

## Zipf's law, power laws and maximum entropy

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2013 New J. Phys. 15 043021

(<http://iopscience.iop.org/1367-2630/15/4/043021>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

### Download details:

This content was downloaded by: count0

IP Address: 134.102.186.160

This content was downloaded on 13/03/2014 at 18:01

Please note that [terms and conditions apply](#).

## Zipf's law, power laws and maximum entropy

**Matt Visser**

School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

E-mail: [matt.visser@msor.vuw.ac.nz](mailto:matt.visser@msor.vuw.ac.nz)

*New Journal of Physics* **15** (2013) 043021 (13pp)

Received 17 January 2013

Published 16 April 2013

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/15/4/043021

**Abstract.** Zipf's law, and power laws in general, have attracted and continue to attract considerable attention in a wide variety of disciplines—from astronomy to demographics to software structure to economics to linguistics to zoology, and even warfare. A recent model of *random group formation* (RGF) attempts a general explanation of such phenomena based on Jaynes' notion of *maximum entropy* applied to a particular choice of *cost function*. In the present paper I argue that the specific cost function used in the RGF model is in fact unnecessarily complicated, and that power laws can be obtained in a much simpler way by applying maximum entropy ideas directly to the Shannon entropy subject only to a single constraint: that the average of the logarithm of the observable quantity is specified.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Contents**

<b>1. Introduction</b>	<b>2</b>
<b>2. Power laws in infinite state space</b>	<b>3</b>
<b>3. Power laws in finite state space</b>	<b>5</b>
<b>4. Zipf's law in finite state space</b>	<b>6</b>
<b>5. Hybrid geometric/power models in infinite state space</b>	<b>7</b>
<b>6. Zipf's law: geometric version in infinite state space</b>	<b>8</b>
<b>7. Very general Gibbs-like model</b>	<b>9</b>
<b>8. Non-Shannon entropies</b>	<b>10</b>
<b>9. Summary and discussion</b>	<b>11</b>
<b>Acknowledgments</b>	<b>12</b>
<b>References</b>	<b>12</b>

**1. Introduction**

Zipf's law [1–3], and power laws in general [4–6], have and continue to attract considerable attention in a wide variety of disciplines—from astronomy to demographics to software structure to economics to zoology, and even to warfare [7]. Typically one is dealing with integer-valued observables (numbers of objects, people, cities, words, animals, corpses), with  $n \in \{1, 2, 3, \dots\}$ . Sometimes the range of values is allowed to be infinite (at least in principle), sometimes a hard upper bound  $N$  is fixed (e.g., total population if one is interested in subdividing a fixed population into sub-classes). Particularly interesting probability distributions are probability laws of the form

- Zipf's law:  $p_n \propto 1/n$ ;
- power laws:  $p_n \propto 1/n^z$ ;
- hybrid geometric/power laws:  $p_n \propto w^n/n^z$ .

Specifically, a recent model of *random group formation* (RGF), see [8], attempts a general explanation of such phenomena based on Jaynes' notion of *maximum entropy* [9–13] applied to a particular choice of *cost function* [8]. (For recent related work largely in a demographic context see [14–18]. For related work in a fractal context implemented using an iterative framework see [19].)

In the present paper I shall argue that the specific cost function used in the RGF model is in fact unnecessarily complicated (in fact RGF most typically leads to a hybrid geometric/power law, not a pure power law), and that power laws can be obtained in a much simpler way by applying maximum entropy ideas directly to the Shannon entropy itself [20, 21] subject only to a single constraint: that the average of the logarithm of the observable quantity is specified. Similarly, I would argue that (at least as long as the main issue one is interested in is 'merely' the minimum requirements for obtaining a power law) the appeal to a fractal framework and the iterative model adopted by Pastor-Satorras and Wagensberg [19] is also unnecessarily complicated.

To place this observation in perspective, I will explore several variations on this theme, modifying both the relevant state space and the number of constraints, and will briefly discuss the relevant special functions of mathematical physics that one encounters (zeta functions, harmonic series, poly-logarithms). I shall also discuss an extremely general Gibbs-like model, and the use of non-Shannon entropies (the Rényi [22] and Tsallis [23] entropies and their generalizations). There is a very definite trade-off between simplicity and generality, and I shall very much focus on keeping the discussion as technically simple as possible, and on identifying the simplest model with minimalist assumptions.

## 2. Power laws in infinite state space

Let us define the set of observable quantities to be positive integers  $n \in \{1, 2, 3, \dots\}$ , without any *a priori* upper bound. The maximum entropy approach [9–13] seeks to estimate the probabilities  $p_n$  by maximizing the Shannon entropy [20, 21]

$$S = - \sum_n p_n \ln p_n \quad (1)$$

subject to a (small) number of constraints/cost functions—representing our limited state of knowledge regarding the underlying process. For example, the RGF model of Baek *et al* [8] uses one constraint and one (relatively complicated) cost function [8], in addition to the ‘trivial’ normalization constraint  $\sum_n p_n = 1$ . Instead, let us consider the *single* constraint

$$\langle \ln n \rangle \equiv \sum_{n=1}^{\infty} p_n \ln n = \chi. \quad (2)$$

Let us now maximize the Shannon entropy subject to this constraint. This is best done by introducing a Lagrange multiplier  $z$  corresponding to the constraint  $\langle \ln n \rangle$ , plus a second Lagrange multiplier  $\lambda$  corresponding to the ‘trivial’ normalization constraint, and considering the quantity

$$\hat{S} = -z \left( \sum_{n=1}^{\infty} p_n \ln n - \chi \right) - \lambda \left( \sum_{n=1}^{\infty} p_n - 1 \right) - \sum_{n=1}^{\infty} p_n \ln p_n. \quad (3)$$

Of course there is no loss of generality in redefining the Lagrange multiplier  $\lambda \rightarrow Z$  as follows:

$$\hat{S} = -z \left( \sum_{n=1}^{\infty} p_n \ln n - \chi \right) - (\ln Z - 1) \left( \sum_{n=1}^{\infty} p_n - 1 \right) - \sum_{n=1}^{\infty} p_n \ln p_n. \quad (4)$$

Varying with respect to the  $p_n$  yields the extremality condition

$$-z \ln n - \ln Z - \ln p_n = 0 \quad (5)$$

with explicit solution

$$p_n = \frac{n^{-z}}{\zeta(z)}, \quad Z = \zeta(z), \quad z > 1. \quad (6)$$

Here  $\zeta(z)$  is the Riemann zeta function [24–27], a relatively common and well-known special function, and the condition  $z > 1$  is required to make the sum  $\sum_{n=1}^{\infty} n^{-z} = \zeta(z)$  converge. This is enough to tell you that one will never *exactly* reproduce Zipf’s law ( $z = 1$ ) in this particular

manner, though one can get arbitrarily close. The value of the Lagrange multiplier  $z$  (which becomes the exponent in the power law) is determined self-consistently in terms of  $\chi$  by demanding

$$\chi(z) = \langle \ln n \rangle = \frac{\sum_{n=1}^{\infty} n^{-z} \ln n}{\zeta(z)} = -\frac{d\zeta(z)/dz}{\zeta(z)} = -\frac{d \ln \zeta(z)}{dz}. \quad (7)$$

Then  $z \in (1, \infty)$  while  $\chi \in (0, \infty)$ . For practical calculations it is often best to view the exponent  $z$  as the single free parameter and  $\chi(z)$  as the derived quantity, but this viewpoint can easily be inverted if desired. Near  $z = 1$  we have the analytic estimate

$$\chi(z) = \langle \ln n \rangle = \frac{1}{z-1} - \gamma + \mathcal{O}(z-1), \quad (8)$$

where  $\gamma$  denotes Euler's constant. At maximum entropy, imposing the extremality condition and summing, we have

$$\hat{S}(z) = S(z) = -\sum_{n=1}^{\infty} p_n \ln p_n = \ln \zeta(z) + z\chi(z). \quad (9)$$

Near  $z = 1$  we have the analytic estimate

$$\hat{S}(z) = S(z) = \frac{1}{z-1} + \ln\left(\frac{1}{z-1}\right) + 1 - \gamma + \mathcal{O}(z-1). \quad (10)$$

For future reference, it is useful to observe that [27, see p 118]

$$\zeta(z) = \frac{1}{z-1} + \sum_{m=1}^{\infty} \gamma_m \frac{(1-z)^m}{m!}, \quad (11)$$

where the Stieltjes constants  $\gamma_m$  satisfy

$$\sum_{n=1}^N \frac{(\ln n)^m}{n} = \frac{(\ln N)^{m+1}}{m+1} + \gamma_m + o(1), \quad \gamma_0 = \gamma. \quad (12)$$

A better estimate, using Euler–Mclaurin summation, is

$$\sum_{n=1}^N \frac{(\ln n)^m}{n} = \frac{\{\ln(N + \frac{1}{2})\}^{m+1}}{m+1} + \gamma_m + \mathcal{O}\left(\frac{\{\ln(N + \frac{1}{2})\}^m}{(N + \frac{1}{2})^2}\right). \quad (13)$$

The quick lesson to take is this: by applying maximum entropy considerations to the *single* constraint  $\langle \ln n \rangle = \chi$  you can get a pure power law with *any* exponent  $z > 1$ . Furthermore, note that the quantity

$$\exp\langle \ln n \rangle = \prod_{n=1}^{\infty} n^{p_n} \quad (14)$$

is the *geometric mean* of the integers  $\{1, 2, 3, \dots\}$  with the exponents weighted by the probabilities  $p_n$ . So one can just as easily obtain the pure power laws considered above by maximizing the entropy subject to the constraint that this geometric mean takes on a specified value.

### 3. Power laws in finite state space

If one desires an exact Zipf law (exponent  $z = 1$ , so that  $p_n \propto 1/n$ ), then because the harmonic series diverges,  $\sum_{n=1}^{\infty} 1/n = \infty$ , it is clear that something in the above formulation needs to change. Perhaps the easiest thing to do is to introduce an explicit maximum value of  $n$ , call it  $N$ , so that we take the set of observables to be positive integers  $n \in \{1, 2, 3, \dots, N\}$ . (Physicists would call this an infra-red cutoff, or large-distance cutoff.) The maximum entropy approach now amounts to considering

$$\hat{S} = -z \left( \sum_{n=1}^N p_n \ln n - \chi \right) - (\ln Z - 1) \left( \sum_{n=1}^N p_n - 1 \right) - \sum_{n=1}^N p_n \ln p_n. \quad (15)$$

Varying with respect to the  $p_n$  and maximizing again yields the same extremality condition

$$-z \ln n - \ln Z - \ln p_n = 0 \quad (16)$$

but now implying

$$p_n = \frac{n^{-z}}{H_N(z)}, \quad Z = H_N(z). \quad (17)$$

Here  $H_N(z)$  is the (reasonably well known) generalized harmonic function [27]

$$H_N(z) = \sum_{n=1}^N \frac{1}{n^z}. \quad (18)$$

Compared with the previous case the only real difference lies in the normalization function. However, because the sum is now always finite, there is no longer any constraint on the value of the exponent  $z$ , in fact we can have  $z \in (-\infty, \infty)$ . The case  $z = 1$  is Zipf's law, while  $z = 0$  is a uniform distribution, and  $z < 0$  corresponds to an 'inverted hierarchy' where large values are more common than small values. The price paid for this extra flexibility is that the model now has two free parameters, which can be chosen to be  $z$  and  $N$ . One has the self-consistency constraint

$$\chi(z, N) = \langle \ln n \rangle = \frac{\sum_{n=1}^N n^{-z} \ln n}{H_N(z)} = -\frac{dH_N(z)/dz}{H_N(z)} = -\frac{d \ln H_N(z)}{dz}. \quad (19)$$

It is easy to check that  $\chi$  is now bounded by  $\chi \in (0, \ln N)$ . At maximum entropy we now have

$$\hat{S}(z, N) = S(z, N) = -\sum_{n=1}^N p_n \ln p_n = \ln H_N(z) + z\chi(z, N). \quad (20)$$

Because this is now a two-parameter model, it will always (naively) be a 'better' fit to observational data than a single-parameter model. Sometimes (for  $z \leq 1$ ) retreating to this two-parameter model is necessary, but for  $z > 1$  the one-parameter model of the previous section ( $N \rightarrow \infty$ ) should be preferred.

#### 4. Zipf's law in finite state space

If for observational or theoretical reasons one is *certain* that  $z = 1$  (Zipf's law), then the model reduces as follows: the state space is  $n \in \{1, 2, 3, \dots, N\}$  where  $N$  is now the *only* free parameter. Then explicitly forcing  $z \rightarrow 1$  one considers

$$\hat{S} = - \left( \sum_{n=1}^N p_n \ln n - \chi \right) - (\ln Z - 1) \left( \sum_{n=1}^N p_n - 1 \right) - \sum_{n=1}^N p_n \ln p_n. \quad (21)$$

This is completely equivalent to considering

$$\hat{S} = \chi - (\ln Z - 1) \left( \sum_{n=1}^N p_n - 1 \right) - \sum_{n=1}^N p_n \ln(n p_n) \quad (22)$$

but writing the quantity to be maximized in this way hides the role of the Shannon entropy. Varying with respect to the  $p_n$  and maximizing now yields a (very) slightly different extremality condition

$$-\ln n - \ln Z - \ln p_n = 0 \quad (23)$$

and so

$$p_n = \frac{1}{H_N} \frac{1}{n}, \quad Z = H_N. \quad (24)$$

Here  $H_N$  is the (ordinary) harmonic number [27]

$$H_N = \sum_{n=1}^N \frac{1}{n}. \quad (25)$$

Then

$$\chi(N) = \langle \ln n \rangle = \frac{1}{H_N} \sum_{n=1}^N \frac{\ln n}{n}. \quad (26)$$

Furthermore, at maximum entropy

$$\hat{S}(N) = S(N) = - \sum_{n=1}^N p_n \ln p_n = \ln H_N + \chi(N). \quad (27)$$

Now we have already seen

$$H_N = \ln \left( N + \frac{1}{2} \right) + \gamma + \mathcal{O} \left( \frac{1}{(N + \frac{1}{2})^2} \right) \quad (28)$$

and

$$\sum_{n=1}^N \frac{\ln n}{n} = \frac{1}{2} \left\{ \ln \left( N + \frac{1}{2} \right) \right\}^2 + \gamma_1 + \mathcal{O} \left( \frac{\ln \left( N + \frac{1}{2} \right)}{(N + \frac{1}{2})^2} \right). \quad (29)$$

Therefore

$$\chi(N) = \langle \ln n \rangle = \frac{1}{2} \ln \left( N + \frac{1}{2} \right) + o(1) \quad (30)$$

and at maximum entropy

$$\hat{S}(N) = S(N) = \frac{1}{2} \ln \left( N + \frac{1}{2} \right) + \ln \ln \left( N + \frac{1}{2} \right) + o(1). \quad (31)$$

Note that you can use this to estimate the size  $N$  of the state space one needs to adopt in order to be compatible with the (observed) logarithmic average  $\langle \ln n \rangle$ . Indeed

$$N \approx \exp\{2\langle \ln n \rangle\} = \{\exp\langle \ln n \rangle\}^2. \quad (32)$$

This relates the size of the required state space  $N$  to the *square* of the geometric mean  $\exp\langle \ln n \rangle$ .

## 5. Hybrid geometric/power models in infinite state space

To generate a hybrid geometric/power law model, similar in output to the RGF model [8], but with considerably simpler input assumptions, simply take *both* the logarithmic average  $\langle \ln n \rangle = \chi$ , and the arithmetic average  $\langle n \rangle = \mu$ , to be specified—and then maximize the Shannon entropy subject to these two constraints, (plus the trivial normalization constraint). That is, introduce two Lagrange multipliers  $z$  and  $w$ , and maximize

$$\hat{S} = -z \left( \sum_{n=1}^{\infty} p_n \ln n - \chi \right) + \ln w \left( \sum_{n=1}^{\infty} p_n n - \mu \right) - (\ln Z - 1) \left( \sum_{n=1}^{\infty} p_n - 1 \right) - \sum_{n=1}^{\infty} p_n \ln p_n. \quad (33)$$

Varying with respect to the  $p_n$  yields

$$-z \ln n + n \ln w - \ln Z - \ln p_n = 0 \quad (34)$$

with solution

$$p_n = \frac{w^n n^{-z}}{\text{Li}_z(w)}, \quad Z = \text{Li}_z(w), \quad w < 1. \quad (35)$$

Here the normalizing constant is the well-known poly-logarithm function [24–27]

$$\text{Li}_z(w) = \sum_{n=1}^{\infty} \frac{w^n}{n^z}, \quad \text{Li}_1(w) = -\ln(1-w). \quad (36)$$

Then

$$\chi(w, z) \equiv \langle \ln n \rangle = \frac{1}{\text{Li}_z(w)} \sum_{n=1}^{\infty} \frac{w^n \ln n}{n^z} = -\frac{d \ln \text{Li}_z(w)}{dz}, \quad (37)$$

while

$$\mu(w, z) \equiv \langle n \rangle = \frac{1}{\text{Li}_z(w)} \sum_{n=1}^{\infty} \frac{w^n n}{n^z} = \frac{\text{Li}_{z-1}(w)}{\text{Li}_z(w)} = \frac{d \ln \text{Li}_z(w)}{d \ln z}. \quad (38)$$

Furthermore, at maximum entropy

$$\hat{S}(w, z) = S(w, z) = - \sum_{n=1}^{\infty} p_n \ln p_n = \ln \text{Li}_z(w) + z \chi(w, z) - \mu(w, z) \ln w. \quad (39)$$

Note that the probability function arising in this model is fully as general as that arising in the RGF model [8], but with what is perhaps a somewhat clearer interpretation.



This is because the RGF model uses what may be viewed as an unnecessarily complicated ‘cost function’, with an unnecessary degeneracy in the parameters. Indeed, from [8] one sees (in their notation, slightly different from current notation)

$$I_{\text{cost}} = \sum_k P(k) \ln[kN(k)], \quad P(k) = N(k)/N. \quad (40)$$

That is

$$I_{\text{cost}} = \sum_k P(k) \ln[kN P(k)] \quad (41)$$

$$= \sum_k P(k) \ln k + \sum_k P(k) \ln N + \sum_k P(k) \ln P(k) \quad (42)$$

$$= \langle \ln k \rangle + \ln N - S. \quad (43)$$

So the RGF cost function [8] is simply a linear combination of Shannon entropy, the logarithmic mean  $\langle \ln k \rangle$ , and a redundant constant offset  $\ln N$ . (Unfortunately the  $N$  of Baek *et al* [8] is not the same as the  $N$  used in this paper, the RGF parameter  $N$  corresponds to the number of independent realizations of the underlying statistical process one considers—it is the number of simulations, or number of universes in the statistical ensemble.) The additional complexity implicit in the RGF model can to some extent be viewed as a side-effect of forcing data into discrete ordinal boxes when one does not necessarily have good physical/mathematical/demographic reasons for knowing which particular boxes are ‘best’, how the boxes are to be assigned ordinal numbers, and where the box boundaries should most profitably be placed. Apart from the issues raised above, one could in addition explicitly restrict the state space to be finite, adding yet another free parameter ( $M$  in the language of Baek *et al* [8],  $N$  in the language of this note), but there is little purpose in doing so—the key insight is this: once the data are assigned to ordinal boxes, hybrid geometric/power laws drop out automatically and straightforwardly by maximizing the Shannon entropy subject to the two very simple constraints  $\langle \ln n \rangle = \chi$  and  $\langle n \rangle = \mu$ .

## 6. Zipf’s law: geometric version in infinite state space

If for observational or theoretical reasons one is *certain* that  $z = 1$  (Zipf’s law), but for whatever reason feels a finite state space cutoff  $N$  is inappropriate, then a geometric version of Zipf’s law can be extracted from the hybrid model. Setting  $z = 1$  the model reduces as follows: the state space is now  $n \in \{1, 2, 3, \dots\}$  while

$$\hat{S} = - \left( \sum_{n=1}^{\infty} p_n \ln n - \chi \right) + \ln w \left( \sum_{n=1}^{\infty} p_n n - \mu \right) - (\ln Z - 1) \left( \sum_{n=1}^{\infty} p_n - 1 \right) - \sum_{n=1}^{\infty} p_n \ln p_n. \quad (44)$$

This is completely equivalent to maximizing

$$\hat{S} = \chi + \ln w \left( \sum_{n=1}^{\infty} p_n n - \mu \right) - (\ln Z - 1) \left( \sum_{n=1}^{\infty} p_n - 1 \right) - \sum_{n=1}^{\infty} p_n \ln(n p_n) \quad (45)$$

but writing the quantity to be maximized in this way hides the role of the Shannon entropy. Varying with respect to the  $p_n$  yields

$$-\ln n + n \ln w - \ln Z - \ln p_n = 0 \quad (46)$$

with solution

$$p_n = \frac{1}{|\ln(1-w)|} \frac{w^n}{n}, \quad Z = |\ln(1-w)|, \quad w \in (0, 1). \quad (47)$$

Note the normalizing function is now extremely simple—the natural logarithm. Then

$$\chi(w) \equiv \langle \ln n \rangle = \frac{1}{|\ln(1-w)|} \sum_{n=1}^{\infty} \frac{w^n \ln n}{n}, \quad w \in (0, 1), \quad (48)$$

while

$$\mu(w) \equiv \langle n \rangle = \frac{1}{|\ln(1-w)|} \sum_{n=1}^{\infty} w^n = \frac{w}{(1-w)|\ln(1-w)|}. \quad (49)$$

Furthermore, at maximum entropy

$$\hat{S}(w) = S(w) = - \sum_{n=1}^{\infty} p_n \ln p_n = \ln |\ln(1-w)| + \chi(w) - \mu(w) \ln w. \quad (50)$$

This is a one-parameter model with a geometrical cutoff, which for  $w \lesssim 1$  (or more precisely  $w \rightarrow 1^-$ ), approximates the naive un-normalizable Zipf law with arbitrary accuracy.

## 7. Very general Gibbs-like model

Let us now consider an arbitrary number of constraints of the form

$$\langle \ln g_i(n) \rangle \equiv \sum_n p_n \ln g_i(n) = \chi_i, \quad i \in (1, \#_g) \quad (51)$$

and

$$\langle f_a(n) \rangle \equiv \sum_n p_n f_a(n) = \mu_a, \quad a \in (1, \#_f). \quad (52)$$

One could always transform a  $g$ -type constraint into an  $f$ -type constraint or vice versa, but as we shall soon see there are advantages to keeping the logarithm explicit. Applying the maximum entropy principle amounts to considering

$$\begin{aligned} \hat{S} = & - \sum_i z_i \left( \sum_n p_n \ln g_i(n) - \chi \right) - \sum_a \beta_a \left( \sum_n p_n f_a(n) - \mu \right) \\ & - (\ln Z - 1) \left( \sum_n p_n - 1 \right) - \sum_n p_n \ln p_n, \end{aligned} \quad (53)$$

where with malice aforethought we have now relabelled the Lagrange multipliers for the  $f$  constraints as follows:  $\ln w \rightarrow -\beta$ . Then maximizing over the  $p_n$  we have the extremality condition

$$- \sum_i z_i \ln g_i(n) - \sum_a \beta_a f_a(n) - \ln Z - \ln p_n = 0 \quad (54)$$

with explicit solution

$$p_n = \frac{1}{Z} \left\{ \prod_i g_i(n)^{-z_i} \right\} \exp \left\{ - \sum_a \beta_a f_a(n) \right\}, \quad (55)$$

where now the normalizing constant is

$$Z(\vec{z}, \vec{\beta}) = \sum_n \left[ \left\{ \prod_i g_i(n)^{-z_i} \right\} \exp \left\{ - \sum_a \beta_a f_a(n) \right\} \right]. \quad (56)$$

This can be viewed as a generalization/modification of the Gibbs distribution where we have explicitly pulled out some of the constraints (the  $g$ -type constraints) to make them look power-law-like, while the remaining constraints (the  $f$ -type constraints) are left in Boltzmann-like form. Then

$$\chi_i(\vec{z}, \vec{\beta}) \equiv \langle \ln g_i(n) \rangle = \sum_n p_n \ln g_i(n) = - \frac{d \ln Z}{dz_i}, \quad (57)$$

while

$$\mu_a(\vec{z}, \vec{\beta}) \equiv \langle f_a(n) \rangle = \sum_n p_n f_a(n) = - \frac{d \ln Z}{d\beta_a}. \quad (58)$$

Furthermore, at maximum entropy

$$\hat{S}(\vec{z}, \vec{\beta}) = S(\vec{z}, \vec{\beta}) = - \sum_n p_n \ln p_n = \ln Z(\vec{z}, \vec{\beta}) + \sum_i z_i \chi_i(\vec{z}, \vec{\beta}) + \sum_a \beta_a \mu_a(\vec{z}, \vec{\beta}). \quad (59)$$

It is only once one specifies particular choices for the functions  $f_a$  and  $g_i$  that the model becomes concrete, and only at that stage might one need to focus on particular special functions of mathematical physics. The model is extremely general—the drawback is that, because it can fit almost anything, it can ‘explain’ almost anything, and so can predict almost nothing.

## 8. Non-Shannon entropies

Shannon’s entropy is by far the best motivated of the entropy functions infesting the literature. Without making any particular commitment to the *advisability* of doing so, we can certainly ask what happens if we apply maximum entropy ideas to non-Shannon entropies (such as the Rényi [22] or Tsallis [23] entropies and their generalizations). Let us define an entropic zeta function by

$$\zeta_S(s) = \sum_n (p_n)^s, \quad (60)$$

which certainly converges for  $s \geq 1$  and *may* converge on a wider region. Then

$$S_{\text{Rényi}}(1+a) = - \frac{\ln \zeta_S(1+a)}{a}, \quad S_{\text{Tsallis}}(1+a) = \frac{1 - \zeta_S(1+a)}{a}, \quad (61)$$

where in both cases the Shannon entropy is recovered in the limit  $a \rightarrow 0$ . More generally let us consider a generalized entropy of the form

$$S(a) = -f(\zeta_S(1+a)) \quad (62)$$

for an arbitrary smooth function  $f(\cdot)$ . Let us further impose a constraint on the  $b$ th moment

$$\langle n^b \rangle \equiv \sum_n p_n n^b = \mu_b. \quad (63)$$

With power laws being explicitly built in as input into both the generalized entropy and the constraint, it is perhaps not too surprising that we will manage to get power laws dropping out. One is now interested in maximizing

$$\hat{S} = \lambda \left( \sum_n p_n n^b - \mu_b \right) + w \left( \sum_n p_n - 1 \right) - f(\zeta_S(1+a)). \quad (64)$$

Varying the  $p_n$  leads to

$$\lambda n^b + w - f'(\zeta_S(1+a)) (1+a)(p_n)^a = 0 \quad (65)$$

with solution

$$p_n = \frac{(w + \lambda n^b)^{1/a}}{Z}, \quad Z = \sum_n (w + \lambda n^b)^{1/a}. \quad (66)$$

(An overall factor of  $(1+a)f'(\zeta_S(1+a))$  simply drops out of the calculation.) If the number of states is finite, then we cannot *a priori* discard the parameter  $w$ , and we have derived a distorted power law. (The probability distribution then interpolates between a pure power law for  $w = 0$  and a uniform distribution for  $w = \infty$ .) If on the other hand, the number of states is infinite then normalizability enforces  $w \rightarrow 0$ , and  $\lambda$  factors out. We then have a pure power law

$$p_n \rightarrow \frac{n^{b/a}}{Z}, \quad Z \rightarrow \sum_n n^{b/a}. \quad (67)$$

If the state space is the positive integers  $n \in \{1, 2, 3, \dots\}$  then  $Z \rightarrow \zeta(-b/a)$ , the Riemann zeta function, and the sum converges only for  $-b/a > 1$ . So one of the two parameters  $(a, b)$  must be negative. In this situation

$$\mu_b = \langle n^b \rangle = \frac{\zeta(-b-b/a)}{\zeta(-b/a)} \quad (68)$$

now requiring both  $-b-b/a > 1$  and  $-b/a > 1$ , while at maximum entropy

$$\hat{S} = S \rightarrow f(\zeta_S(1+a)) = f\left(\frac{\zeta(-b-b/a)}{\zeta(-b/a)^{1+a}}\right). \quad (69)$$

So yes, one can also extract power laws from maximum entropy applied to non-Shannon entropies (in particular, the generalized Rényi–Tsallis entropies), but the process is (at best) rather clumsy, and seems an exercise in overkill. Apart from the whole question of whether or not non-Shannon entropies are particularly interesting, one should ask whether the result is particularly useful? This derivation does not seem to be in any way an improvement over the simpler one based directly on the Shannon entropy, so its utility is dubious.

## 9. Summary and discussion

The main point of this paper is that power laws (and their variants, including hybrid geometric/power laws) have a very natural and straightforward interpretation in terms of the maximum entropy formalism pioneered by Jaynes. The key to obtaining a pure power law in the simplest possible manner lies in maximizing the Shannon entropy while imposing the simple constraint  $\langle \ln n \rangle = \chi$ . Depending on other features of the specific model under consideration, detailed analysis leads to certain of the special functions of mathematics (the Riemann zeta

function, generalized harmonic functions, poly-logarithms or even ordinary logarithms), but these are relatively well-known mathematical objects, which are still tolerably simple to deal with.

Adding additional features (finite size state space, extra constraints) can (and typically will) modify both the functional form and the normalization constants appearing in the probability distribution. As always there is a trade-off between simplicity and flexibility. A more complicated model (with more free parameters) has a more flexible probability distribution, but this comes at a real (if often unacknowledged) cost in terms of internal complexity. A rather general Gibbs-like model is laid out and briefly discussed. We also briefly discuss applying maximum entropy ideas to non-Shannon entropies (such as the Rényi and Tsallis entropies). There is very definitely a trade-off in both elegance and plausibility, and I would argue strongly that the simplest and most elegant model consists of the Shannon entropy, a constraint on  $\langle \ln n \rangle = \chi$ , and a trivial normalization constraint on the sum of probabilities.

The fact that the logarithmic average  $\langle \ln n \rangle$  plays such an important role in power laws seems to have a connection with the fact that logarithmic scales are ubiquitous in classifying various natural and social phenomena. For instance

- stellar magnitudes are logarithmic in stellar luminosity,
- earthquake magnitudes (modified Richter scale) are logarithmic in energy release,
- sound intensity decibels are logarithmic in pressure,
- the acidity/alkalinity pH scale is logarithmic in hydrogen ion concentration,
- musical octaves are logarithmic in frequency,
- war severity can be characterized as being logarithmic in casualty count [7].

In many cases the utility of a logarithmic scale can be traced back to an approximate logarithmic sensitivity in human perceptual systems, but it is very easy to confound cause and effect. After all, in the presence of power-law distributed external stimuli, there is a significant disadvantage in having the human perceptual system overwhelmed by large numbers of low-impact events, suggesting an evolutionary pressure towards suppressing sensitivity to low-impact events. *This suggests that logarithmic sensitivity in human (and animal) perceptual systems is evolutionarily preferred for those senses that are subject to an external bath of power-law distributed stimuli.*

Fortunately, for the purposes of applying maximum entropy ideas one does not need to know which is the cause and which is the effect—one is ‘merely’ using Bayesian principle to *estimate* underlying probabilities in the presence of limited knowledge; for current purposes this is most typically the single piece of information that  $\langle \ln n \rangle = \chi$ .

## Acknowledgments

MV acknowledges support via the Marsden Fund, and via a James Cook Fellowship, both administered by the Royal Society of New Zealand.

## References

- [1] Zipf G K 1935 *The Psychobiology of Language* (Cambridge, MA: Houghton-Mifflin)
- [2] Zipf G K 1949 *Human Behavior and the Principle of Least Effort* (Reading, MA: Addison-Wesley)

- [3] Saichev A, Malevergne Y and Sornette D 2009 *Theory of Zipf's Law and Beyond (Lecture Notes in Economics and Mathematical Systems* vol 632) (Berlin: Springer)
- [4] Newman M E J 2005 Power laws, Pareto distributions and Zipf's law *Contemp. Phys.* **46** 323–51
- [5] Clauset A, Shalizi C R and Newman M E J 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703
- [6] Baxter G, Freen M, Noble J, Rickerby M, Smith H, Visser M, Melton H and Tempero E 2006 Understanding the shape of Java software *OOPSLA, Proc. 21st Annual ACM SIGPLAN Conf. on Object-Oriented Programming Systems, Languages and Applications* ed P L Tarr and W R Cook (New York: ACM) pp 397–412
- [7] Fry Richardson L 1960 *Statistics of Deadly Quarrels* (Pacific Grove, CA: Boxwood Press)
- [8] Baek S K, Bernhardsson S and Minnhagen P 2011 Zipf's law unzipped *New J. Phys.* **4** 043004
- [9] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev. Ser. II* **106** 620–30
- [10] Jaynes E T 1957 Information theory and statistical mechanics: II *Phys. Rev. Ser. II* **108** 171–90
- [11] Jaynes E T 1968 Prior probabilities *IEEE Trans. Syst. Sci. Cybern.* **4** 227–41
- [12] Jaynes E T 1988 The relation of Bayesian and maximum entropy methods *Maximum-Entropy and Bayesian Methods in Science and Engineering* vol 1 ed G Erickson and C R Smith (Dordrecht: Kluwer/Academic) pp 25–9
- [13] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press) pp 351–5
- [14] Cristelli M, Batty M and Pietronero L 2012 There is more than a power law in Zipf *Sci. Rep.* **2** 812
- [15] Hernando A, Plastino A and Plastino A R 2012 MaxEnt and dynamical information *Eur. Phys. J. B* **85** 147
- [16] Hernando A and Plastino A 2012 Variational principle underlying scale invariant social systems *Eur. Phys. J. B* **85** 293
- [17] Hernando A and Plastino A 2012 The thermodynamics of urban population flows *Phys. Rev. E* **86** 066105
- [18] Hernando A, Hernando R, Plastino A and Plastino A R 2013 The workings of the maximum entropy principle in collective human behaviour *J. R. Soc. Interface* **10** 20120758
- [19] Pastor-Satorras R and Wagensberg J 1998 The maximum entropy principle and the nature of fractals *Physica A* **251** 291–302
- [20] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423
- [21] Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Champaign, IL: University of Illinois Press)
- [22] Rényi A 1960 On measures of information and entropy *Proc. 4th Berkeley Symp. on Mathematics, Statistics and Probability (1961)* pp 547–561
- [23] Tsallis C 1988 Possible generalization of Boltzmann–Gibbs statistics *J. Stat. Phys.* **52** 479–87
- [24] Hardy G H 1908 *A Course in Pure Mathematics* (Cambridge: Cambridge University Press) (reprinted 2006)
- [25] Hardy G H and Riesz M 1915 *The General Theory of Dirichlet's Series* (Cambridge: Cambridge University Press) (republished by Cornell University Library 1991)
- [26] Shilov G E 1996 *Elementary Real and Complex Analysis* (New York: Dover)
- [27] Havil J 2003 *Gamma: Exploring Euler's Constant* (Princeton, NJ: Princeton University Press)