} \frac{p^x}{(a+x)^c}. \tag{1.1}$$

In Section 2 certain Zipf distributions are listed, considered always as special cases of the Lerch distribution. The general moments of the Lerch distribution are considered in Section 3. In Sections 4 and 5 we derive estimators and illustrate them by means of empirical frequencies.

Correspondence to: Dr. P. Zörnig, Rohdesdiek 45, 44357 Dortmund, Germany.

0167-9473/95/$09.50
SSDI 0167-9473(94)00009-8

2. The Lerch distribution and some of its special cases

1. The most general distribution of Zipf's type, following immediately from (1.1), has the probability function

$$f(x) = \frac{p^x}{T(a+x)^c}, \quad x = 1, 2, 3, \ldots, \tag{2.1a}$$

with

$$T = \sum_{i=1}^{\infty} \frac{p^i}{(a+i)^c}. \tag{2.1b}$$

Using (1.1) we can represent T as

$$T = \Phi(p, a, c).$$

The probability generating function (pgf) is

$$G(t) = \sum_{x=1}^{\infty} p_x t^x = \frac{1}{T}\Phi(pt, a, c). \tag{2.2}$$

The distribution (2.1) has not been considered in the literature; we will call it the *Lerch distribution*. Its right-truncated variant is

$$f(x) = \frac{p^x}{T(a+x)^c}, \quad x = 1, \ldots, n, \tag{2.3a}$$

with

$$T = \Phi(p, a, c) - p^n \phi(p, a+n, c). \tag{2.3b}$$

The pgf here is

$$G(t) = \frac{1}{T}\left[\Phi(pt, a, c) - (pt)^n \phi(pt, a+n, c)\right]. \tag{2.4}$$

We will now state some important special cases of (2.1) and (2.3) which have been intensively studied in the literature.

2. For $p = 1$, $a = 0$, $c = 1$, we obtain the *Estoup distribution* [13,59], existing only in truncated form, since the harmonic series is not convergent:

$$f(x) = \frac{1}{Tx}, \quad x = 1, 2, \ldots, n, \tag{2.5a}$$

with

$$T = \sum_{x=1}^{n} \frac{1}{x}. \tag{2.5b}$$

3. Choosing $p = 1$, $a = 0$, $c = 2$ yields the *Lotka distribution* [1,2,4,5,9,14,22, 31,35,39,40–42,45–47,51–53,61]:

$$f(x) = \frac{1}{Tx^2}, \quad x = 1, 2, \ldots, \tag{2.6a}$$

with

$$T = \Phi(1, 0, 2) = \tfrac{1}{6}\pi^2. \tag{2.6b}$$

4. For $p = 1$, $a = 0$, $c > 1$ we obtain the zeta distribution, known also as the *Joos model* or the *discrete Pareto distribution* [6,7,10,16,18,23,25–28,30,37,43,44, 48–50,54,57–60,62]:

$$f(x) = \frac{1}{Tx^c}, \quad x = 1, 2, \ldots, \tag{2.7a}$$

with

$$T = \Phi(1, 0, c). \tag{2.7b}$$

5. The *Zipf–Mandelbrot distribution* [3,17,20,32–34,36,55,56,62–65] is obtained for $p = 1$, $a > 0$, $c > 1$:

$$f(x) = \frac{1}{T(a+x)^c}, \quad x = 1, 2, \ldots, \tag{2.8a}$$

with

$$T = \Phi(1, a, c). \tag{2.8b}$$

The distributions (2.5)–(2.8) have been mainly used for describing ranking problems in linguistics with x as a given rank and $f(x)$ as the relative frequency of the given unit at rank x. They are also useful for describing the frequency structure of a set of units e.g. words in texts. In documentation and scientometrics they are used in the study of publication activity, citation frequency etc.

6. The *Good distribution* [11,19,24,29] arises from $0 < p < 1$, $a = 0$, $c \in \mathbb{R}$:

$$f(x) = \frac{p^x}{Tx^c}, \quad x = 1, 2, \ldots, \tag{2.9a}$$

with

$$T = \Phi(p, 0, c). \tag{2.9b}$$

This model has been used in linguistics as the distribution of word frequences.

7. For $0 < p < 1$, $a = 0$, $c = 1$ we obtain the *logarithmic distribution* (cf. [38] for extensive literature):

$$f(x) = \frac{p^x}{Tx}, \quad x = 1, 2, \ldots, \tag{2.10a}$$

with

$$T = \Phi(p, 0, 1) = -\ln(1-p). \tag{2.10b}$$

The logarithmic distribution attracts especially the attention of biologists in the investigation of the distribution of species and genera.

8. The *geometric distribution* [38] can be considered as a Lerch distribution as well for $a \in \mathbb{R}$, $0 < p < 1$, $q = 1 - p$ and $c = 0$:

$$f(x) = \frac{p^x}{T} = qp^{x-1}, \quad x = 1, 2, \ldots, \tag{2.11a}$$

with

$$T = \Phi(p, 0, 0) = \frac{p}{q}. \tag{2.11b}$$

The geometric distribution, the best known member of the family, is used e.g. in physics (known as the Furry distribution), in linguistics as a model for the distribution of gaps and ranks, and various modified forms are used in the theory of queucing.

9. Moreover the discrete *rectangular distribution* [38] is obtained for $p = 1$, $a \in \mathbb{R}$, $c = 0$:

$$f(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n. \tag{2.12}$$

10. The only published distribution with a negative parameter c is the *discrete Pearson distribution, type III* [21,38]:

$$f(x) = A(x - 1 + B)^c \, e^{-b(x-1)}, \quad x = 1, 2, \ldots \tag{2.13}$$

Using the notations $T := (A\, e^b)^{-1}$, $a := B - 1$, $p := e^{-b}$ yields:

$$f(x) = \frac{(x + a)^c p^x}{T}, \quad x = 1, 2, \ldots, \tag{2.14a}$$

with

$$T = \Phi(p, a, -c). \tag{2.14b}$$

Other distributions, permitting a negative c, are the truncated zeta distribution, the truncated Zipf–Mandelbrot distribution and both variants of the Good distribution.

The respective parameter values for each of the above distributions are listed in Table 1.

Table 1
Special cases of the Lerch distribution

distribution	parameters		
	p	a	c
Estoup	1	0	1
Lotka	1	0	2
zeta	1	0	c
Zipf-Mandelbrot	1	a	c
Good	p	0	c
logarithmic	p	0	1
geometric	p	–	0
rectangular			
Pearson	p	a	$c < 0$

Some of the distributions in Table 1 can also be represented using other functions, e.g. the generalized Riemann zeta function, the bigamma or trigamma function or the hypergeometric function.

3. Moments

The general moments of the Lerch distribution can be represented in terms of the Lerch zeta function:

$$\begin{aligned}\mu_r^c &= \sum_{x=1}^{\infty} \frac{(x-c)^r p^x}{T(a+x)^c}\\ &= \frac{1}{T}\sum_{x=1}^{\infty}(x+a-c-a)^r \frac{p^x}{(a+x)^c}\\ &= \frac{1}{T}\sum_{x=1}^{\infty}\sum_{i=1}^{r}\binom{r}{i}(x+a)^i(-c-a)^{r-i}\frac{p^x}{(a+x)^c}\\ &= \frac{1}{T}\sum_{i=1}^{r}\binom{r}{i}(-c-a)^{r-i}\sum_{x=1}^{\infty}\frac{p^x}{(a+x)^{c-i}}\\ &= \frac{1}{T}\sum_{i=1}^{r}\binom{r}{i}(-c-a)^{r-i}\Phi(p, a, c-i).\end{aligned} \tag{3.1}$$

The moments of the truncated case are analogous:

$$\begin{aligned}\mu_r^c &= \frac{1}{T}\sum_{i=1}^{r}\binom{r}{i}(-c-a)^{r-i}\sum_{x=1}^{n}\frac{p^x}{(a+x)^{c-i}}\\ &= \frac{1}{T}\sum_{i=1}^{r}\binom{r}{i}(-c-a)^{r-i}\cdot[\Phi(p, a, c-i)-p^n\Phi(p, a+n, c-i)].\end{aligned} \tag{3.2}$$

4. Estimation by means of empirical frequencies

In this section we will derive estimators for the parameters of the general Lerch distribution (2.1). We will confine ourselves to fitting the Lerch distribution to data sets with "very fast" decreasing frequencies (cf. Section 5).

Estimators based on probability classes can be derived as follows. We choose the parameters p, a, c such that

$$\hat{p}_i = \frac{p^i}{(a+i)^c} \tag{4.1}$$

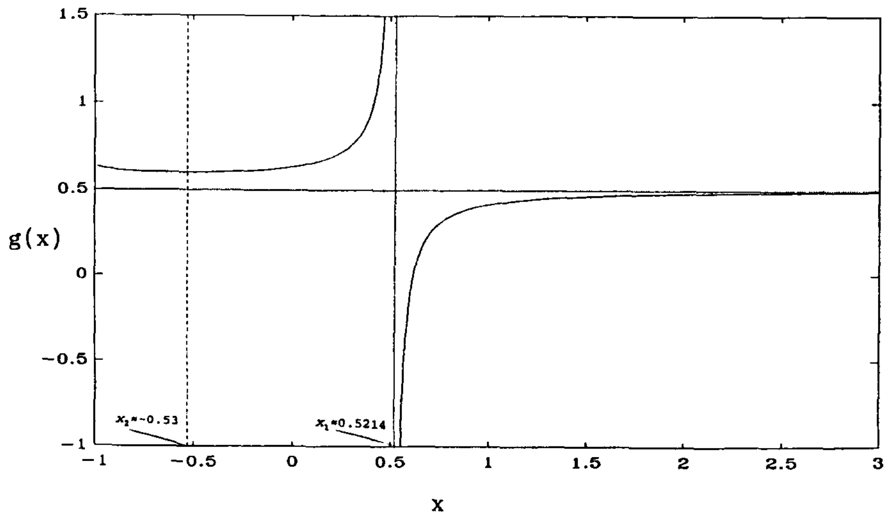

Fig. 1. Illustration of the function g in (4.2).

holds for $i = 1, 2$ or for $i = 1, 2, 3$, where the numbers $\hat{p}_1$ denote the relative empirical frequencies. On grounds of the above assumption on the empirical data, (4.1) "normally" implies

$$T = \sum_{i=1}^{\infty} \frac{p^i}{(a+i)^c} \approx \sum_{i=1}^{\infty} \hat{p}_i = 1$$

(cf. Section 5), i.e. (4.1) insures that the first two or three theoretical probabilities $p_i = p^i / T(a+i)^c$ coincide approximately with the corresponding empirical values $\hat{p}_i$.

In order to solve a nonlinear system of equations of the form (4.1) we define the function g by

$$g(x) = \frac{\ln \frac{(x+1)^2}{x+2}}{\ln \frac{(x+1)^3}{x+3}}, \tag{4.2}$$

for $x > -1$, $x \neq x_1 \approx 0.5214$, where x_1 denotes the unique zero of the denominator in (4.2) (cf. Fig. 1). Let further $x_2 \approx -0.53$ denote the relative minimum of g (for $x < x_1$).

Using the rule of the l'Hôspital it can be shown that $\lim_{x \to -1} g(x) = \frac{2}{3}$ and $\lim_{x \to \infty} g(x) = \frac{1}{2}$.

Theorem 1. *Let p_1, p_2, p_3 be any positive real numbers with $p_2 \neq p_1^2$, $p_3 \neq p_1^3$. Let the function g and the number x_2 be defined as above and*

$$Q := \frac{\ln \frac{p_2}{p_1^2}}{\ln \frac{p_3}{p_1^3}}.$$

The system of equations

$$p_i = \frac{p^i}{(a+i)^c} \quad (i = 1, 2, 3), \tag{4.3}$$

with variables $p \in \mathbb{R}$, $a > -1$, $c \neq 0$ *is unsolvable for*

$$0.5 \leqslant Q < g(x_2) \approx 0.598. \tag{4.4}$$

If (4.4) *does not hold, any solution of* (4.3) *is given by*

$$a = g^{-1}(Q),$$
$$c = \frac{\ln \frac{p_2}{p_1^2}}{\ln \frac{(a+1)^2}{a+2}}, \tag{4.5}$$
$$p = p_1(a+1)^c.$$

Proof. The assertion follows from the following equivalent transformations of system (4.3), where the above conditions on p_1, p_2, p_3 insure that all occurring logarithms are $\neq 0$:

Substituting $p = p_1(a+1)^c$ for p in the second and the third equation yields the system

$$p = p_1(a+1)^c, \quad p_2 = p_1^2\left(\frac{(a+1)^2}{a+2}\right)^c, \quad p_3 = p_1^3\left(\frac{(a+1)^3}{a+3}\right)^c,$$

which is equivalent to

$$p = p_1(a+1)^c, \quad c = \frac{\ln \frac{p_2}{p_1^2}}{\ln \frac{(a+1)^2}{a+2}}, \quad c = \frac{\ln \frac{p_3}{p_1^3}}{\ln \frac{(a+1)^3}{a+3}}.$$

Substituting the right side of the second equation for c in the last equation yields

$$p = p_1(a+1)^c, \quad c = \frac{\ln \frac{p_2}{p_1^2}}{\ln \frac{(a+1)^2}{a+2}}, \quad Q = g(a).$$

It follows that the system (4.3) has exactly one solution for $Q < \frac{1}{2}$, $Q = g(x_2) \approx 0.598$ or $Q \geqslant \frac{2}{3}$ (since $g(x) = Q$ has exactly one solution in these cases, cf. Fig. 1). For $g(x_2) < Q < \frac{2}{3}$ exactly two solutions exist for (4.3). In the remaining cases, i.e. for $\frac{1}{2} \leqslant Q < g(x_2)$, no solution exists. □

Theorem 2. *Let p_1, p_2 be positive real numbers with $p_2 \neq p_1^2$, $p_2 < p_1$.*

(a) *Any solution of the system*

$$p_i = \frac{p^i}{(a+i)^c}, \quad (i = 1, 2;\ a > -1)$$

has the form

$$a > -1, \qquad c = \frac{\ln \frac{p_2}{p_1^2}}{\ln \frac{(a+1)^2}{a+2}}, \qquad p = p_1(a+1)^c. \tag{4.6}$$

(b) *The (continuous) solution function*

$$f(x) = \frac{p^x}{(a+x)^c} \tag{4.7}$$

(p, a, c as in (4.6)) is strictly monotone decreasing for $x > 1$ if

$$h(a) := \frac{1}{\alpha \ln \frac{(a+1)^2}{a+2} + \ln(a+1)} - a \leqslant 1$$

$$\left(\alpha := \frac{\ln p_1}{\ln \frac{p_2}{p_1^2}};\quad cf.\ Fig.\ 2\right) \tag{4.8}$$

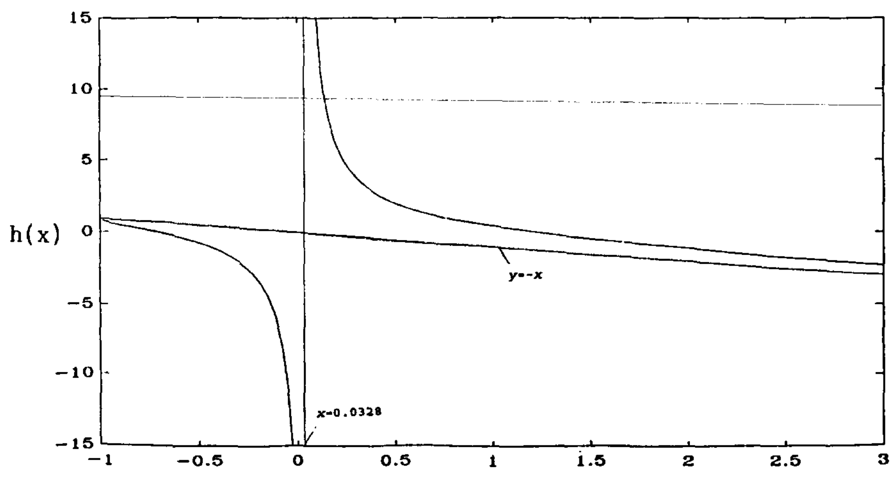

Fig. 2. Illustration of the function h in (4.8) for $\alpha = 0.05$.

Proof. Part (a) is analogous to the proof in Theorem 1. (b) By means of elementary analysis we can show that

$$f'(x) = 0 \quad \Leftrightarrow \quad x = \frac{c}{\ln p} - a.$$

Using (4.6), the last equation can be transformed as follows:

$$x = \frac{c}{\ln p_1 + c \ln(a+1)} - a,$$

$$x = \frac{1}{\dfrac{\ln p_1}{c} + \ln(a+1)} - a,$$

$$x = \frac{1}{\alpha \ln \dfrac{(a+1)^2}{a+2} + \ln(a+1)} - a = h(a).$$

The above transformations show that $f'(x)$ equals zero if $x = h(a)$. Thus f is monotone decreasing for $x > 1$ if $h(a) \leqslant 1$. □

The preceding theorems yield two possible ways to compute estimators. On the one hand we can choose p, a, c as a solution of (4.5), where p_1, p_2, p_3

Table 2
Estimators and optimal parameters in the fit of the Lerch distribution to an observed surname distribution

District	Estimators			Optimal parameters		
	p	a	c	p	a	c
1	0.4431	−0.5	0.8126	1	0.4097	3.2736
2	0.4126	−0.5	1.0465	1	1.6957	6.3904
3	0.3753	−0.5	1.2419	0.5488	−0.3143	1.9424
4	0.3381	−0.5	1.4340	0.5156	−0.6918	1.3816
5	0.3557	−0.5	1.3447	0.2437	−0.9998	0.1246
6	0.3541	−0.5	1.3331	0.3306	−0.9999	0.1464
7	0.3615	−0.5	1.3075	0.4023	−0.8826	0.6879
8	0.4006	−0.5	1.0930	0.2971	−0.9996	0.1116

District	X^2	prob./ES	d.f.	$h(a)$	T
1	9.5056	0.30	8	−0.4984	0.9830
2	1.4076	0.24	1	−0.6823	1.0022
3	0.0221	0.88	1	−0.7671	0.9954
4	$3.8 * 10^{-6}$	0.0001	0	−0.8224	0.9907
5	0.0267	0.0102	0	−0.8010	0.9940
6	1.1928	0.27	1	−0.7840	0.9822
7	0.2692	0.60	1	−0.7852	0.9906
8	0.4311	0.51	1	−0.6948	0.9907

denote the empirical relative frequencies. This alternative is not always suitable, since (4.5) can be inconsistent and the solution function $f(x)$ is not always monotone decreasing. On the other hand we can choose p, a, c as given in (4.6). If a satisfies the condition $h(a) < 1$ (cf. Theorem 2(b); this is always possible since $h(a)$ is monotone decreasing for sufficiently large a), equations (4.6) yield a monotone decreasing solution function $f(x)$.

5. Fitting the model to empirical data

In this section we will fit the Lerch distribution (2.1) to empirical data and illustrate the estimation in Section 4. Therefore we use the observed data in Panaretos [37] who studied the distributions of surnames in eight districts and tried to fit the Yule and the zeta distribution to these data.

Table 3
Observed (upper entries) and expected (lower entries) frequencies f(x) of the occurrences of surnames

x	District							
	1	2	3	4	5	6	7	8
1	832	329	292	243	234	281	349	282
	829.87	328.29	291.99	243.00	233.57	280.45	348.89	281.62
2	151	43	28	17	19	23	30	34
	143.48	43.713	27.925	17.002	19.695	24.074	29.792	34.941
3	39	11	6	4	5	9	7	11
	46.057	9.4620	6.2016	4.0004	4.4027	7.1909	7.7213	9.6083
4	20	1	2	2	0	1	3	2
	19.845	2.7553	1.8404	1.2544	1.0201	2.2403	2.3806	2.7284
5	11	0	0	0	1	0	1	0
	10.164	0.9801	0.6336	0.4490	0.2398	0.7101	0.7909	0.7850
6	2	1	0	0	0	0	0	0
	5.8335	0.4027	0.2388	0.1735	0.0568	0.2272	0.2740	0.2275
7	4	0	1	0	0	0	0	0
	3.6292	0.1845	0.0957	0.0705	0.0135	0.0731	0.0975	0.0662
8	5	0	0	0	0	1	0	0
	2.3979	0.0920	0.0401	0.0297	0.0032	0.0236	0.0353	0.0193
9	0	1	0	0	0	0	0	0
	1.6600	0.0491	0.0173	0.0128	0.0008	0.0077	0.0130	0.0057
10	1	0	0	0	0	0	0	0
	1.1926	0.0278	0.0077	0.0056	0.0002	0.0025	0.0048	0.0017
11	0	0	0	0	0	0	0	0
	0.8833	0.0164	0.0035	0.0025	0.0000	0.0008	0.0018	0.0005
12	2	0	0	0	0	0	0	1
	0.6709	0.0101	0.0016	0.0011	0.0000	0.0003	0.0007	0.0001
$\geqslant$	2	0	0	0	0	0	0	0
13	3.3099	0.0210	0.0014	0.0010	0.0000	0.0001	0.0004	0.0001

The empirical data are listed in Table 3 (upper entries) where $f(x)$ denotes the number of surnames occurring x times in the given district. We have computed the estimators according to (4.6). Practical tests with the present data have shown that $a = -0.5$ yields good starting values ($h(a) \leqslant 1$ is satisfied in particular; cf. Theorem 2). The estimators p, a, c are listed in Table 2. Starting with these values we minimized the chi-square value by means of the optimization method of Nelder and Mead and obtained the optimal parameters in the right part of Table 2. The chi-square values for the Lerch distribution with optimal parameters, the corresponding probabilities and the numbers of degrees of freedom (d.f.) are also contained in Table 2. We have set $p = 1$ for the districts 1 and 2, since the optimal parameters obtained by the procedure of Nelder and Mead lead to probability functions which are not monotone decreasing. For the districts 4 and 5 we have computed Cohen's [8] effect size coefficient instead of the chi-square value, because there are 0 degrees of freedom in these cases.

The "monotony criterion" $h(a)$ and the number T (cf. (2.1b) and the introductory remarks in Section 4) are listed in the last two columns. The theoretical frequencies are also listed in Table 3 (lower entries).

The investigations show that the fit of the Lerch distribution is better than in the case of the discrete Pareto distribution or the Yule distribution, used in Panaretos [37].

References

[1] Allison, P.D., D. de S. Price, B.C. Griffith, M.J. Moravcsik and J.A. Stewart, Lotka's law: a problem in its interpretation and application, *Social Studies of Science* **6** (1976) 269–276.

[2] Allison, P.D. and J.A. Stewart, Productivity differences among scientists: Evidence for accumulative advantage, *Amer. Sociol. Review* **39** (1974) 596–606.

[3] Arapov, M.V., E.N. Efimova and J.A. Shreyder, O smysle rangovykh raspredeleniy, *Nauchno-tekhnicheskaya informacia Ser. 2, Nr.* **1** (1975) 9–20.

[4] Bookstein, A., The bibliometric distributions, *Library Quarterly* **46** (1976) 416–423.

[5] Bookstein, A., Patterns of scientific productivity and social change: A discussion of Lotka's law and bibliometric symmetry, *J. Amer. Soc. Inform. Sci.*, **28** (1977) 206–210.

[6] Brookes, B.C., The derivation and application of the Bradford–Zipf distribution, *J. Document.* **24** (1968) 247–265.

[7] Chen, W-Ch., On the weak form of the Zipf's law, *J. Appl. Probab.* **17** (1980) 611–622.

[8] Cohen, J., *Statistical power analysis for the behavioral sciences* (Academic Press, New York, 1977).

[9] Coile, R.C., Lotka's frequency distribution of scientific productivity, *J. Amer. Soc. Inform. Sci.*, **28** (1977) 366–370.

[10] Cressie, N. Empirical Bayes estimation for discrete distributions. *South African Statist. J.*, **16** (1982) 25–37.

[11] Engen, S., *Stochastic abundance models* (Chapman and Hall, London, 1978).

[12] Erdélyi, A. et al. (eds.), *Higher transcendental functions I* (McGraw-Hill, New York, 1953).

[13] Estoup, J.B., Les gammes sténographiques (Paris, Institut Sténographique, Paris, 1916).

[14] Fairthorne, R.A., Empirical hyperbolic distributions (Bradford–Zipf–Mandelbrot) for bibliometric description and prediction, *J. Document.* **25** (1969) 319–349.

[15] Feller, W., *An introduction to probability theory and its applications*, Vol. 1 (Wiley, New York, 1962(7)).
[16] Fox, W.R., G.W. Lasker, *The distribution of surname frequencies, Intern. Statist. Rev.* **51** (1983) 81–87.
[17] Frumkina, R.M., K voprosu o tak nazyvaemom zakone Cipfa, *Voprosy Jazykoznania, Nr.* **2** (1961) 117–122.
[18] Gani, J., A sampling problem with application to the type and token relationship, *Utilitas Mathematica* 21 B (1982) 333–340.
[19] Good, I.J., The population frequencies of species and the estimation of population parameters, *Biometrika* **40** (1953) 237–264.
[20] Guiter, H., M.V. Arapov (eds.), *Studies on Zipf's law* (Brockmeyer, Bochum, 1982).
[21] Haight, F.A., Queueing with balking, *Biometrika* **44** (1957) 360–369.
[22] Haitun, S.D., Stationary scientometric distributions Part I. Different approximations, *Scientometrics* **4** (1982) 5–25.
[23] Herdan, G., *Language as choice and chance* (Nordhoff, Groningen, 1956).
[24] Herdan, G., The relation between the dictionary distribution and the occurrence distribution of word length and its importance for the study of quantitative linguistics, *Biometrika* **5** (1958) 222–228.
[25] Hill, B., Zipf's law and prior distributions for the composition of a population, *J. Amer. Statist. Assoc.* **65** (1970) 1220–1232.
[26] Hill, B. The rank-frequency form of Zipf's law, *J. Amer. Statist. Assoc.* **69** (1974) 1017–1026.
[27] Hill, B., B. Woodroofe, Stronger forms of Zipf's law, *J. Amer. Statist. Assoc.* **70** (1975) 212–219.
[28] Hubert, J.J., General bibliometric models, *Library trends* **30** (1981) 65–81.
[29] Ijiri, Y., H.A. Simon, *Skew distributions and the sizes of business firms* (North-Holland, Amsterdam, 1977).
[30] Johnson, N.L., S. Kotz, *Discrete distributions* (Houghton, Mifflin, Boston, MA, 1969).
[31] Lotka, A.J., Théorie analytique des associations biologiques II, *Actualités scientifique et industrielles* (1939) 780.
[32] Mandelbrot, B. An information theory of the statistical structure of language, in: W. Jackson (ed.), *Communication theory* (Academic Press, New York, 1953) pp. 503–512.
[33] Mandelbrot, B. A note on a class of skew distribution functions. Analysis and critique of a paper of H.A. Simon, *Inform. Control* **2**, 90–99.
[34] Miller, G.A., Some effects of intermittent silence, *Amer. J. Psych.* **70** (1957) 311–314.
[35] Murphy, L.J. Lotka's law in the humanities? *J. Amer. Soc. Inform. Sci.* **24** (1973) 460–462.
[36] Orlov, J.K., M.G. Boroda, I.Sh. Nadareishvili, *Sprache, Text, Kunst. Quantitative Analysen* (Brockmeyer, Bochum, 1982).
[37] Panaretos, J. On the evolution of surnames. *Intern. Statist. Rev.* 57/2 (1989) 161–179.
[38] Patil, G.P., M.T. Boswell, S.W. Joshi, M.W. Ratnaparkhi, *Dictionary and classified bibliography of statistical distributions in scientific work. Volume I: Discrete models* (International Cooperative House, Fairland, MD, 1984).
[39] Platz, A., Psychology of the scientist: XI. Lotka's law and research visibility, *Psychol. Reports* **16** (1965) 566–568.
[40] Potter, W.G., Lotka's law revisited. *Library Trends* **30** (1981) 21–39.
[41] Radhakrishnan, T., R. Kerinzan, Lotka's law and computer science literature, *J. Amer. Soc. Inform. Sci.* **30** (1979) 51–54.
[42] Rao, I.K.R., The distribution of scientific productivity and social change, *J. Amer. Soc. Inform. Sci.* **31** (1980) 111–122.
[43] Rider, P.R., The zeta distribution, in: G.P. Patil, (ed.), *Classical and contagious discrete distributions* (Statistical Publishing House and New York, Calcutta Pergamon, 1965). pp. 443–444.
[44] Scarrott, G., Will Zipf join Gauss? *New Scientist* **62** (1974) 402–404.

[45] Schorr, A.E., Lotka's law and the history of legal medicine, *Research in Librarianship* **30** (1975a) 205–209.
[46] Schorr, A.E., Lotka's law and map librarianship, *J. Amer. Soc. Inform. Sci* **26** (1975b) 189–190.
[47] Schubert, A., W. Glänzel, A dynamic look at a class of skew distributions. A model with scientometric applications, *Scientometrics* **6** (1984) 149–167.
[48] Seal, H.L., A probability distribution of death at age x when policies are counted instead of lives, *Skandinavisk Aktuarietidskrift* **30** (1947) 18–43.
[49] Seal, H.L., The maximum likelihood fitting of the discrete Pareto law, *J. Instit. of Actuaries* **78** (1952) 115–121.
[50] Simon, H.A., On a class of skew distribution functions, *Biometrika* **42** (1955) 425–440.
[51] Vlachy, J., Variable factors in scientific communities (observations on Lotka's law), *Teorie a Metoda* **4** (1972) 91–120.
[52] Vlachy, J., Time factors in Lotka's law, *Probleme de Informare si Documentare* **10** (1976) 44–87.
[53] Vlachy, J., Frequency distributions of scientific performance. A bibliography of Lotka's law and related phenomena, *Scientometrics* **1** (1978) 109–130.
[54] Walther, A., Anschauliches zur Riemannschen Zetafunktion, *Acta Mathematica* **48** (1926) 393–400.
[55] Wen Chen, On Zipf's law, Dissed. (University of Michigan, 1978).
[56] Wen Chen, On the weak form of Zipf's law, *J. Appl. Probab.* **17** (1980) 611–622.
[57] Williams, C.B., A note on the statistical analysis of sentence length as a criterion of literary style, *Biometrika* **31** (1940) 356.
[58] Woodroofe, B., B. Hill, On Zipf's law, *J. Appl. Probab.* **12** (1975) 425–434.
[59] Woronczak, J., On an attempt to generalize Mandelbrot's distribution, in: *To Honor Roman Jakobson, Vol. 3.* (Mouton, The Hague, Neth., 1967) pp. 2254–2268.
[60] Wyllys, R.E., Empirical and theoretical bases of Zipf's law, *Library Trends* **30** (1981) 53–64.
[61] Yablonsky, A.I., On fundamental regularities of the distribution of scientific productivity, *Scientometrics* **2** (1980) 3–34.
[62] Zipf, G.K., *Human behavior and the principle of least effort* (Addison-Wesley, Reading, MA, 1949).
[63] Zörnig, P., G. Altmann, The repeat rate of phoneme frequencies and the Zipf–Mandelbrot Law, *Glottometrika* **5** (Brockmeyer, Bochum, 1983) pp. 205–211.
[64] Zörnig, P., G. Altmann, The entropy of phoneme frequencies and the Zipf–Mandelbrot Law, *Glottometrika* **6** (Brockmeyer, Bochum, 1984) pp. 41–47.
[65] Zörnig, P., M. Boroda, The Zipf–Mandelbrot law and the interdependencies between frequency structure and frequency distribution in coherent texts, *Glottometrika* **13** (Brockmeyer, Bochum, 1992) pp. 205–218.