scientific reports

Check for updates

OPEN

Lognormals, power laws and double power laws in the distribution of frequencies of harmonic codewords from classical music

Marc Serra-Peralta[1,2], Joan Serrà[3] & Álvaro Corral[1,4,5]✉

Zipf's law is a paradigm describing the importance of different elements in communication systems, especially in linguistics. Despite the complexity of the hierarchical structure of language, music has in some sense an even more complex structure, due to its multidimensional character (melody, harmony, rhythm, timbre, etc.). Thus, the relevance of Zipf's law in music is still an open question. Using discrete codewords representing harmonic content obtained from a large-scale analysis of classical composers, we show that a nearly universal Zipf-like law holds at a qualitative level. However, in an in-depth quantitative analysis, where we introduce the double power-law distribution as a new player in the classical debate between the superiority of Zipf's (power) law and that of the lognormal distribution, we conclude not only that universality does not hold, but also that there is not a unique probability distribution that best describes the usage of the different codewords by each composer.

For centuries, physics has dealt with deterministic mathematical laws, such as Newton's laws of mechanics, the laws of electromagnetism, or the laws of relativity[1]. It were Maxwell and Boltzmann who, in the nineteenth century, discovered probabilistic or statistical laws in the study of the motion of the (hypothetical) particles constituting a gas. The so-famous Planck's radiation law can be understood as another instance of a probabilistic law. The great insight of Maxwell, Boltzmann, and Planck (and others, like Einstein) was the introduction of probability to infer the mechanics of the constituents of matter and radiation. That insight has been one of the most successful knowledge programs in the history of humankind. Although the just-mentioned examples of probabilistic laws are valid in the "ideal case" (no interaction between the constituents), an interaction (at least with the surroundings) has to be present to ensure the existence of a state of equilibrium. In general, the study of how macroscopic behavior emerges from microscopic interactions (either in equilibrium or out of equilibrium) is the goal of statistical physics.

In ecology, in the social sciences (sociology, economics, demography), and in the study of technological and information networks, one is, in some sense, in a situation similar to statistical physics, in which there is an enormous number of individual entities whose behavior depends on each other, leading to an emerging collective behavior[2,3]. Despite the lack of well-defined underlying microscopic laws, it is remarkable that one may find regular statistical laws describing the aggregated properties of the constituent entities, and even more remarkable that it is the same law which seems to capture a particular but important aspect of many of these systems[4–7]. This ubiquitous and nearly universal law is Zipf's law[8], which describes how the constituent entities, or tokens, are distributed into larger groups, or types. In this way, Zipf's law states that the size distribution of these groups (measured in terms of constituent entities) follows a power-law distribution with a loosely constrained value of the exponent, close to 2 (for the probability mass function). In an additional twist in favor of the meaningfulness of statistical "natural" laws beyond physics, Zipf's law emerges again when one considers quantitative

[1]Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra, 08193 Barcelona, Spain. [2]Departament de Fsica Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain. [3]Dolby Laboratories, 08018 Barcelona, Spain. [4]Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain. [5]Complexity Science Hub Vienna, Josefstädter Straßbe 39, Vienna 1080, Austria. ✉email: alvaro.corral@uab.es

approaches to human sciences (mainly the study of language[9–13]), where the nature of the interactions between the constituent entities is not so clear[14,15].

However, there have been serious issues with the Zipf's paradigm, and with power laws in general, being the most important of them the lack of generality of the results[13,16] and low statistical rigor[17–22]. In the first case, a very small number of datasets are usually analyzed in order to establish the validity of Zipf's law in every particular system. For instance, in quantitative linguistics, research articles are usually focused in about a dozen (or even less) texts[23,24], with the selection of them seeming rather arbitrary. Therefore, many published claims should be considered as anecdotic examples, or conjectures, rather than well-established facts.

Regarding statistical rigor, proper fitting methods and goodness-of-fit tests have seldomly been used, being replaced many times by visual, qualitative checks. Moreover, some apparently rigorous procedures[19] have been found to yield inconsistent results[25,26]. Therefore, it is not yet established for which systems Zipf's law is a rough or even bad approximation and for which systems it is a good description in some range or limit. An annoying side effect of the lack of proper statistical tools is the recurrent debate about if the lognormal distribution is superior or not to the power law to describe (some) Zipfian systems[27,28]. A further concern is some ambiguity in the definition of Zipf's law[13,29,30], which admits several mathematical formulations not strictly equivalent between them.

In recent years, several authors have tried to overcome the problems of Zipf's law in linguistics. For instance, Moreno-Sánchez et al.[13] analyzed different mathematical alternatives to Zipf's law, using all English texts (tens of thousands) available from the Project Gutenberg digital library. In a sort of complementary study, Mehri and Jamaati[31] considered just one text (the Bible) but in its translation to one hundred distinct languages. An alternative approach, instead of analyzing many different individual texts, has been to use big corpora (collections such as the British National Corpus formed by gathering many text fragments). Although this is also a valid procedure, one cannot assert that results (for instance, a claimed double power-law distribution comprising Zipf's law for large word frequencies[32–34]) are not an artifact arising from the mixture of rather different sort of texts[35]. In other words, the statistical properties of the British National Corpus could be different if the corpus were compiled in a different way (e.g., changing the length of the selected fragments), and very different also to the ones of a hypothetical text of the same length from a single author.

In the last years, diverse forms of artistic expressions have been approached through the eyes of complex-systems science[36]. In this paper, we deal with music. Music seems to be a necessary and sufficient condition for "humanity"[37], in the sense that music has been present across all human societies in all times, and other animal species do not seem to have real musical capabilities. Thus, music is a uniquely human attribute (needless to say, if any extraterrestrial intelligence were ever discovered, one of the first questions to figure out would be about its relationship with some sort of music[38,39]). Even more, music is one of the human activities that attracts more public interest (e.g., at the time of writing this, out of the 10 Twitter accounts with more followers, 6 correspond to popular musicians or musical performers[40]). Certain parallelisms between natural language and music have been noted in the literature, where music has been sometimes categorized as a "language"[41]; nevertheless, there is no clear notion of grammar and semantic content in music[37,41] although there exist relations in terms of rhythm, pitch, syntax and meaning[42].

In any case, one can conceptualize music as a succession (in time) of some musical descriptors or symbols, which can be counted in a Zipf-like manner[43], with the frequency of appearance of each different symbol playing the role of the size of the groups (types) in which Zipfian systems are organized. A remarkable problem is that, in contrast to language[24], the individual entities to analyze in music can be extraordinarily elusive to establish. For instance, Manaris et al.[44] mention several possibilities involving different combinations of pitch and duration, as well as pitch differences, illustrating the multidimensional character of music. This, together with some technicalities to deal with musical datasets, may explain the fact that the study of Zipf's law and other linguistic laws in music has been substantially limited in comparison to natural language. Nevertheless, some precedents are of interest. Using a simple pitch-duration pair as a metric, Zanette[41] fitted a variant of Zipf's law to four classical musical pieces, finding a rather high power-law exponent (up to 4.6 for the probability mass function of frequency, except for a piece by Arnold Schönberg). Later, Liu et al.[45] did the statistics of pitch jumps for five classical composers to find even higher power-law exponents. In a more large-scale study, Beltrán del Río et al.[46] analyzed plain pitches (of which there is a maximum "vocabulary" of 128) in more than 1800 MIDI files, containing classical music, jazz, and rock, to fit a generalization of Zipf's law. This, in general, resulted in a rather high power-law exponent. In any case, these publications did not use very high statistical standards.

In a more recent attempt, studying popular Western music, Serrà et al.[47] considered the combination of pitch classes present in short time intervals (harmonic and melodic chords, in some sense) to construct discretized chromas. Aggregating individual pieces for fixed lustrums of the twentieth century, and using maximum-likelihood estimation with the necessary goodness-of-fit tests, they found a robust exponent for the tail close to 2.2 (in agreement with Zipf's law), which remained stable across the different historical periods that were analyzed. Distributions of timbral indicators were also explored in that work, as well as by Haro et al.[48]; intriguingly, the latter reference found that Zipf's law for timbre is not only fulfilled by music but also by speech and natural sounds (such as rain, wind, and fire).

In this paper, we analyze classical music using its "crystallization" into electronic MIDI scores, by means of a rather large database. One could argue that, in order to access genuine expressions of music, audio recordings are preferable to MIDI scores, due to the fact that the latter may lack the richness and nuances of interpretation[49] (although there are MIDI files created from the life performance of a musical piece). Nevertheless, for our purposes, scores contain the essence of music, and, in the case of music previous to the twentieth century, they are our best remainder of the original intention of the composer. Moreover, as to undertake statistical analysis we need to deal with discretized elements, scores provide an objective first step in such discretization.

In the next sections, we present the characteristics of the corpus used ("Data, processing, and elementary statistics"), describe the extraction of harmonic codewords from the MIDI files ("Data, processing, and elementary

G. Dufay:	**ln**, dpl;	J. Desprez:	**ln**, dpl;	C. de Morales:	**dpl**, ln;	G. P. da Palestrina:	**ln**, dpl
O. Lassus:	**ln**, dpl;	T. L. de Victoria:	**dpl**, ln;	W. Byrd:	**dpl**, ln;	C. Gesualdo:	**dpl**, ln
J. Dowland:	**dpl**, ln;	C. Monteverdi:	**ln**, pl;	G. Frescobaldi:	**dpl**, ln;	S. Scheidt:	**ln**, pl
J. J. Froberger:	**dpl**, ln;	J. B. Lully:	**dpl**, ln;	J.-H. d'Anglebert:	**ln**, dpl;	D. Buxtehude:	**dpl**, ln
J. Pachelbel:	**ln**;	F. Couperin:	**dpl**;	D. Zipoli:	**ln**, dpl;	A. Vivaldi:	**dpl**, ln
J.-F. Dandrieu:	**dpl**;	T. Albinoni:	**dpl**, ln;	J. S. Bach:	–;	D. Scarlatti:	**dpl**, ln
G. F. Händel:	**dpl**;	J.-P. Rameau:	**dpl**, ln;	G. P. Telemann:	**dpl**, ln;	J. Haydn:	**dpl**
J. G. Albrechtsberger:	**ln**, pl;	W. A. Mozart:	**dpl**;	M. Clementi:	**dpl**, ln;	L. van Beethoven:	**dpl**
N. Paganini:	**pl**, ln;	F. Schubert:	**ln**;	J. B. Cramer:	**ln**, pl;	F. Mendelssohn:	**ln**, dpl
F. Chopin:	**dpl**, ln;	R. Schumann:	**dpl**, ln;	H. Berlioz:	**ln**, pl;	F. Liszt:	**ln**, dpl
L. M. Gottschalk:	**ln**, pl;	C.-V. Alkan:	**dpl**;	C. Franck:	**ln**;	G. Bizet:	**ln**, pl
A. Bruckner:	**ln**;	M. Mússorgsky:	**dpl**, ln;	J. Brahms:	**ln**;	P. I. Tchaikovsky:	**ln**, dpl
A. Dvorák:	**ln**, dpl;	A. Guilmant:	**pl**, ln;	E. Grieg:	**ln**, dpl;	C. Saint-Saëns:	**dpl**, ln
I. Albéniz:	**ln**, dpl;	G. U. Fauré:	**ln**;	G. Mahler:	**dpl**, ln;	C. Debussy:	**ln**, dpl
L. Janácek:	**ln**, pl;	S. Joplin:	**pl**, ln;	A. Scriabin:	**ln**, dpl;	M. Reger:	**dpl**, ln
F. Busoni:	**dpl**, ln;	E. Satie:	**ln**, pl;	L. Godowsky:	**dpl**, ln;	S. Karg-Elert:	**pl**, ln
M. Ravel:	**dpl**, ln;	O. Respighi:	**ln**, pl;	S. Rajmáninov:	**ln**, pl;	A. Schoenberg:	**pl**, ln
B. Bartók:	**pl**, ln;	N. Médtner:	**ln**, pl;	G. Gershwin:	**ln**, pl;	S. Prokófiev:	**ln**, pl
Í. Stravinsky:	**pl**, ln;	P. Hindemith:	**ln**, dpl;	D. Shostakóvich:	**dpl**;	O. Messiaen:	**dpl**, ln

Table 1. Name of the 76 classical composers in the *Kunstderfuge* corpus analyzed in this paper. The order is chronological (from left to right and from top to bottom, established by the average between birth and death). The results of the best fit (in bold), followed by other good fits are included. pl = (simple) power law, dpl = double power law, ln = lognormal. Number of composers fitted by a unique distribution: 14 (**pl** 0; **dpl** 8; **ln** 6). Number of composers with two good fits: 61 (**pl**, ln 7; **dpl**, ln 25; **ln**, pl 14; **ln**, dpl 15).

statistics" and Supplementary Information, SI), introduce the probability distributions to fit the codewords counts (power law, double power law, and lognormal, "Power-law, double power-law and lognormal fits"), explain the statistical procedure ("Power-law, double power-law and lognormal fits" and SI), and present the results ("Results"). Naturally, we end with some conclusions. The code used in this paper is available in Github[50].

Data, processing, and elementary statistics

Data. As a corpus of classical music scores we use the *Kunstderfuge* database[51], including 17,419 MIDI scores of composers from the twelfth to the twentieth century. Different aspects of this musical corpus have been analyzed elsewhere[52,53]. We perform a preliminary cleaning in which we identify and remove traditional songs, anthems, anonymous pieces, and also MIDIs arising from live performances, leading to a remainder of 10,523 files. In general, these files contain the name of the composer and an indication of the name of the piece. Further removing files that we are not able to process, files for which we cannot obtain the bar (and cannot determine therefore the temporal unit), files corresponding to very short pieces, and files corresponding to repeated pieces, we retain 9327 of them corresponding to 76 composers, ranging from Guillaume Dufay (1397–1474) to Olivier Messiaen (1908–1992). The complete list of composers, in chronological order, is provided in Table 1. Details of the data cleaning and the method of detection of repeated pieces in the corpus are provided in the Supplementary Information.

Harmonic codewords. Our analysis focuses on the harmonic content of music, understood as the combination of pitches across all instruments in short time frames. A complete summary for the obtention of the elementary units in which we decompose music is provided in the Supplementary Information (see also Ref.[53]); Fig. 1 provides a simple illustration. The different steps are:

(i) Conversion of each MIDI file (corresponding to a piece) into a text file (containing the time occurrence, duration, and pitch of each note).
(ii) Transformation of pitches into (twelve) pitch classes (i.e., collapse into a unique octave).
(iii) Segmentation into elementary time intervals (given by the score beats).
(iv) Construction of chromas: 12-dimensional vectors $(C, C\#, \ldots G\#, A, A\#, B)$, counting the contribution of notes from each pitch class for each elementary time interval and for each stave. That is, we collapse all staves in the piece into a single chroma sequence.
(v) Discretization of chromas (using a discretization threshold), which yields the harmonic codewords to analyze. These are 12-dimensional vectors of binary elements (0, 1).
(vi) Transposition to C major (major pieces) or A minor (minor pieces).
(vii) For each composer, aggregation of all the time series of transposed discretized chromas (corresponding to each piece) into a unique dataset.

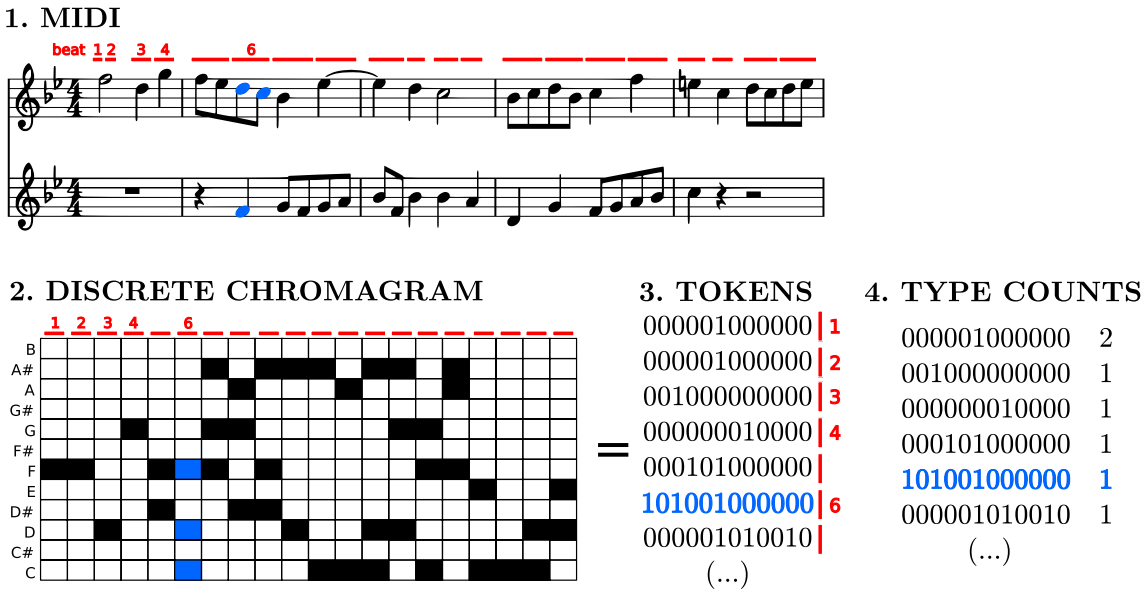

Figure 1. Scheme showing the correspondence between a score (represented by two parallel staves) and its representation in terms of discretized chromas. Counts for each type (number of tokens) are also shown.

This aggregation of pieces of the same composer is done in order to get significant statistics. Although under the framework of some models explaining Zipf's law[54], aggregation makes little sense, for other models there are no such restrictions[55]. In any case our approach is model free. Notice that aggregation can be done in several ways. Ours is equivalent to aggregate the frequencies n of each type, but one could also aggregate the counts $f(n)$ of each frequency n, and this could be done also for relative frequencies. The latter two options would change the results, as they lead to mixtures of the distributions corresponding to each piece.

The obtained codewords contain information about the melody and, mostly, the harmony of the pieces. The most common codewords in the studied corpus are listed in Table 6 of Ref.[53] and they are consistent with characteristic harmonic features, i.e., they correspond to harmonic sets of pitches. Therefore, the results arising from the analysis of the codewords can be associated with the harmonic characteristics of the compositions, especially in the range of high type frequencies.

Elementary statistics. For each dataset (corresponding to a composer), we count the repetitions or absolute frequency n of each type (discretized chromas, see Fig. 1). This absolute frequency is our random variable, and the number of appearances of each value of the frequency (frequencies of frequencies, then) constitutes an empirical estimation of the probability mass function of the frequency, which we may denote as $f(n)$. However, due to the broad range of the distributions (with n ranging from one to hundreds of thousands), it is more convenient to treat n as a continuous random variable and estimate its empirical probability density using logarithmic binning (for visualization purposes only)[20]; so $f(n)$ denotes in fact a probability density, as well as its empirical estimation.

The number of different types present in a dataset (the types with $n \geq 1$) is what we call the vocabulary of the dataset, denoted by V (this is bounded by $2^{12} = 4096$). The sum of all the frequencies of all types yields the total number of tokens, which corresponds, by construction, to the dataset length L measured in terms of the elementary time unit (number of beats, by default). In a formula, $\sum_{i=1}^{V} n_i = L$, where i labels the types.

Power-law, double power-law and lognormal fits

Probability densities and rescaling. As a summary of the empirical probability densities of type frequency, Fig. 2a shows all of them (77 in total, one for each composer plus the global one in which all composers are aggregated). All distributions present many types that only occur once ($n = 1$, the so-called hapax legomena in linguistics[9]), as well as types with very high frequencies ($n > 10^5$ in the global dataset), with a rather smooth decaying curve linking both extremes.

When rescaled, a roughly similar shape is unveiled, as seen in Fig. 2b. Rescaling is done in the following way: $n \to n\langle n\rangle/\langle n^2\rangle$ and $f(n) \to f(n)\langle n^2\rangle^2/\langle n\rangle^3$, with $\langle n\rangle$ and $\langle n^2\rangle$ denoting the first (mean) and second empirical moments of the distribution. The reason behind such rescaling is the assumption of a hypothetical scaling form for $f(n)$,

$$f_{\mathrm{sca}}(n) = \frac{1}{a}\left(\frac{a}{\theta}\right)^{\beta_1} G\left(\frac{n}{\theta}\right),$$

defined for $n \geq a$, with a a (fixed) lower cutoff, θ a scale parameter, and G a scaling function behaving as a decreasing power law with exponent β_1 for small arguments ($1 < \beta_1 < 2$) and decaying fast enough for large arguments (in order that $\langle n\rangle$ and $\langle n^2\rangle$ exist). Then, $\langle n\rangle \propto a(\theta/a)^{2-\beta_1}$ and $\langle n^2\rangle \propto a^2(\theta/a)^{3-\beta_1}$, and isolating, $\theta \propto \langle n^2\rangle/\langle n\rangle$

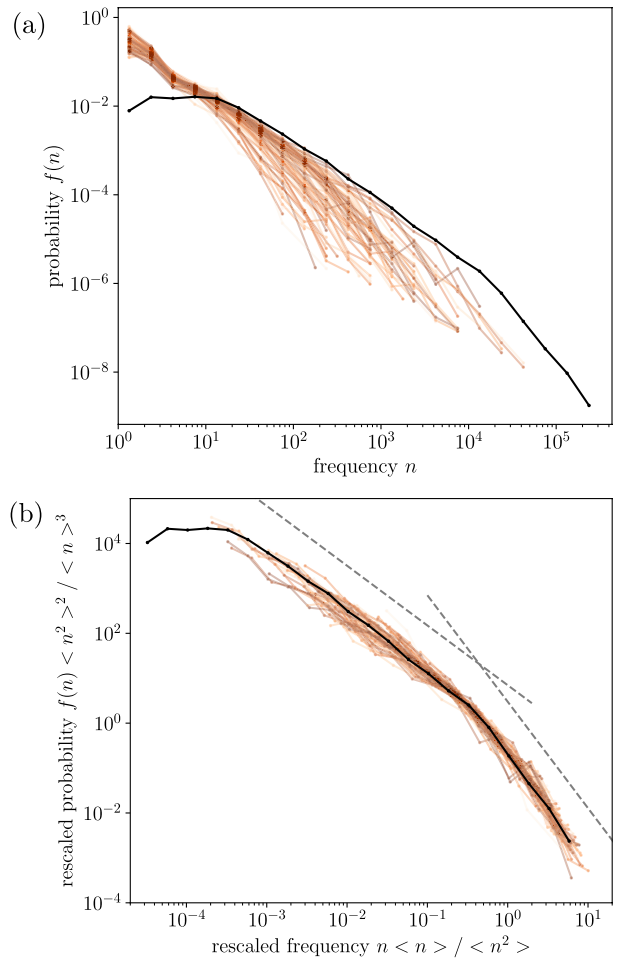

Figure 2. (**a**) Empirical probability densities of codeword frequency for the 76 composers, individually (orange-brown lines), and with all of them aggregated (black line). (**b**) Same distributions under rescaling of the axes, where a roughly similar shape emerges for every composer and the global corpus. Dashed straight lines are power laws with exponents 1.33 and 2.4, derived from the fit of the global dataset.

and $\theta^{\beta_1} \propto a^{\beta_1 - 1}\langle n^2\rangle^2/\langle n\rangle^3$, which justifies the rescaling and allows its verification without knowledge of the values of θ and β_1, see Ref.[56]. When a literary piece is broken into different parts, this rescaling is equivalent to $n \to n/L$ and $f(n) \to f(n)LV$ (see Ref.[57]); however, this is not the case here, and the latter scaling form (in contrast to the former one) does not work well (not shown). Visual inspection of the rescaled plot based on the ratio of moments (Fig. 2b) suggests that the lognormal and the double power law seem appropriate candidate distributions to fit the data.

In this work, we consider three different fitting distributions, all of them continuous; thus, the frequency n is assumed to be a continuous random variable. The explanation of the three fitting distributions follows, with special emphasis in the double power law.

Simple power-law distribution. The simple power-law distribution, referred to as untruncated power law or simply as power-law (pl) distribution, has a probability density

$$f_{\mathrm{pl}}(n) = \frac{\beta - 1}{a}\left(\frac{a}{n}\right)^{\beta} \text{ for } n \geq a,$$

and zero otherwise. The exponent β fulfils $\beta > 1$ and a is the lower cut-off, fulfilling $a > 0$. Considering a as fixed, there is only one free parameter, which is β. In the particular case in which β is in the range $1.8 \leq \beta \leq 2.2$ we will talk about the fulfillment of Zipf's law.

Double power-law distribution. The second fitting distribution is the one we call the double power-law (dpl) distribution[22], whose probability density is given by

$$f_{\rm dpl}(n) = (1-q)\frac{\beta_1 - 1}{\theta}\frac{1}{c^{1-\beta_1}-1}\left(\frac{\theta}{n}\right)^{\beta_1} \text{ for } a \le n \le \theta,$$

$$f_{\rm dpl}(n) = q\frac{\beta_2 - 1}{\theta}\left(\frac{\theta}{n}\right)^{\beta_2} \text{ for } n \ge \theta,$$

and zero for $n < a$. Naturally, the lower cut-off a may take a different value than for the simple power law, although at this point we use the same symbol for simplicity. The two exponents β_1 and β_2 fulfill $-\infty < \beta_1 < \infty$ with $\beta_1 \neq 1$ and $\beta_2 > 1$; θ is a scale parameter fulfilling $\theta \ge a$; and the lower cut-off a fulfills $a \ge 0$ if $\beta_1 < 1$ and $a > 0$ if $\beta_1 > 1$. The auxiliary parameter c is defined as $c = a/\theta$ and the parameter q is not free either but ensures continuity at $n = \theta$ between the two regimes, leading to

$$q = \frac{\beta_1 - 1}{(\beta_2 - 1)c^{1-\beta_1} - (\beta_2 - \beta_1)},$$

which gives $0 < q \le 1$. As the expressions that multiply $1 - q$ and q in $f_{\rm dpl}(n)$ are normalized in their respective ranges, q turns out to be the fraction of probability contained in the range $n \ge \theta$. If a is fixed, the free parameters are β_1, β_2, and θ. The usual (untruncated) power-law distribution is recovered either in the limits $\theta = a$ (equivalent to $q = 1$) or $\beta_1 = \beta_2 > 1$.

Note that the double power-law distribution has two contributions: on the left ($n \le \theta$) we have a truncated (from above) power-law distribution, with weight $1 - q$; on the right ($n \ge \theta$) we have an untruncated power law, with weight q. We will take advantage of this fact to fit the double power law to the empirical data, fitting, separately, a truncated power law in the range $a \le n \le b$ (where we have redefined θ as b), and fitting an untruncated power law in $n \ge a_2$, (redefining θ as a_2)[58]. In each fit, a_2 and b are fixed and considered different, in general.

The method used for the fit is, in both cases, the one explained in Refs.[20,22] (see the Supplementary Information). If both fits are accepted in some range (in the sense that they cannot be rejected, with p—value greater than 0.20) and their ranges overlap (in the sense that the upper cut-off b of the truncated power law is above the lower cut-off a_2 of the untruncated power law, and both power laws cross each other), the double power-law fit is not rejected (provided $a \le 32$, see below). The resulting value of θ is given by the value of n at which both fits cross, which turns out to be

$$\theta = \left[\frac{q}{1-q}\,\frac{\beta_2 - 1}{\beta_1 - 1}\left(\left(\frac{a}{b}\right)^{1-\beta_1} - 1\right)\frac{a_2^{\beta_2 - 1}}{b^{\beta_1 - 1}}\right]^{\frac{1}{\beta_2 - \beta_1}}.$$

In the case in which the double power-law fit works well, we expect $a_2 \simeq \theta \simeq b$ (but with $a_2 \le \theta \le b$). The replacements $b \to \theta$ and $a_2 \to \theta$ can lead to small changes in the fitted values of β_1 and β_2; nevertheless, the good visual performance of the fits allows us to disregard such changes.

Although we could have fitted the power laws in the discrete case[13,30], we have considered the continuous case instead, in order to compare on equal footing with the (truncated) lognormal distribution defined below, which is continuous. For high enough values of n, the distinction between continuous and discrete random variables becomes irrelevant, but not for small values of n.

The sudden change of exponent of the double power law at $n = \theta$ may seem "unphysical", but the distribution works quite well for the number of data we are dealing with (in the last section we discuss an extension of the double power law that avoids this "unphysicality"). The case of interest for us is when $0 < \beta_1 < \beta_2$; specifically, when β_2 is between 1.8 and 2.2, the resulting power-law tail is in correspondence with Zipf's law; then, we will refer to this particular case of a double power-law distribution as "double Zipf"[32].

Truncated lognormal distribution. The third fitting distribution that we deal with is the (lower) truncated lognormal (ln), whose probability density is

$$f_{\rm ln}(n) = \sqrt{\frac{2}{\pi}}\left[\operatorname{erfc}\left(\frac{\ln a - \mu}{\sqrt{2}\sigma}\right)\right]^{-1}\frac{1}{\sigma n}\exp\left(-\frac{(\ln n - \mu)^2}{2\sigma^2}\right) \text{ for } n \ge a$$

and zero otherwise, with $a \ge 0$ and μ and σ the two free parameters (being the mean and the standard deviation of the associated untruncated normal distribution); $\operatorname{erfc}(y) = \frac{2}{\sqrt{\pi}}\int_y^\infty e^{-x^2}dx$ is the complementary error function. The fitting procedure[22] proceeds in exactly the same way as for the untruncated power law (with the only difference that the lognormal involves two free parameters, μ and σ, when a is considered fixed).

Fitting method and model selection. The random variable to fit is the absolute frequency n of the codewords (types); this choice is not obvious in Zipfian systems, see Ref.[30]. The fitting method is the one in Refs.[20,22], consisting in maximum-likelihood estimation and the Kolmogorov-Smirnov goodness-of-fit test for different values of the lower cut-off a (and the upper cutoff b for the first regime of the double power law). The selected value of a (and b) is the one that yields the largest number of types in the fitting range (i.e., the one comprising more realizations of the random variable) among all the fits that lead to a p—value larger than 0.20. If the resulting value of a is larger than $10^{3/2} \simeq 32$ (corresponding to more than 1.5 orders of magnitude from $n = 1$ to a), the fit is rejected, otherwise it is "accepted" and considered a "good fit".

For the comparison of the fits provided by the different distributions, we have to deal with different subsets of the data (as the values of the lower cut-off a will be different in each case, in general). We take the simple

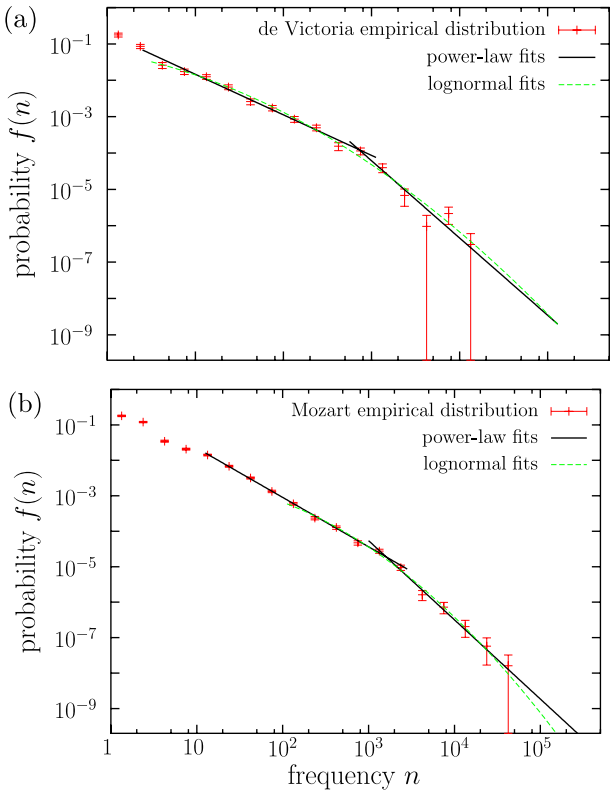

Figure 3. Empirical probability density of codeword frequency for (**a**) de Victoria and (**b**) Mozart, together with double power-law and lognormal fits. Among all the composers, de Victoria and Mozart yield the largest logarithmic span of the fitting range, n_{max}/a, for the double power law. Note that de Victoria also yields the lognormal fit with the largest n_{max}/a (comparable to the value for the double power law).

criterion of selecting the fitting distribution that yields the smaller value of a, i.e., the model that explains the larger portion of the data. The result is what we refer to as the "best fit". This is explained in detail in the Supplementary Information.

Results

Simple power-law and double power-law fits. We start by comparing the results of fitting a simple power law and a double power law. We find that for the majority of the composers (48 of them, 63%) the double power law provides a good fit, which is obviously preferable to the simple power law (due to the fact that the double power law covers a larger range of data than the simple power law, as the latter is a part of the former). In fact, when the double power law fits the data, the simple power law is rejected, as this only fits the tail, which is a too small fraction of the data (with $a_{\rm pl} \gg 32$). A couple of particularly good double power-law fits are shown as an illustration in Fig. 3, where empirical probability densities and their fits are plotted together; there, it is also clear how the simple power law fits a rather small part of the data ($n > a_{\rm pl} \simeq 10^3$).

Figure 4 shows, for each composer for which the double power-law fit is not rejected, the corresponding fitting ranges and exponents. We see that when the cut-offs are expressed in terms of the relative frequency, these are much more stable between different composers than when expressed in absolute frequencies (except for the relative minimum cut-off, $a_{\rm dpl}/L$). As a rough summary of the figure, the relative scale parameter of the dpl is $\theta/L \approx 0.005$, and the maximum relative frequency is $n_{max}/L \approx 0.05$ (both with considerable dispersion). Curiously, 0.05 is also, approximately, the relative frequency of the most common word in English, which is "the"[59].

We also see in the figure that the exponent β_1 ranges between 1 and 2, for most of the composers. This power-law regime coincides qualitatively with what has been found in linguistics, using large corpora (where the exponent β_1 seems to be between 1.4[33,34] and 1.6[32,60,61]), but we are not aware that it was reported before in music. In addition, we observe that β_2 ranges mostly between 2 and 3, (remember that in order to consider that we have a Zipfian tail, β_2 should be between 1.8 and 2.2, roughly). As the composers are ranked by decreasing L, we also observe that the smaller L, the larger the dispersion in the values of β_2 and β.

However, there are a number of cases (28, 37%) in which the double power-law fit is not appropriate, and this can be due to two main reasons: either the two power-law regimes do not overlap (6 composers) and thus the fit is rejected, or the first power-law regime is meaningless and thus the fit is also rejected. The latter can arise from two subcases: from a too short fitting range (2 composers), or from the fact that both exponents are nearly the same, i.e., $\beta_1 \simeq \beta_2$, and then the existence of two power-law regimes cannot be established (20 composers).

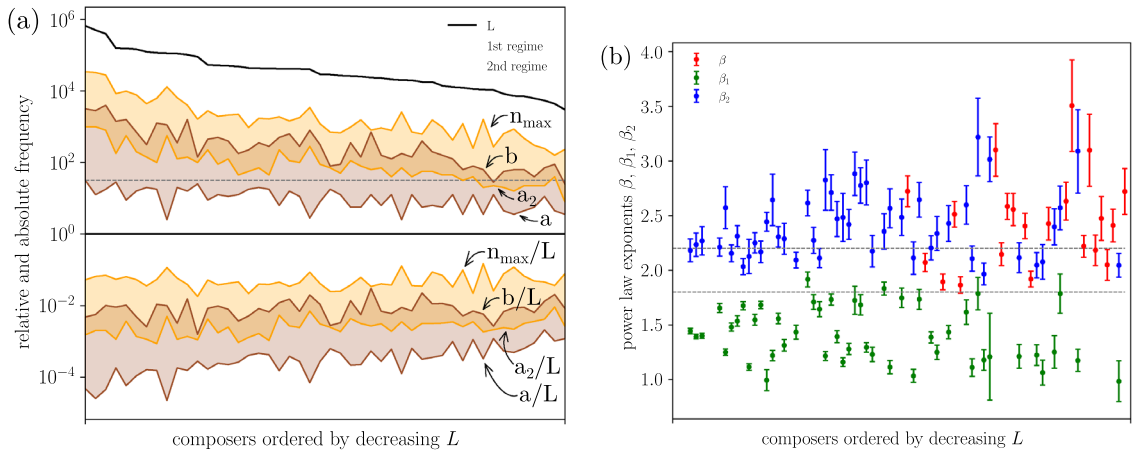

Figure 4. Results of the double power-law fit for each composer for which this is not rejected. (**a**) Fitting ranges sorted by decreasing L. Relative fitting ranges, in terms of cut-off frequencies divided by L, are also shown below. The dashed line marks the $10^{1.5} \simeq 32$ limit. The middle region not in the legend corresponds to the overlap between both regimes. (**b**) Double power-law exponents β_1 and β_2. The value of the exponent β for the simple power law (when this is not rejected) is also included. Horizontal dashed lines delimit the Zipf's range for the exponents β_2 and β.

Composer	V	a	b	β_1	a_2	n_{max}	ℓ	v_2	β_2
Monteverdi	247	8	319	1.895 ± 0.126	10	319	1.50	68	**2.048 ± 0.142**
Scheidt	233	11	475	2.105 ± 0.155	11	475	1.63	76	**2.184 ± 0.139**
Albrechtsberger	511	16	282	2.634 ± 0.205	16	282	1.25	84	2.632 ± 0.172
Paganini	713	10	351	1.638 ± 0.097	8	351	1.65	166	**1.919 ± 0.072**
Cramer	657	11	56	1.786 ± 0.249	20	218	1.04	50	3.098 ± 0.330
Berlioz	805	11	398	1.844 ± 0.097	18	602	1.53	109	**2.146 ± 0.104**
Gottschalk	1035	18	794	1.908 ± 0.086	20	2410	2.08	174	**2.072 ± 0.084**
Bizet	782	11	751	1.709 ± 0.080	13	751	1.78	155	**1.864 ± 0.075**
Bruckner	1602	40	891	2.338 ± 0.157	35	895	1.40	132	2.436 ± 0.119
Guilmant	522	9	276	2.349 ± 0.159	9	276	1.49	92	2.410 ± 0.149
Janáček	874	16	251	2.323 ± 0.148	18	251	1.15	125	2.556 ± 0.147
Joplin	634	8	325	2.154 ± 0.116	8	325	1.61	119	2.220 ± 0.099
Satie	912	11	781	1.728 ± 0.073	13	781	1.79	180	**1.894 ± 0.073**
Karg-Elert	752	9	101	2.594 ± 0.225	9	101	1.05	75	2.720 ± 0.210
Respighi	891	16	348	2.304 ± 0.139	16	348	1.34	122	2.404 ± 0.116
Rajmáninov	2231	13	380	2.375 ± 0.099	18	380	1.33	146	2.512 ± 0.117
Schoenberg	997	10	335	2.261 ± 0.212	9	335	1.58	68	2.474 ± 0.200
Bartók	1766	13	226	2.503 ± 0.136	13	226	1.25	153	2.584 ± 0.119
Médtner	1233	13	74	2.825 ± 0.295	16	74	0.67	49	3.504 ± 0.418
Gershwin	1827	16	50	2.398 ± 0.291	22	154	0.84	81	3.100 ± 0.241
Prokófiev	1171	13	291	2.345 ± 0.164	13	291	1.36	93	2.426 ± 0.149
Stravinsky	2442	28	296	2.574 ± 0.159	28	296	1.02	154	2.723 ± 0.141
All 76	4085	40	3980	1.334 ± 0.017	10,000	254,000	1.40	118	2.369 ± 0.120

Table 2. Results of the double power-law fit when this fit is rejected, either because the truncated power-law regime has a too short range or because there are not two different power-law regimes (case of non-overlapping fitting ranges is not included, except for the global dataset)[58]. The simple power law fit is accepted (with $\beta = \beta_2$) for all composers but one (Bruckner, for which $a_{pl} = a_2 = 35$). The number of orders of magnitude of this fit is given by $\ell = \log_{10}(n_{max}/a_2)$. The number of types included in the fit is v_2. The uncertainties in β_1 and β_2 are given by the standard deviation of the maximum likelihood estimation. The rest of variables are explained in the main text. The Zipfian subcases ($1.8 \leq \beta \leq 2.2$) are marked in bold.

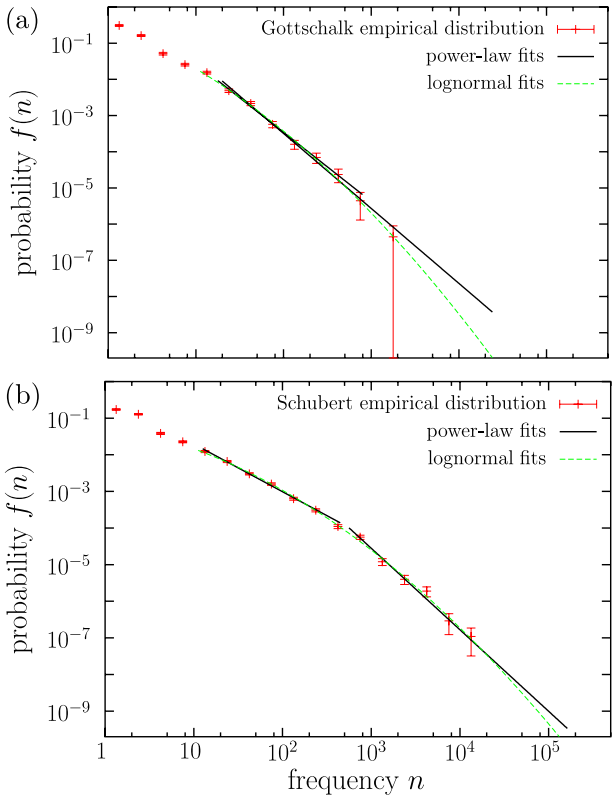

Figure 5. Empirical probability density of codeword frequency for (**a**) Gottschalk and (**b**) Schubert, together with their fits. Gottschalk is the composer with the largest logarithmic fitting range for the simple power law, nevertheless, the (logarithmic) fitting range for the lognormal is larger. Schubert has the largest number of types in the lognormal fitting range (1102). Failed double-power law fits are also shown.

Table 2 displays the results for these two subcases, clearly showing the failure of the double power law ($b/a_{\rm dpl}$ small or $\beta_1 \simeq \beta_2$; the table makes these statements quantitative). The table also shows that, when the first power-law regime (the truncated one) is meaningless, the simple power law provides a good fit for all composers but one (21 composers; Bruckner, with $a_{\rm pl} = 35$, is excluded). In most of the cases, the fitted power-law exponent ranges from $\beta \simeq 1.9$ to 2.7 (also displayed at Fig. 4). Figure 5a shows one case of the failure of the double power-law fit and the validity of the simple power law.

Coming back to the case when the double power law is rejected because no overlap between the two regimes exists (6 composers only, as Schubert in Fig. 5b), the simple power law is rejected as well, as the value of $a_{\rm pl}$ turns out to be too high ($a_{\rm pl} \gg 32$) and the power-law fit only includes the tail of the empirical distribution. Nevertheless, we will see that in 5 of these cases the lognormal provides a good fit (see Fig. 5b); the lognormal becomes the best fit then, as represented in Fig. 6. The exception to this is given by Bach, who turns out to be the only composer for which none of the three distributions is able to fit the data with $a \leq 32$.

We show Bach's empirical distribution of frequencies together with its different (failed) fits in Fig. 7a; there one can see that the two power-law regimes are far from overlapping, and that the tail exponent is rather large ($\beta \simeq 3.7$) and limited to a narrow range in frequency. Note that Bach's distribution could be fitted by a power-law body followed (with overlap) by a lognormal tail, but such a distribution is not considered in this paper (of course, other distributions with a supercritical bump could be considered as well[62]).

As another example of a failed double power-law fit, due to the non-overlapping of the two regimes, we consider in Fig. 7b the global dataset of the 76 composers, although it is remarkable that in this case the double power-law fit is not far from being "accepted" (the region of no overlap has a rather short range). The two power law exponents would be $\beta_1 = 1.33$ and $\beta_2 = 2.4$ (this case has also been included in Table 2).

Lognormal fit. Regarding the lognormal fit, it is rejected for only 9 composers (12%), due to a value of $a_{\rm ln}$ too large (greater than 32). These are Couperin, Dandrieu, Bach, Händel, Haydn, Mozart, Beethoven, Alkan, and Shostakóvich, for which the double-power law provides a good fit, except for Bach. The values of the cut-offs for the remaining 67 are shown in Fig. 6 (this includes all cases that were not fitted by the double power law, except Bach). When comparing the double power law with the lognormal, we find that in many cases the two distributions provide roughly similar fits (see for instance Fig. 3a; in Fig. 3b, despite some similarity, the fitting ranges are clearly different). Out of the 40 composers that are fitted by both distributions, for 25 of them the double power

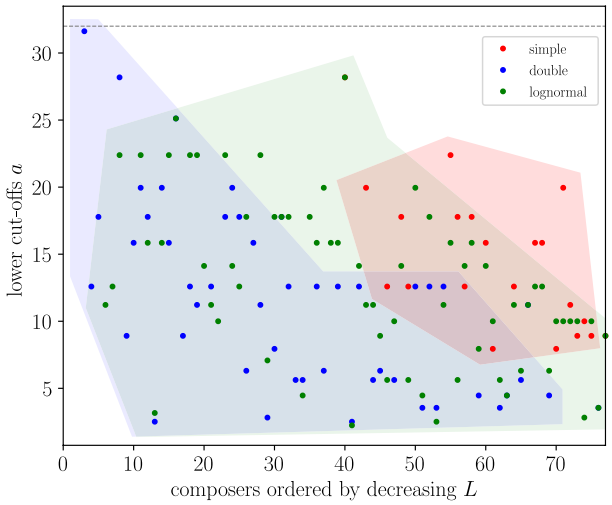

Figure 6. Results for each composer of the lower cut-off for each fit, when this is not rejected. Shadowed regions group the results for each fit. A lower value of a implies a larger fitting range; the distribution with the lowest value of a is the preferred best fit for each composer.

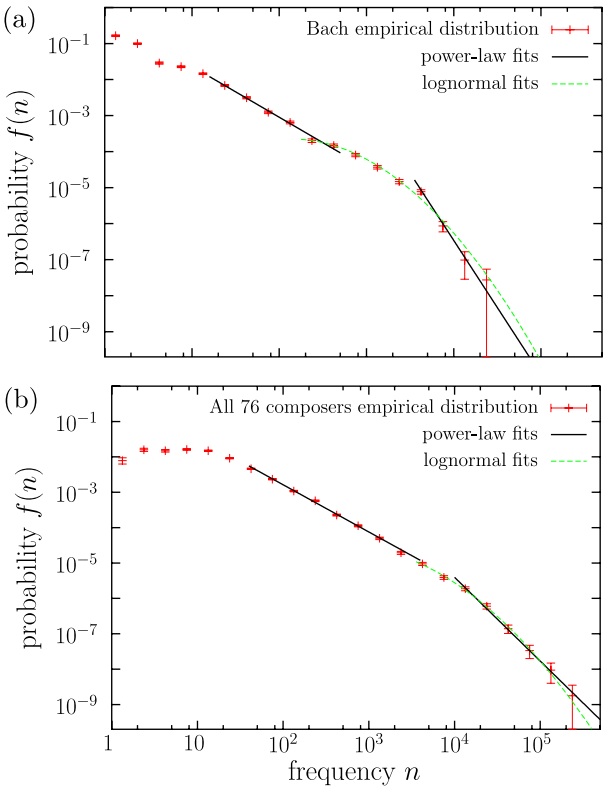

Figure 7. Empirical probability density of codeword frequency for (a) Bach, and for (b) the global dataset with all the 76 composers, together with power-law fits and lognormal fit. In both cases the double-power law fit is rejected, as the two fitted regimes do not overlap, and also the lognormal fit is unsatisfactory, as it excludes a significant part of the types (i.e., the value of a turns out to be too large).

	Simple pl	Double pl	Lognormal
Best fits	7	33	35
Best Zipf	1	8	–
Good fits	21	48	67
Good Zipf	7	15	–
Reason for rejection	$a_{\mathrm{pl}} > 32$ (55)	No overlap (6)	$a_{\mathrm{ln}} > 32$ (9)
		$b \approx a_{\mathrm{dpl}}$ (2)	
		$\beta_1 \approx \beta_2$ (20)	

Table 3. Summary of the fits for the 76 composers. First row counts distributions yielding the best fits (in terms of fitting more data); Second row counts the Zipf subcases among the best fits (exponents β and β_2 between 1.8 and 2.2); third counts distributions yielding good fits (fit not rejected but not necessarily the best); and fourth counts the Zipf subcases among the good fits. The statistics for the reason of rejection (bad fits) are also included (in brackets).

law provides the best fit (in the sense that it fits a larger fraction of the data), and for 15 of them the situation is reversed (Fig. 6).

All composers fitted by the simple power law (21) are also fitted by the lognormal, and the fit provided by the lognormal is better for 14 of these 21 composers (for the remaining 7 composers the simple power law is able to fit a larger range than the lognormal; this can be seen in Fig. 6). Figure 5a is a good illustration of the case when the lognormal outperforms the simple power law. Figure 5b shows another case, in which both the simple and the double power-law fail but not the lognormal.

Discussion and conclusions

Table 3 summarizes our results. Recapitulating, the lognormal is the distribution that gives a good fit for more composers (67 out of 76, 88%), followed by the double power law (48, 63%), and by the simple power law (21, 28%). If one considers the simple power law as a special case of the double power law, one would obtain a larger number for the latter distribution (69, 91%), thus accounting for roughly the same percentage of fits as the lognormal. Note that good simple power-law fits arise for composers with low representativity in the corpus (low value of L and smaller absolute frequencies); so, we conjecture that increasing the number of pieces by these composers would unveil a range of smaller relative frequencies which could be well fitted by another power-law regime and thus a double power law would emerge in these cases (in other words, for the number of pieces in the corpus of these composers, the value of θ would be too small to be detectable).

Among the double power-law fits, there are only 15 cases that can be considered good "double Zipf" fits (with $1.8 \leq \beta_2 \leq 2.2$), and among the simple power-law fits, there are 7 with an exponent in the Zipf range. So, out of 76 composers only 22 (29%) can be considered to follow Zipf's law in the sense of a good fit (although a lognormal can provide a better fit for some of them). Considering best fits only, Zipf's law only arises for 9 composers (12%).

If we ask instead, not which distributions fit well the data but for the distribution that yields the best fit (in the sense of fitting more data, remember), Table 3 shows that the lognormal does it for 35 composers, the double power law for 33, and the simple power law for 7; one case (Bach) is not fitted well by any of the three distributions. Figure 8 quantifies these differences in chronological order. Except for the fifteenth century, the double power law slightly dominates about the first half of the data, but after Beethoven, the lognormal takes over. Interestingly, we observe that the simple power law only starts to appear as a best fit in the second part of the corpus, around the eighteenth century, and corresponds to composers with little representation in the corpus (low L). Nevertheless, composers with low L can also be found in the first part of the corpus, but the simple power law does not provide the best fit in any of such cases.

For a small number of composers (six, as well as for the global aggregated dataset), the double power law is rejected because the two power-law regimes do not overlap. This is an unfortunate situation, as the two power-law regimes exist, but the double power law turns out to be too sharp in its transition from one power-law regime to the other at $n = \theta$. A smoother version of the double power law (sdpl) could be used instead,

$$f_{\mathrm{sdpl}}(n) = \frac{\gamma}{B\theta}\left(\frac{\theta}{n}\right)^{\beta_1}\left[\frac{1}{1+(n/\theta)^{\gamma}}\right]^{\frac{\beta_2-\beta_1}{\gamma}}, \text{ for } n \geq a$$

where B refers to the incomplete beta function

$$B\left(\frac{\theta^{\gamma}}{\theta^{\gamma}+a^{\gamma}}; \frac{\beta_2-1}{\gamma}, \frac{1-\beta_1}{\gamma}\right)$$

and the extra parameter $\gamma > 0$ controls the sharpness of the transition[63]. The limit $\gamma \to \infty$ recovers the (infinitely sharp) double power law and the limit $a = \beta_1 = 0$ with $\gamma = 1$ leads to the Pareto distribution. In general, one could fix $\gamma = 1$ to reduce the number of free parameters, but this seems a rather arbitrary decision. An even more convenient option would be to consider the discrete version of this distribution (as the data are discrete). These extensions are left for future research.

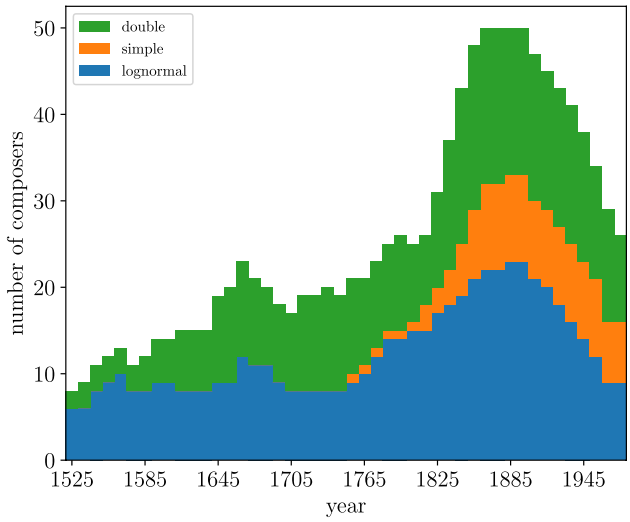

Figure 8. Number of composers best fitted by each fit, as a function of year (year is determined as the mean between birth plus 20 and death). Numbers have been smoothed: for each interval of 10 years we take a window of 150 years and count the number of fits in the window.

A relevant issue in our analysis is the construction of the harmonic codewords from the scores. The discretization procedure looks for the presence (1) or absence (0) of each pitch class in every beat of the score. This involves two somewhat arbitrary decisions: first, presence and absence are decided in terms of an arbitrary threshold applied to the (nonbinary, continuous) chromas; second, it is taken for granted that the fundamental time unit is the beat. We have tested the robustness of our results against the change in these arbitrary parameters, repeating the process for different discretization thresholds and different selections of the elementary time unit. It is clear that the increase of the time unit (e.g., from one beat to two beats) leads to a reduction in the number of tokens L and subsequently in the values of the frequencies n. Obviously, scale parameters (such as θ and e^{μ}) are strongly affected under such a change; however, a rescaled plot such as the one in Fig. 2b shows that the shape of the distributions is robust and remains nearly the same, also when the threshold is changed. As the distinction between lognormals and power laws depends on the shape and not on the scale of the distributions, our conclusions regarding the lack of universality and the poor fulfilment of Zipf's law in music do not change (in a previous study utilizing the same discretization method, we also found the definition of the codeword threshold and the temporal unit rather irrelevant given a reasonable range[53]).

In summary, we find that the usage of harmonic vocabulary in classical composers may seem universal-like at a qualitative level (see Fig. 2b), but this paradigm fails when one approaches the issue in a quantitative statistical way. Not only universal parameters to describe the distributions do not exist, but different distributions (lognormal and power laws) fit better different composers. In particular, the Zipf picture (a power-law tail with exponent in the approximate range 1.8–2.2) only applies to a reduced subset of composers (9 best fits, out of 76 composers, and 22 good fits, out of 69 power law and double power-law good fits and 76 composers, see Table 3). Although some degree of universality has been claimed in complex systems in analogy to statistical physics[64], detailed analyses in some particular systems have shown a diversity of parameters and distributions for some particular systems[22,65].

Our work can be put into the wider context of quantifying the universality of scaling laws. This is directly related to the use of proper statistical methodologies in complex-systems science, where some controversies have arisen in recent years. For example, the application of the ideas of allometric scaling to urban science[66] has raised important concerns[67,68] (see also Ref.[69] for a revision of the original problem). Methods of fitting power-law distributions have been criticized[19] and re-criticized[25,26,70,71]. The important role played by statistical dependence when fitting has been pointed out in Ref.[21]. In general, one important lesson that emerges from the study of complex systems is that these have to be characterized in probabilistic terms, so a precise description of their stochastic or probabilistic properties, together with adequate statistical tools, becomes mandatory.

Received: 28 October 2021; Accepted: 21 January 2022
Published online: 16 February 2022

References

1. Longair, M. S. *Theoretical Concepts in Physics: An Alternative View of Theoretical Reasoning in Physics* 2nd edn. (Cambridge University Press, 2003).
2. Watts, D. J. A twenty-first century science. *Nature* **445**, 489–489 (2007).
3. West, G. *Scale: The Universal Laws of Life and Death in Organisms, Cities and Companies* (Penguin Press, 2017).
4. Camacho, J. & Solé, R. V. Scaling in ecological size spectra. *Europhys. Lett.* **55**, 774–780 (2001).
5. Axtell, R. L. Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820 (2001).

6. Adamic, L. A. & Huberman, B. A. Zipf's law and the Internet. *Glottometrics* **3**, 143–150 (2002).
7. Pueyo, S. & Jovani, R. Comment on "A keystone mutualism drives pattern in a power function". *Science* **313**, 1739c–1740c (2006).
8. Li, W. Zipf's law everywhere. *Glottometrics* **5**, 14–21 (2002).
9. Baayen, H. *Word Frequency Distributions* (Kluwer, 2001).
10. Baroni, M. Distributions in text. In *Corpus Linguistics: An International Handbook* Vol. 2 (eds Lüdeling, A. & Kytö, M.) 803–821 (Mouton de Gruyter, 2009).
11. Zanette, D. Statistical patterns in written language. arXiv:14123336v1, (2014).
12. Piantadosi, S. T. Zipf's law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
13. Moreno-Sánchez, I., Font-Clos, F. & Corral, A. Large-scale analysis of Zipf's law in English texts. *PLoS ONE* **11**(1), e0147073 (2016).
14. Stephens, G. J. & Bialek, W. Statistical mechanics of letters in words. *Phys. Rev. E* **81**, 066119 (2010).
15. Corral, A. & García del Muro, M. From Boltzmann to Zipf through Shannon and Jaynes. *Entropy* **22**, 179 (2020).
16. Gerlach, M. & Font-Clos, F. A standardized Project Gutenberg Corpus for statistical analysis of natural language and quantitative linguistics. *Entropy* **22**(1), 126 (2020).
17. Bauke, H. Parameter estimation for power-law distributions by maximum likelihood methods. *Eur. Phys. J.* **58**, 167–173 (2007).
18. White, E. P., Enquist, B. J. & Green, J. L. On estimating the exponent of power-law frequency distributions. *Ecology* **89**, 905–912 (2008).
19. Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
20. Deluca, A. & Corral, A. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophys.* **61**, 1351–1394 (2013).
21. Gerlach, M. & Altmann, E. G. Testing statistical laws in complex systems. *Phys. Rev. Lett.* **122**, 168301 (2019).
22. Corral, A. & González, A. Power law distributions in geoscience revisited. *Earth Space Sci.* **6**(5), 673–697 (2019).
23. Font-Clos, F., Boleda, G. & Corral, A. A scaling law beyond Zipf's law and its relation to Heaps' law. *New J. Phys.* **15**, 093033 (2013).
24. Corral, A., Boleda, G. & Ferrer-i-Cancho, R. Zipf's law for word frequencies: Word forms versus lemmas in long texts. *PLoS ONE* **10**(7), e0129031 (2015).
25. Corral, A., Font, F. & Camacho, J. Non-characteristic half-lives in radioactive decay. *Phys. Rev. E* **83**, 066103 (2011).
26. Voitalov, I., van der Hoorn, P., van der Hofstad, R. & Krioukov, D. Scale-free networks well done. *Phys. Rev. Res.* **1**, 033034 (2019).
27. Malevergne, Y., Pisarenko, V. & Sornette, D. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Phys. Rev. E* **83**, 036111 (2011).
28. Corral, A., Udina, F. & Arcaute, E. Truncated lognormal distributions and scaling in the size of naturally defined population clusters. *Phys. Rev. E* **101**, 042312 (2020).
29. Mandelbrot, B. On the theory of word frequencies and on related Markovian models of discourse. In *Structure of Language and its Mathematical Aspects* (ed. Jakobson, R.) 190–219 (American Mathematical Society, 1961).
30. Corral, A., Serra, I. & Ferrer-i-Cancho, R. Distinct flavors of Zipf's law and its maximum likelihood fitting: Rank-size and size-distribution representations. *Phys. Rev. E* **102**, 052113 (2020).
31. Mehri, A. & Jamaati, M. Variation of Zipf's exponent in one hundred live languages: A study of the Holy Bible translations. *Phys. Lett. A* **381**(31), 2470–2477 (2017).
32. Ferrer i Cancho, R. & Solé, R. V. Two regimes in the frequency of words and the origin of complex lexicons: Zipf's law revisited. *J. Quant. Linguist.* **8**(3), 165–173 (2001).
33. Montemurro, M. A. Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Physica A* **300**(3–4), 567–578 (2001).
34. Corral, A. & Serra, I. The brevity law as a scaling law, and a possible origin of Zipf's law for word frequencies. *Entropy* **22**(2), 224 (2020).
35. Williams, J. R., Bagrow, J. P., Danforth, C. M. & Dodds, P. S. Text mixing shapes the anatomy of rank-frequency distributions. *Phys. Rev. E* **91**, 052811 (2015).
36. Perc, M. Beauty in artistic expressions through the eyes of networks and physics. *J. R. Soc. Interface* **17**(164), 20190686 (2020).
37. Ball, P. *The Music Instinct* (Oxford Univ. Press, 2010).
38. Jet Propulsion Laboratory, NASA. The golden record. https://voyager.jpl.nasa.gov/golden-record/.
39. Wikipedia. Voyager golden record. https://en.wikipedia.org/wiki/Voyager_Golden_Record.
40. Wikipedia. List of most-followed Twitter accounts. https://en.wikipedia.org/wiki/List_of_most-followed_Twitter_accounts.
41. Zanette, D. H. Zipf's law and the creation of musical context. *Mus. Sci.* **10**, 3–18 (2004).
42. Patel, A. D. *Music, Language, and the Brain* (Oxford University Press, 2008).
43. Zanette, D. Playing by numbers. *Nature* **453**, 988–989 (2008).
44. Manaris, B., Purewal, T. & McCormick, C. Progress towards recognizing and classifying beautiful music with computers: MIDI-encoded music and the Zipf-Mandelbrot law. In *Proceedings IEEE SoutheastCon 2002 (Cat. No.02CH37283)* 52–57 (2002).
45. Liu, L., Wei, J., Zhang, H., Xin, J. & Huang, J. A statistical physics view of pitch fluctuations in the classical music from Bach to Chopin: Evidence for scaling. *PLoS ONE* **8**(3), 1–6 (2013).
46. Beltrán del Río, M., Cocho, G. & Naumis, G. G. Universality in the tail of musical note rank distribution. *Physica A* **387**, 5552–5560 (2008).
47. Serrà, J., Corral, A., Boguñá, M., Haro, M. & Arcos, JLl. Measuring the evolution of contemporary western popular music. *Sci. Rep.* **2**, 521 (2012).
48. Haro, M., Serrà, J., Herrera, P. & Corral, A. Zipf's law in short-time timbral codings of speech, music, and environmental sound signals. *PLoS ONE* **7**, e33993 (2012).
49. Hennig, H., Fleischmann, R. & Geisel, T. Musical rhythms: The science of being slightly off. *Phys. Today* **65**(7), 64–65 (2012).
50. Chromagramer. https://github.com/MarcSerraPeralta/chromagramer.
51. kunstderfuge.com. The largest resouce of classical music in .mid files. http://www.kunstderfuge.com.
52. González-Espinoza, A., Martínez-Mekler, G. & Lacasa, L. Arrow of time across five centuries of classical music. *Phys. Rev. Res.* **2**, 033166 (2020).
53. Serra-Peralta, M., Serrà, J. & Corral, A. Heaps' law and vocabulary richness in the history of classical music harmony. *EPJ Data Sci.* **10**, 40 (2021).
54. Simon, H. A. On a class of skew distribution functions. *Biometrics* **42**, 425–440 (1955).
55. Cattuto, C., Loreto, V. & Pietronero, L. Semiotic dynamics and collaborative tagging. *Proc. Natl. Acad. Sci. USA* **104**(5), 1461–1464 (2007).
56. Corral, A. Scaling in the timing of extreme events. *Chaos. Solit. Fract.* **74**, 99–112 (2015).
57. Corral, A. & Font-Clos, F. Dependence of exponents on text length versus finite-size scaling for word-frequency distributions. *Phys. Rev. E* **96**, 022318 (2017).
58. In the fitting of the truncated power law, the method we use [20,22] considers the possibility of $b > n_{\max}$ (the truncation point could be above the maximum value of the random variable). for the case we are interested in (the truncated power law giving one of the regimes of the double power law), this does not apply, and therefore we have ignored this fact.
59. Google Books Ngram Viewer. https://books.google.com/ngrams.
60. Petersen, A. M., Tenenbaum, J. N., Havlin, S., Stanley, H. E. & Perc, M. Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Sci. Rep.* **2**, 943 (2012).
61. Gerlach, M. & Altmann, E. G. Stochastic model for the vocabulary growth in natural languages. *Phys. Rev. X* **3**, 021006 (2013).

62. Corral, A., Garcia-Millan, R., Moloney, N. R. & Font-Clos, F. Phase transition, scaling of moments, and order-parameter distributions in Brownian particles and branching processes with finite-size effects. *Phys. Rev. E* **97**, 062156 (2018).
63. The incomplete gamma function arises immediately with the change of variables $t = n^\gamma/(\theta^\gamma + n^\gamma)$.
64. Stanley, H. E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
65. Hantson, S., Pueyo, S. & Chuvieco, E. Global fire size distribution: From power law to log-normal. *Int. J. Wildl. Fire* **25**, 403–412 (2016).
66. Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913 (2010).
67. Arcaute, E. *et al.* Constructing cities, deconstructing scaling laws. *J. R. Soc. Interface* **12**(102), 20140745 (2015).
68. Leitão, J. C., Miotto, J. M., Gerlach, M. & Altmann, E. G. Is this scaling nonlinear?. *R. Soc. Open Sci.* **3**(7), 150649 (2016).
69. Ballesteros, F. J. *et al.* On the thermodynamic origin of metabolic scaling. *Sci. Rep.* **8**, 1448 (2018).
70. Barabási, A.-L. Love is all you need. Clauset's fruitless search for scale-free networks. https://www.barabasilab.com/post/love-is-all-you-need (2018).
71. Corral, A. Tail of the distribution of fatalities in epidemics. *Phys. Rev. E* **103**, 022315 (2021).

Acknowledgements

Josep Lluís Arcos provided support to start this collaboration; Juan Camacho contributed with discussions about statistical physics; V. Navas-Portella facilitated his R-code for fitting; I. Moreno-Sánchez participated in a preliminary stage of this research. Support from projects FIS2015-71851-P and PGC-FIS2018-099629-B-I00 from Spanish MINECO and MICINN, as well as from CEX2020-001084-M from the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D, is acknowledged. M. S.-P.'s participation has been possible thanks to the Internship Program of the Centre de Recerca Matemàtica and Treball Campus of the UAB.

Author contributions

J. S. and A. C. discussed the original idea, M. S.-P. wrote the code and did a first analysis of the data, A. C. did a second analysis, J. S. and A. C. supervised the results, all authors interpreted the results. A.C. wrote the first draft of the manuscript and all authors revised the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-06137-3.

Correspondence and requests for materials should be addressed to Á.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.