Zipf's law against the text size: A half-rational model

Łukasz Dębowski
Institute of Computer Science, Polish Academy of Sciences
ldebowsk@ipipan.waw.pl, http://www.ipipan.waw.pl/~ldebowsk/

Abstract

In this article, we consider Zipf-Mandelbrot law as applied to texts in natural languages. We present a simple model of dependence of the law on the text size, which is featured by variable power-law tail and constant ratio of the most frequent words. As a result we derive several closed formulas, which accord with empirical data qualitatively and partially quantitatively. For example, there appears to be a minimal length of literary texts equal to ≈ 159 word tokens for English.

Keywords: Zipf's law

1 Introduction

For a definite object in which we can identify tokens and count them as instances of some identifiable types, Zipf's law (Zipf, 1935, 1949) is a statement that the frequency $f(w)$ of all tokens belonging to given type w is roughly inversely proportional to rank $r(w)$ of the type,

$$f(w) \approx \frac{\text{const}}{r(w)}. \tag{1}$$

Rank $r(w)$ is defined as the ordinal number of w on the list of all empirical types sorted in descending order according to $f(w)$.

Zipf's law forms a beautiful example of quasi-interdisciplinary empirical regularity which possesses the following features:

1. Regularity is observed in data resuming phenomena investigated in various scientific disciplines. Examples are biology (Camacho & Solé, 1999), economics (Pareto, 1897), linguistics (Estoup, 1916), physics (Tsallis, 2000). (In non-linguistic applications, rank $r(w)$ can be proportional to some simple variable describing physical size or magnitude of w, such as income in the distribution of personal incomes.)

2. Regularity inspires a multitude of half-explanations introducing assumptions which do not seem to be so universal as the regularity itself. Examples are random-typing text model (Belevitch, 1956; Li, 1992), effects of artificial ranking of sample taken from distribution with large variance (Günther, Levitin, Schapiro, & Wagner, 1996), eco-system's dynamics (Camacho & Solé, 1999), fractal vocabulary model (Mandelbrot, 1983), new thermodynamic formalism (Denisov, 1997), sampling of LNRE distributions (large number of rare events) (Khmaladze, 1987; Baayen, 2001),

information theoretic models of language learning (Harremoës & Topsøe, preprint; Dębowski, 2002).

3. Regularity is roughly described by a simple mathematical formula, but domain specific investigations usually uncover many finer significant departures. Examples include initial bend and final power law (Mandelbrot, 1954), better fit of tail by the inverse-square distribution of frequencies of frequencies (Kornai, 1999) (also independently observed by us and by M. Montemurro), another bend for very large ranks (Montemurro, 2001) (also observed in economical data), parameter dependence on the size of texts (Orlov, 1982), or investigated fraction of a text (Baayen, 2001).

In dealing with empirical regularities such as Zipf's law, for which there is no unique, accurate and enchanting theory, one usually follows either of often disjoint ways: empirical or rational. In the empirical approach, one seeks for the formula most tightly interpolating given experimental data, even at the cost of introducing obscure additional parameters and worsening the extrapolations. The rational method consists in deriving the regularity from simpler principles or other empirical facts, even at the cost of worsening the fit of already observed data in comparison to more elaborated but obscure models.

Nonetheless, the possibility of combining empirical and rational approaches arises sometimes. For example, one can complicate a formula in concern to have additional apparently random parameters and to fit better some portion of data, but the formula with the same class of parameters plus some conditions on their variability can fit a much larger scope of data than before the modification. In fact, it is the way how many fundamental theories in natural sciences have been born.

The aim of this article is to present a modest example of combined empirical-rational approach given by a simple model of Zipf's law variability across texts of different length. We will speak only of Zipf's law applied to texts in natural languages. We are going to show that Mandelbrot's modification (2) of formula (1), introducing two unknown parameters, can be perceived as less arbitrary if we let the values of the parameters be linked with the text length according to several common-sense postulates.

2 The model

In quantitative linguistics, Zipf's law (1) is formulated for types w being types of words (word-forms or lemmas) encountered in some finite text. The tokens are occurrences of words at consecutive positions in the text. It is also this text against which both counts $f(w)$ and ranks $r(w)$ are computed. (In the case of words with the same count, we assign them distinct ranks.) It is important to note that language texts treat on various subjects, so rank $r(w)$ of particular word w strongly depends on the particular text. The exception for this rule is a group of words constantly occupying the lowest ranks and identifiable with functional (grammatically auxiliary) words.

Mandelbrot (1954) observed that instead of formula (1), formula

$$f(r) \approx \left\lfloor \left[\frac{V+\rho}{r+\rho} \right]^{1+\epsilon} \right\rfloor , \qquad (2)$$

where we abbreviate $f(r) = f(w(r))$ for $r = r(w)$ and $\lfloor x \rfloor$ is the greatest integer smaller than x, approximates statistics of words better. We have $r \in \{1, 2, ..., V\}$, where V is the size of vocabulary for the given text. Formula (2) fits the whole range of finite-text data better than (1), but there are some departures still (Baayen, 2001). The formula contains also two new parameters to estimate: ϵ and ρ. (For very large texts, $0 < \epsilon \ll 1$ and $0 < \rho < 10$.)

It is important to note that parameters V, ρ, and ϵ depend quite regularly on the size N of the text, i.e. N being the number of word tokens in the text. Especially, $\epsilon < 0$ for $N < N_0$ and $\epsilon > 0$ for $N > N_0$, where N_0 is some characteristic text length, called Zipfian size (Orlov, 1982).

Orlov (1982) described this phenomenon and gave it some mathematical model in terms of interpolation formulas for random sample (urn model, or IID process) drawn from LNRE distribution. See also Khmaladze (1987), Baayen (2001) for more elaborate calculations. We had learned of the article by Orlov (1982) from an article by Sambor (1988). By the time we collected a copy of Orlov (1982), we had found out a very different heuristic model of Zipf's law variability which we introduce here.

If one considers an ensemble of texts of variable size N written in the same language, it is reasonable to assume that the same grammar is obeyed in the whole ensemble. The conservation of grammar across the ensemble may imply the stability of probability estimates for the functional words which occupy constantly the same lowest ranks. Thus

$$\frac{f(r)}{N} \approx \text{const} \quad \text{for} \quad r = 1, 2, ..., K, \tag{3}$$

for the majority of texts, where K is some small natural number and N is assumed to be the length of the text in the ensemble. If $N \to \infty$, however, $f(r)/N \to \text{const}$ for all r. Thus

$$\epsilon \to \text{const} \quad \text{and} \quad \rho \to \text{const} \quad \text{for} \quad N \to \infty. \tag{4}$$

Postulates (3), (4) when applied to (2) can be approximated by the following set of three postulates:

1. There is such $N = N_0$ that $\epsilon = 0$.
2. For all N, it is $f(0)/N = \text{const}$.
3. For all N, it is $f'(0)/N = \text{const}$.

(Formula (2) allows us to define the value $f(0)$ and the derivative $f'(0)$.)

In the further reasoning, we will assume $\rho \gg 1$, despite the empirical data. Using (2), one can compute the number of tokens N in the text as

$$N = \int_0^V f(r)dr \approx \int_0^V \left[\frac{V+\rho}{r+\rho}\right]^{1+\epsilon} dr = \frac{\rho}{\epsilon}\left[\frac{V+\rho}{\rho}\right]^{1+\epsilon}\left[1 - \left[\frac{\rho}{V+\rho}\right]^{\epsilon}\right]. \tag{5}$$

For $N = N_0$, let us write $V = V_0$, $\rho = \rho_0$. Combining postulates 2 and 3, one obtains $f(0)/f'(0) = \text{const}$. Inserting (2) for any N and for $N = N_0$, and preserving terms linear in $1/\rho$ yields

$$\rho = (1+\epsilon)\rho_0 \quad \text{for} \quad \rho \gg 1. \tag{6}$$

Parameter N_0 can be rewritten by means of V_0 and ρ_0 as

$$N_0 \approx \int_0^{V_0} \left[\frac{V_0+\rho_0}{r+\rho_0}\right] dr = \rho_0 \left[\frac{V+\rho_0}{\rho_0}\right] \ln\left[\frac{V_0+\rho_0}{\rho_0}\right]. \tag{7}$$

(Formula (5) can be applied directly for $\epsilon \neq 0$. Formula (7) is its limit for $\epsilon \to 0$.) Using (5), (7), postulate 2 with (2) for any $N \neq N_0$ and for $N = N_0$ gives

$$\rho_0 \ln\left[\frac{\rho_0}{V_0+\rho_0}\right] = \frac{\rho}{\epsilon}\left[\left[\frac{\rho}{V+\rho}\right]^{\epsilon} - 1\right]. \tag{8}$$

It is convenient to define

$$\lambda = 1 - \frac{\epsilon x}{1+\epsilon}, \tag{9}$$

$$x = \ln\left[\frac{V_0+\rho_0}{\rho_0}\right]. \tag{10}$$

Then

$$1+\epsilon = \frac{x}{x-1+\lambda}. \tag{11}$$

Equations (8) and (6) yield

$$\left[\frac{V+\rho}{\rho}\right]^{1+\epsilon} = \left[1 - \left[\frac{\epsilon}{1+\epsilon}\right]\ln\left[\frac{V_0+\rho_0}{\rho_0}\right]\right]^{-\frac{1+\epsilon}{\epsilon}} = [e(\lambda)]^x, \tag{12}$$

where function $e(\lambda)$ is defined as

$$e(\lambda) = \lambda^{1/(\lambda-1)}. \tag{13}$$

Resuming, one obtains

$$f(r) \approx [e(\lambda)]^x \left[\frac{\rho_0\left[\frac{x}{x-1+\lambda}\right]}{r+\rho_0\left[\frac{x}{x-1+\lambda}\right]}\right]^{\left[\frac{x}{x-1+\lambda}\right]}, \tag{14}$$

$$N \approx [e(\lambda)]^x \rho_0 x, \tag{15}$$

$$V \approx \left[\frac{[e(\lambda)]^{(x-1+\lambda)} - 1}{x-1+\lambda}\right]\rho_0 x, \tag{16}$$

where V was computed from property $f(V) = 1$.

In equations (14)–(16), three parameters appear: ρ_0, x and λ. The status of ρ_0 and x is different from λ. Parameters ρ_0 and x, where $\rho_0 > 0$, $x > 1$, should be the properties of a given language, i.e. they should be constant in the whole ensemble of texts in that language. Parameter λ, where $\lambda > 0$, is a function of the size of text N and the two other parameters. Since $e(\lambda) > 1$, our model can be only applied if $N \geq \rho_0 x$.

Both text size N and vocabulary size V are monotonically decreasing functions of λ, For $\lambda \to \infty$, it is $N, V \to N_{\min} = \rho_0 x$. For $\lambda \to 0$, it is $N, V \to \infty$. Value $\lambda = 1$ corresponds to $\epsilon = 0$ and $N_0 = \rho_0 x e^x$. We can see that $\epsilon < 0$ for $N < N_0$ and $\epsilon > 0$ for $N > N_0$ actually holds for our model. (By the way, condition $x > 1$ is necessary since we need to have $f(r) > 0$ and monotonically decreasing w.r.t. r also for $\lambda \to 0$.)

If we would like the model to describe texts of any positive length, $N \geq 1$, we might like to fix $\rho_0 = 1/x$ so that the minimum of N and V be $N_{\min} = \rho_0 x = 1$. Then we have just one global parameter x left and $N_0 = e^x$. In this case, there is a simple relation between the Zipfian size N_0 and the Mandelbrot's exponent $1+\epsilon$ for texts of asymptotically infinite length,

$$1+\epsilon = \frac{\ln N_0}{\ln N_0 - 1}. \tag{17}$$

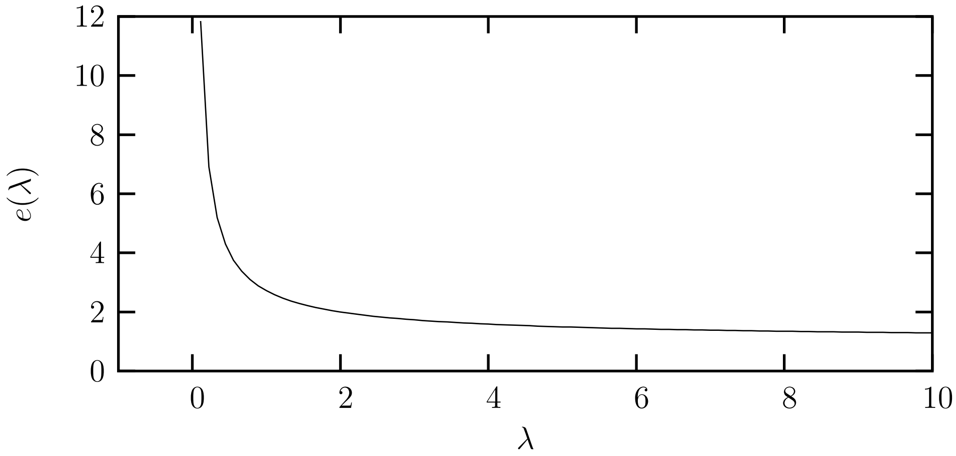

Figure 1: The plot of $e(\lambda) = \lambda^{1/(\lambda-1)}$.

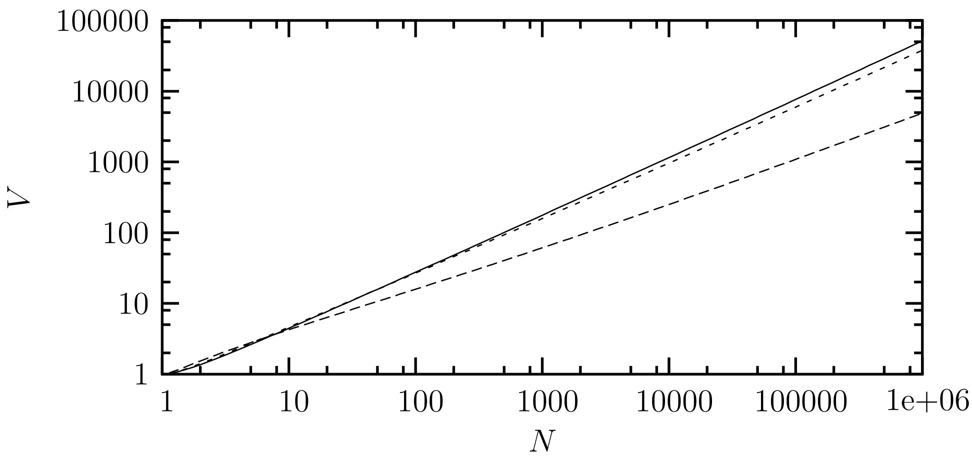

Figure 2: Plots of vocabulary size V as function of text size N for $\rho_0 = 1/x$ and $x = 3.0, 8.0, 10.0$ (curves increasing respectively).

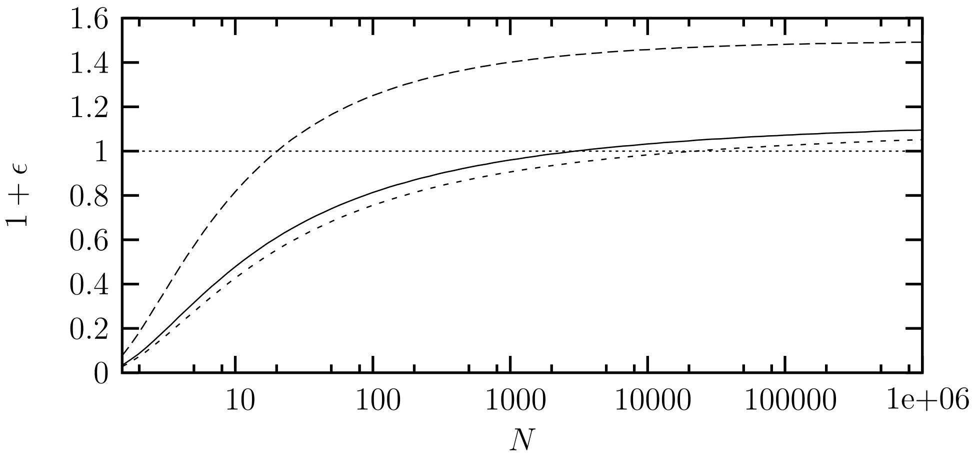

Figure 3: Plots of Mandelbrot exponent $1 + \epsilon$ as function of text size N for $\rho_0 = 1/x$ and $x = 3.0, 8.0, 10.0$ (curves decreasing respectively).

According to Orlov (1982), N_0 for Russian seems to range from 3000 to 20000, so we would get $1+\epsilon$ ranging between 1.143 ($x = 8.006$) and 1.112 ($x = 9.903$) respectively.

In our model, $\log V$ seems to be an almost linear function of $\log N$ but in fact, $\log V$ is a slightly concave function of $\log N$ (with $d\log V/d\log N = 0$ for the minimal point $N = V$). Anyway, since $\log V$ is an almost linear function of $\log N$, $\log V \approx a \log N + b$, we could try to approximate parameters x, ρ_0 using just linear regression for empirical data $(\log N, \log V)$. In fact, we have

$$\frac{x-1}{x} \approx a, \quad \log N_{\min} \approx a \log N_{\min} + b, \quad \rho_0 x = N_{\min}, \tag{18}$$

which can be easily solved for x, $N_{\min}$, ρ_0 given a, b.

In order to compute λ as the function of N, ρ_0, and x, it is necessary to find the inverse of $e(\lambda) = \lambda^{1/(\lambda-1)}$. The inverse of $e(\lambda)$ is not a closed-form function of its argument but there is some good elementary approximation, which is presented in the appendix.

3 Comparison with empirical data

In order to compare our theoretical model with natural language data, we have collected a selection of texts of various sizes which we downloaded from Project Gutenberg website – `http://www.promo.net/pg/index.html`. The full selection is listed in table 1. All the texts are raw English texts, in which we chose the types w to be the graphical words (word-forms) rather than their (disambiguated) lemmas. We ignored the punctuation and turned all word-forms into lower-case. In this way, all text processing and data plotting could be done automatically in several seconds on a PC using simple scripts in Perl and Gnuplot.

For the given e-text data, we have estimated the parameters of our model for two cases: (1) $\rho_0 = 1/x$ (only x is estimated), (2) ρ_0 is variable (both x and ρ_0 are estimated). The resulting values of parameters and implied characteristic constants $x/(x-1)$, $N_{\min}$, N_0 are given in table 2. The estimation of the parameters was done using least-square fit for plot $(\log N, \log V)$ with nonlinear theoretical curves given by chain of formulas (15), (24), (16). (Slightly better than linear regression (18).) The plot of the data including the fits is given in figure 4.

In figure 4, we can see that the model with variable ρ_0 accords with the data consistently better than $\rho_0 = 1/x$. For this model, the ratio of predicted and actual vocabulary size V is almost always about 1, independently of the text size N. For none of the observed data points, the ratio exceeds the range $[0.5, 2]$ (see figure 7).

The curious feature of the model with variable ρ_0 is that it predicts that there is a minimal text length $N_{\min} \approx 159$. Nevertheless, all e-texts considered as data were well-formed literary texts, so this statement need not be so absurd, as it might appear i.e. for random typing texts.

The parameters of the theoretical model for both variable and fixed ρ_0 were estimated using $(\log N, \log V)$ plot only. When we compare the rank-count distributions implied by the same parameters and the empirical rank-count distributions, we might observe greater departures. In fact, it is so. The model with $\rho_0 = 1/x$ seems to predict better the frequencies $f(r)$ for lower ranks r (figures 5, 6). Figure 8 confirms our rational assumption that $f(1)/N$ should be roughly constant across the texts of any length. Nevertheless, the model with variable ρ_0 still better reflects the variability of the power-law tail of $f(r)$ for the highest ranks r (figure 6 as opposed to 5).

Title	Author	Text size N	Vocab. size V
Peach Blossom Shangri-la	T. Yuan Ming	735	332
A Modest Proposal	J. Swift	3427	1092
On the Brain	T. H. Huxley	4017	1078
The Lake Gun	J. F. Cooper	5328	1488
Song Book of Quong Lee...	T. Burke	5440	1612
The Adventure of the Dying...	A. Conan Doyle	5857	1419
The Adventure of the Red Circle	A. Conan Doyle	7407	1668
Everybody's Business...	D. Defoe	7483	1766
Why Go to College?	A. F. Palmer	7847	1915
Dickory Cronke	D. Defoe	10426	2138
The Princess de Montpensier	Lafayette	10904	1881
Bickerstaff-Partridge Papers	J. Swift	13218	2928
The Categories	Aristotle	14488	1394
The New Atlantis	F. Bacon	15769	2750
The City of the Sun	T. Campanella	16855	3239
Alice in Wonderland	L. Carroll	27870	2868
Through the Looking-Glass	L. Carroll	31055	3059
The Battle of the Books...	J. Swift	38944	6068
Utopia	T. More	43633	4624
Around the World in 80 Days	J. Verne	63290	6853
Erewhon	S. Butler	84717	7800
Five Weeks in a Balloon	J. Verne	93252	8524
Eight Hundred Leagues...	J. Verne	95568	8210
20,000 Leagues Under the Sea	J. Verne	100598	8294
Gulliver's Travels	J. Swift	104650	8191
One of Ours	W. Cather	126621	10049
Life of William Carey	G. Smith	143849	11072
Memoirs	Comtesse du Barry	160790	10278
The Mysterious Island	J. Verne	194213	9743
The Journal to Stella	J. Swift	238787	10642
Critical & Historical Essays	Macaulay	296553	18684
The Descent of Man	C. Darwin	308171	14086
Mark Twain, A Biography	A. B. Paine	515597	22572
First Folio/35 Plays	W. Shakespeare	820016	30820
The Complete Memoirs	J. Casanova	1262287	24093

Table 1: The choice of 35 e-texts from Project Gutenberg.

	$\rho_0 = 1/x$	ρ_0 variable
x	11.477	3.1193
$x/(x-1)$	1.0954	1.4719
ρ_0	0.087132	51.108
$N_{\min}$	1	159.42
N_0	96456.	3607.6

Table 2: The parameters resulted for the e-texts in two estimation schemes.

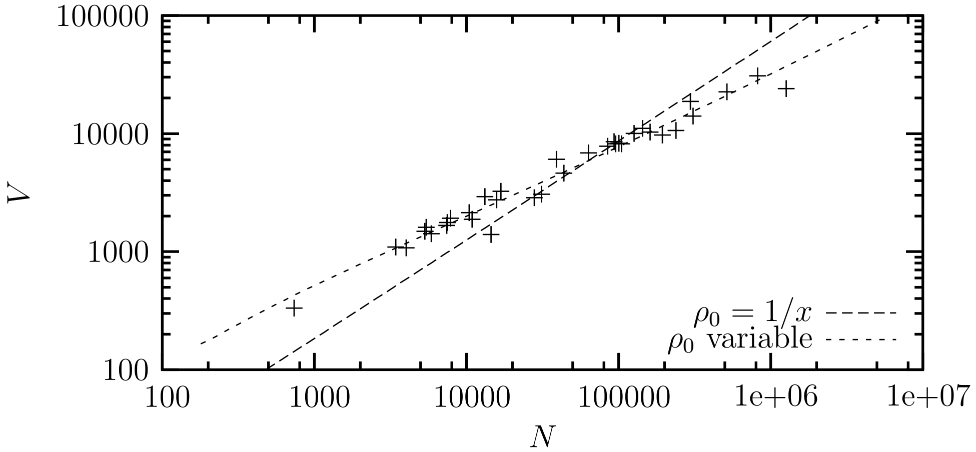

Figure 4: Plots of vocabulary size V as function of text size N for chosen e-texts.

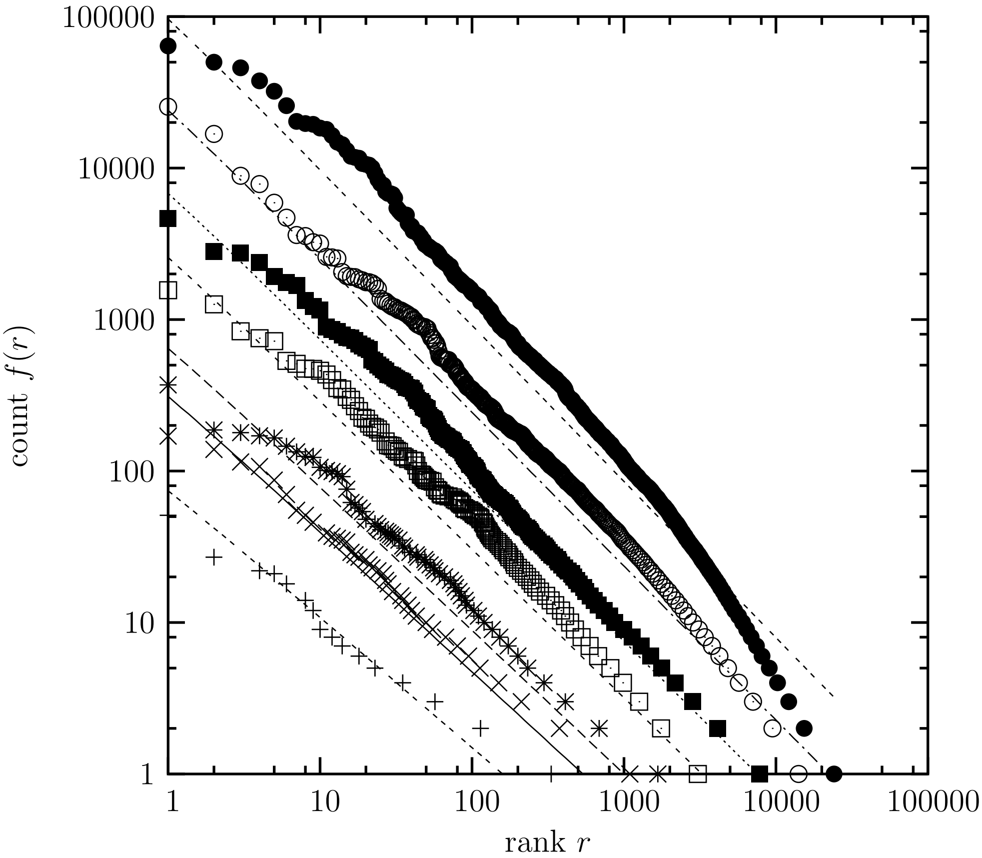

Figure 5: Plots of counts $f(r)$ against ranks r for the following texts: Peach Blossom Shangri-la, $N = 735$; A Modest Proposal, $N = 3427$; The Adventure of the Red Circle, $N = 7407$; Through the Looking-Glass, $N = 31\,055$; Erewhon, $N = 84\,717$; The Descent of Man, $N = 308\,171$; The Complete Memoirs, $N = 1\,262\,287$ (points growing respectively). The smooth curves stand for the count distributions predicted by our model with parameters x, ρ_0 as in the left column of table 2 ($\rho_0 = 1/x$).

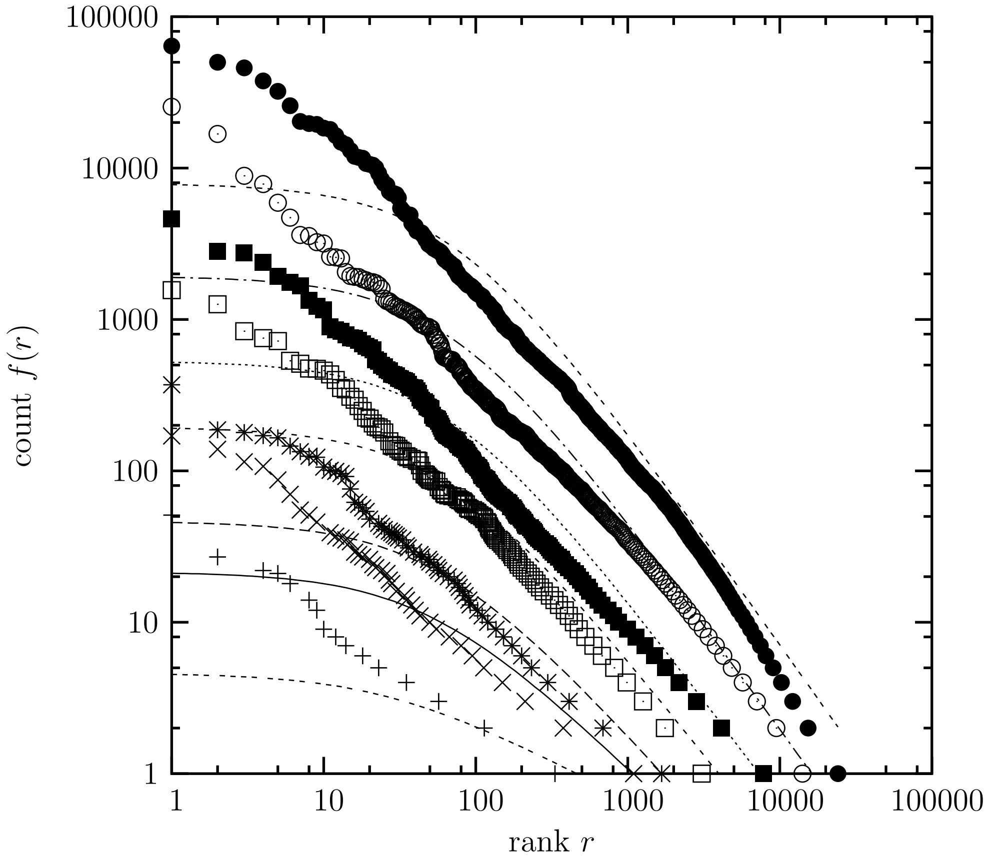

Figure 6: Plots of counts $f(r)$ against ranks r for the following texts: Peach Blossom Shangri-la, $N = 735$; A Modest Proposal, $N = 3427$; The Adventure of the Red Circle, $N = 7407$; Through the Looking-Glass, $N = 31\,055$; Erewhon, $N = 84\,717$; The Descent of Man, $N = 308\,171$; The Complete Memoirs, $N = 1\,262\,287$ (points growing respectively). The smooth curves stand for the count distributions predicted by our model with parameters x, ρ_0 as in the right column of table 2 (variable ρ_0).

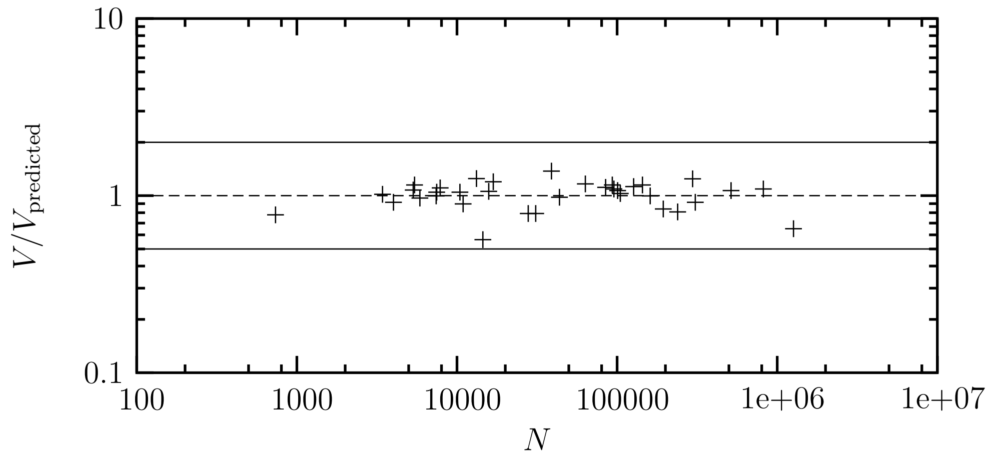

Figure 7: Plots of ratio of empirical vocabulary size V for chosen e-texts to the predicted vocabulary size $V_{\text{predicted}}$ given by our model with parameters x, ρ_0 as in the right column of table 2 (variable ρ_0). The constant lines correspond to $V/V_{\text{predicted}} = 0.5, 1, 2$.

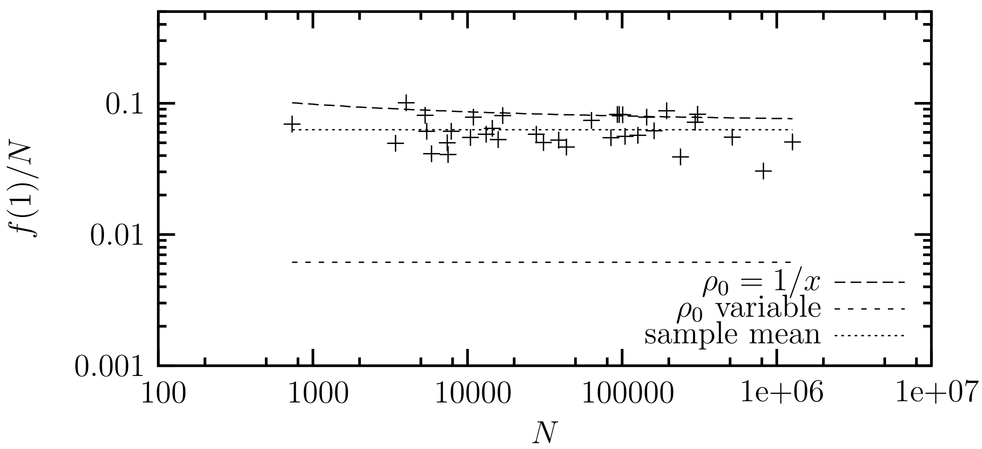

Figure 8: The relative count of the most frequent word against the predicted $f(1)/N$ and the text size N for the whole selection of e-texts.

4 Conclusions

In this article, we have presented a very simple improvement of classical Mandelbrot-Zipf's law for natural language texts. The improvement was done not by introducing new parameters, but by letting the present ones vary with respect to the text size. The newly introduced constraint for the variability of parameters was that the relative counts of the most frequent words be constants independent of the text size.

The resulting model does not fit the data so well as much more complex LNRE models (Baayen, 2001; Orlov, 1982), but it still reproduces Mandelbrot's exponent $1+\epsilon < 1$ for text length $N < N_0$ and $1+\epsilon > 1$ for $N > N_0$. The model also accords with empirical data qualitatively in several other aspects. Here, we have discussed theoretically the probably most complex phenomena in rank-count distribution which are still explainable by simple Mandelbrot-Zipf's formula (2).

Still, we have not checked if the quantitative departures of our model can be decreased if approximations (5), (6), assuming falsely $\rho \gg 1$, were replaced by exact summations and equalities. In this case, we lose pretty closed-form formulas but maybe we could obtain a better fit for self-consistent expressions.

A Approximating the inverse of $e(\lambda) = \lambda^{1/(\lambda-1)}$

The function defined by (13) is related to the definition of base e of natural logarithm. Actually,

$$e(1) = e. \tag{19}$$

Function $e(\lambda)$ is an easily computable function of λ. Unfortunately the inverse is not true. Quantity λ is not a simply computable function of $e(\lambda)$. There is, however, an easily invertible and good approximation $\bar{e}(\lambda)$,

$$\bar{e}(\lambda) = 1 + \frac{e-2}{\sqrt{\lambda}} + \frac{1}{\lambda}. \tag{20}$$

One can define the relative error of $\bar{e}(\lambda)$ as

$$\bar{b}(\lambda) = \frac{\bar{e}(\lambda) - e(\lambda)}{\bar{e}(\lambda)}. \tag{21}$$

Function $e(\lambda)$ has domain $\lambda \in \{0, \infty\}$. In this domain, the following substitution is convenient

$$\lambda = \frac{1-u}{1+u}, \tag{22}$$

where $u \in \{-1, 1\}$. Let $b(u) = \bar{b}(\lambda)$. Then $b(u) = 0$ for $u = -1, 0, 1$. ($b(u)$ for $u = -1, 1$ is defined by the corresponding limits.) Explicitly

$$b(u) = 1 - \frac{\left[\frac{1-u}{1+u}\right]^{-1/2u}}{\sqrt{\frac{1-u}{1+u}} + (e-2) + \sqrt{\frac{1+u}{1-u}}}, \tag{23}$$

so $b(u) = b(-u)$. Furthermore, looking at the plot of $b(u)$ (Figure 9) one can see that $0 \leq b(u) < 0.04$. Resuming, $\bar{e}(\lambda)$ is a good simple approximation of $\lambda^{1/(\lambda-1)}$.

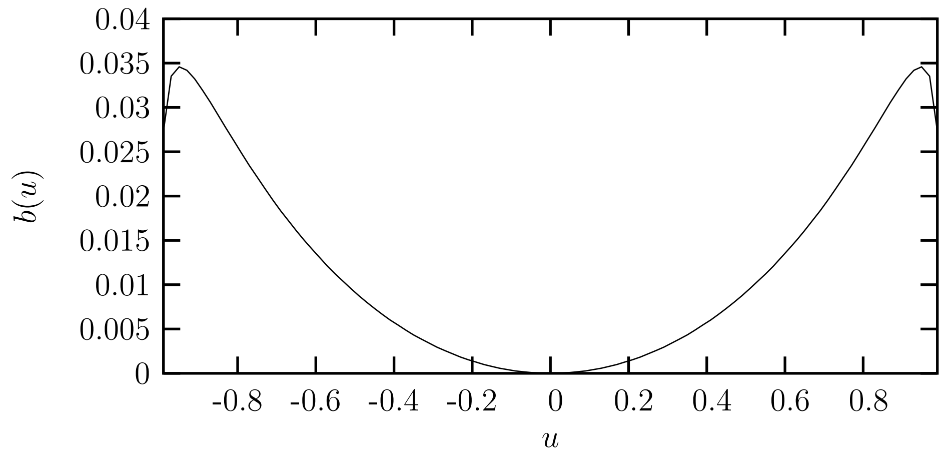

Figure 9: The plot of $b(u)$.

In order to find λ for given $\bar{e}(\lambda)$, one sees that definition (20) is a quadratic equation for $1/\sqrt{\lambda}$ and it can be immediately solved,

$$\lambda = \frac{4}{\left[2 - e + \sqrt{e^2 - 4e + 4\bar{e}(\lambda)}\right]^2}. \tag{24}$$

In formula (24), the one of two solutions was chosen which reproduces $\lambda = 1$ for $\bar{e}(\lambda) = e$.

References

Baayen, H. (2001). *Word frequency distributions.* Kluwer Academic Publishers.

Belevitch, V. (1956). Théorie de l'information et statistique linguistique. *Académie royale de Belgique, Bulletin de la classe des sciences*, 419–436.

Camacho, J., & Solé, R. V. (1999). *Scaling and Zipf's law in ecological size spectra.* (Santa Fe Institute Working Paper 99-12-076)

Dębowski, Ł. (2002). *On the best theories for learningful texts.* (http://www.ipipan.waw.pl/~ldebowsk)

Denisov, S. (1997). Fractal binary sequences: Tsallis thermodynamics and the Zipf's law. *Physics Letters A*, *235*, 447–451.

Estoup, J. B. (1916). *Gammes sténographiques.* Paris: Institut Stenographique de France.

Günther, R., Levitin, L., Schapiro, B., & Wagner, P. (1996). Zipf's law and the effect of ranking on probability distributions. *International Journal of Theoretical Physics*, *15*, 395–417.

Harremoës, P., & Topsøe, F. (preprint). *Zipf's law, hyperbolic distributions and entropy loss.* (`http://www.math.ku.dk/~topsoe/manuscripts.html`)

Khmaladze, E. (1987). *The statistical analysis of large number of rare events.* (Technical Report MS-R8804, Dept. of Mathematical Statistics, CWI. Amsterdam: Center for Mathematics and Computer Science)

Kornai, A. (1999). Zipf's law outside the middle range. In *Proceedings of sixth meeting on mathematics of language* (pp. 347–356). University of Central Florida.

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, *38*, 1842–1845.

Mandelbrot, B. (1954). Structure formelle des textes et communication. *Word*, *10*, 1–27.

Mandelbrot, B. B. (1983). *The fractal geometry of nature.* New York: W. H. Freeman.

Montemurro, M. A. (2001). *Beyond the Zipf-Mandelbrot law in quantitative linguistics.* (`http://xxx.lanl.gov/abs/cond-mat/0104066`)

Orlov, J. K. (1982). Linguostatistik: Aufstellung von Sprachnormen oder Analyse des Redeprozesses? In J. K. Orlov, M. G. Boroda, & I. Š. Nadarejšvili (Eds.), *Sprache, Text, Kunst. Quantitive Analysen* (pp. 1–55). Studienverlag Dr. N. Brockenmeyer, Bochum.

Pareto, V. (1897). *Cours d'économie politique.* Lausanne, Paris: Rouge.

Sambor, J. (1988). Lingwistyka kwantytatywna — stan badań i perspektywy rozwoju. *Biuletyn Polskiego Towarzystwa Językoznawczego*, *XLI*, 47–67.

Tsallis, C. (2000). *Entropic nonextensivity: A possible measure of complexity.* (Santa Fe Institute Working Paper 00-02-043)

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology.* Houghton Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley.