Rényi entropy, abundance distribution and the equivalence of ensembles

Thierry Mora[1], Aleksandra M. Walczak[2]

[1] *Laboratoire de physique statistique, CNRS, UPMC and École normale supérieure, 24, rue Lhomond, Paris, France and*
[2] *Laboratoire de physique théorique, CNRS, UPMC and École normale supérieure, 24, rue Lhomond, Paris, France*

Distributions of abundances or frequencies play an important role in many fields of science, from biology to sociology, as does the Rényi entropy, which measures the diversity of a statistical ensemble. We derive a mathematical relation between the abundance distribution and the Rényi entropy, by analogy with the equivalence of ensembles in thermodynamics. The abundance distribution is mapped onto the density of states, and the Rényi entropy to the free energy. The two quantities are related in the thermodynamic limit by a Legendre transform, by virtue of the equivalence between the micro-canonical and canonical ensembles. In this limit, we show how the Rényi entropy can be constructed geometrically from rank-frequency plots. This mapping predicts that non-concave regions of the rank-frequency curve should result in kinks in the Rényi entropy as a function of its order. We illustrate our results on simple examples, and emphasize the limitations of the equivalence of ensembles when a thermodynamic limit is not well defined. Our results help choose reliable diversity measures based on the experimental accuracy of the abundance distributions in particular frequency ranges.

As an increasing number of large datasets are becoming available in a variety of fields, one often turns to reduced statistics that can capture important properties of the system, or help detect deviations from our expectations. Distributions of abundances have proven useful as such statistics, and have been used in many different contexts, from biology to linguistics, astrophysics and sociology. This notion is best explained when counting biological species from a sample. Say that species 1 was observed n_1 times, species 2 n_2 times, etc. The abundance distribution discards information about the *identity* of the sampled species, and focuses on the distribution of the counts themselves n_1, n_2, etc. This notion is very general and extends well beyond ecology. Counts can refer to the number of times a word is used in a text, to the number of people living in a given city, to the occurence of specific spiking patterns in a population of neurons, or to the abundance of specific lymphocyte clones in the immune system, to give but a few examples.

An equivalent way of representing the abundance distribution is to order the counts from largest to smallest, and plot them as a function of their rank in this ordering. For example, in the English language, one can order words by their frequency of occurence, and study how this frequency decreases with the rank. In 1949 Zipf observed that this dependency roughly followed a power law [1], and similar observations have later been made in a variety of contexts [2]. Because of the ubiquity of these power laws [3], frequency-versus-rank plots are commonly represented on a double logarithmic scale.

Abundance distributions can contain useful, albeit indirect, information about the underlying process at work in the system. In ecology, they are used as a diagnostic tool for detecting deviations from the prediction of a neutral model of population dynamics [4, 5]. The Yule speciation process [6, 7], called the preferential attachment process in the context of networks [8], also predicts a specific form for the abundance distribution, which is consistent with Zipf's law in some limit [7]. The abundance distribution of spike patterns in the retina has been used to study the critical properties of the underlying neural network [9], and a similar analysis was performed on small patches of natural images [10]. The distribution of sizes of lymphocyte clones in the immune system also seems to generically follow power-laws, which puts constraints on the rules of their population dynamics [11].

Abundance distributions are closely related to the notion of diversity. Diversity can be defined in a number of ways: total number of types in the distribution, Shannon's entropy [12], Simpson's diversity index, *etc.* It has long been realized [13] that these different kinds of diversity can all be brought under the common definition of the Rényi entropy [14]. This quantity, which depends on a single parameter called order, generalizes Shannon's and Gibbs' entropy. It is commonly used in ecology to quantify diversity, but has also received increasing attention in condensed matter in the context of quantum entanglement [15].

Here we show how the Rényi entropy can be geometrically constructed from the abundance distribution in the thermodynamic limit. This construction allows one to visualize graphically how different measures of diversity arise from a given abundance distribution. It indicates which are the abundances that are determinant in each diversity measure, and gives a visual assessment of when to trust the estimate of these measures.

Our result relies on the framework of statistical mechanics, piecing together previous observations. The equivalence between rank-frequency curves and the micro-canonical entropy has been previously reported in [16]. The link between micro-canonical entropy and free energy is a classical result of statistical mechanics, known as the equivalence of ensembles [17]. The mapping between free energy and the Rényi entropy has been pointed out recently [18]. By bringing these results in a common framework, we hope to clarify the correspondance between abundance distributions, the density of states and Rényi entropies, and propose a straightforward geomet-

ric method for assessing diversity directly from the abundance distribution represented in an appropriate manner.

Rényi entropy and free energy

Let us define a probability distribution $p(s)$, where s is a state or a type that can take a discrete number of values. $p(s)$ is a relative abundance, or a frequency, so that $\sum_s p(s) = 1$. The variable s can be a spin configuration of a large system, a species, a biological sequence, or a spiking pattern from a population of neurons.

The Rényi entropy of order β is defined as:

$$H(\beta) = \frac{1}{1-\beta} \ln \left[\sum_s p(s)^\beta \right]. \tag{1}$$

This quantity generalizes Shannon's entropy,

$$H_1 = -\sum_s p(s) \ln p(s), \tag{2}$$

to which it reduces in the limit $\beta \to 1$. The Rényi entropy is associated with a family of diversity indices, defined as:

$$D(\beta) = \exp[H(\beta)] = \left(\sum_s p(s)^\beta \right)^{\frac{1}{1-\beta}}. \tag{3}$$

This quantity can be interpreted as an effective number of states. When $\beta = 0$, $D(0)$ is just the raw, total number of possible types in the system. When $\beta = 2$, it is equal to the inverse of Simpson's index, also interpreted as an effective number of types, and related to the Gini-Simpson index (defined as $1 - 1/D(2)$, commonly used to measure inequalities. When $\beta = 1$, $D(1)$ is the exponential of Shannon's entropy, and is sometimes called the true diversity. In Shannon's original work [12], $D(1)$ is the effective number of codewords needed to compress s.

We first derive an equivalence between the Rényi entropy and the free energy of statistical mechanics, as already reported in [18]. We formally rewrite the probability distribution $p(s)$ as a Boltzmann distribution:

$$p(s) = \frac{1}{Z_1} e^{-E(s)}, \tag{4}$$

where the temperature is set to $k_B T = 1$ by definition. For example, this mapping can be realized by defining $E(s) \equiv -\ln P(s)$ and $Z_1 = 1$, but to keep things general we will assume an arbitrary Z_1. We define the free energy at unit temperature as $F_1 = -\ln Z_1$. The Rényi entropy can be rewritten as:

$$H(\beta) = \frac{1}{1-\beta} \ln \left[\sum_s e^{-\beta E(s) + \beta F_1} \right] = \frac{\beta(F_1 - F(\beta))}{1-\beta}, \tag{5}$$

In this formula, $F(\beta)$ is the usual free energy at inverse temperature β:

$$F(\beta) \equiv -\frac{1}{\beta} \ln Z(\beta), \tag{6}$$

where $Z(\beta)$, the normalization factor of the Boltzmann distribution $p_\beta(s) = \exp(-\beta E)/Z(\beta)$ at inverse temperature β,

$$Z(\beta) \equiv \left[\sum_s e^{-\beta E(s)} \right], \tag{7}$$

is called the partition function.

Thus the Rényi entropy is closely related to the free energy, after mapping to the Boltzmann distribution. Note that this mapping is a definition, and does not follow from physical considerations.

Abundance distribution and micro-canonical entropy

There also exists a rigorous analogy between the density of states and the abundance distribution [16]. The abundance distribution is defined as the distribution over p itself:

$$\rho(p) = \sum_s \delta[p(s) - p], \tag{8}$$

where $\delta(\cdot)$ is Dirac's delta function. It is more convenient to work with the cumulative density of p, as it is not plagued with Dirac deltas, and is invariant under reparametrization:

$$C_p(p) = \sum_s \Theta[p(s) - p], \tag{9}$$

where $\Theta(x)$ is the Heaviside function, equal to 1 for $x \geq 0$ and 0 otherwise.

The cumulative distribution of abundances is related to another representation of diversity, the rank-frequency curve or Zipf's plot [1]. In this representation, the system's states are ranked from most abundant to least abundant, and their abundance shown as a descreasing function of the rank. The rank of a given abundance p is given exactly by $C_p(p)$. Hence, rank-frequency graphs are simply plots of p versus $C_p(p)$. In other words, they represent the inverse function of the cumulative abundance distribution.

Since p and E are related by the Boltzmann distribution (4), we can equivalently define the cumulative density of states, which counts all states under a given temperature E.

$$C_E(E) = \sum_s \Theta[E - E(s)]. \tag{10}$$

This cumulative distribution is related to the cumulative distribution of p through $C_E(E) = C_p(e^{-E}/Z_1)$. The usual density of states,

$$\rho(E) = \sum_s \delta(E - E(s)), \tag{11}$$

is obtained as $dC_E(E)/dE$. In order to avoid issues of definition with Dirac delta functions, we define a cumulative micro-canonical entropy as $S(E) = \ln C_E(E)$, rather than the usual micro-canonical entropy. In this definition, for ease of notation we implictly take the Boltzmann constant to be $k_B = 1$.

Equivalence of ensembles

Following textbook statistical mechanics, the partition function (7) can be rewritten entirely as a function of the density of states:

$$Z(\beta) = \int dE \rho(E) e^{-\beta E} = \beta \int dE C_E(E) e^{-\beta E}$$
$$= \beta \int dE e^{S(E) - \beta E}, \qquad (12)$$

where we have used integration by parts in the second equality. In other words, $Z(\beta)$ is the Laplace transform of the density of states.

In the standard thermodynamic limit, where both the entropy and energy are assumed to be extensive, $S(E) \sim E \sim N$, where N is the system's size, this integral can be approximated by its saddle point, an approximation also known as Laplace's method:

$$Z(\beta) \approx \beta \left(\frac{\pi}{|S''(E^*)|} \right)^{1/2} e^{S(E^*) - \beta E^*}, \qquad (13)$$

where E^*, which maximizes the term in the exponential in Eq. 12, is given by the standard thermodynamic relation:

$$\left. \frac{dS}{dE} \right|_{E*} = \beta, \qquad (14)$$

or more classically $dS/dE = 1/T$. The free energy (Eq. 6) then reads:

$$F(\beta) = E^* - \frac{1}{\beta} S(E^*) - \frac{\ln(\beta)}{\beta} - \frac{\ln(\pi)}{2\beta} + \frac{\ln(|S''(E^*)|)}{2\beta}. \qquad (15)$$

In the thermodynamic limit the last three terms are subextensive (scaling sublinearly with the system's size) and therefore dropped, reducing to the usual definition of the free energy, $F = E - TS$. Then the Massieu potential (also called the Helmholtz free entropy) $\Phi(\beta) = -\beta F(\beta)$ and the micro-canonical entropy $S(E)$ are related by a Legendre transform:

$$\Phi(\beta) = \text{extr}_E[S(E) - \beta E], \qquad (16)$$
$$S(E) = \text{extr}_\beta[\Phi(\beta) - \beta E], \qquad (17)$$

in which E and β are conjugate variables. These relations define the equivalence between the micro-canonical and canonical ensembles, which is valid as long as $S(E)$ is a concave function [17]. In this equivalence, different inverse temperatures β are used to sample states of different typical energies, acting as a large-deviation parameter. These relations formally follow from the Boltzmann distribution in the thermodynamic limit, and are the same as in standard thermodynamics.

The saddle-point approximation is more than a computational trick. It also impies that, in the thermodynamic limit, the measure is dominated by just a few states that all have pratically the same energy E^*. There are of the order of $\exp[S(E^*)]$ such states, which each have roughly the same probability $\exp(-\beta E^*)/Z(\beta) = \exp[-S(E^*)]$. Their entropy is then given by Boltzmann's formula:

$$H_1[p_\beta] = -\sum_s p_\beta(s) \ln p_\beta(s) \approx \ln[e^{S(E^*)}] = S(E^*), \qquad (18)$$

where $H_1[p_\beta]$ is the canonical entropy at inverse temperature β, not to be confused with the Rényi entropy $H(\beta)$. The result of Eq. 18 can be shown more rigorously by using the exact identity:

$$H_1[p_\beta] = \beta[\langle E \rangle_\beta - F(\beta)], \qquad (19)$$

with $\langle x \rangle_\beta = \sum_s p_\beta(s) x(s)$, and by showing $\langle E \rangle_\beta \approx E^*$ using Laplace's method.

Legendre construction

The Legendre transform (16) can be constructed geometrically, as illustrated by Fig. 1. In this construction, $F(\beta)$ is obtained as the intercept of the tangent to $S(E)$ of slope β (dashed line in Fig. 1) with the abscissa. To see this, we write the condition for E^* at the point where the tangent of slope β touches the $S(E)$ curve, $dS/dE = \beta$, which is exactly Eq. 14. The equation defining the tangent is then given in (E, S) space by:

$$S = S(E^*) + \beta(E - E^*). \qquad (20)$$

Solving in E for the intercept with the abscissa, $S = 0$, gives $E^* - S(E^*)/\beta = F(\beta)$, which is the result of Eq. 15 up to the sub-extensive terms.

We can generalize this construction to the Rényi entropy, which is obtained as the intersection of two tangents to $S(E)$, of slopes 1 and β respectively. To verify this assertion, one writes the system of two linear equations parametrizing these two tangents in the (S, E) space:

$$S = E - F_1, \qquad (21)$$
$$S = \beta[E - F(\beta)]. \qquad (22)$$

The solution to these two equations in S is $\beta(F_1 - F(\beta)/(1 - \beta)$, which is exactly the Rényi entropy $H(\beta)$ according to (5).

As already mentioned, the Rényi entropy reduces to the classical Shannon or Gibbs entropy, H_1, for $\beta = 1$:

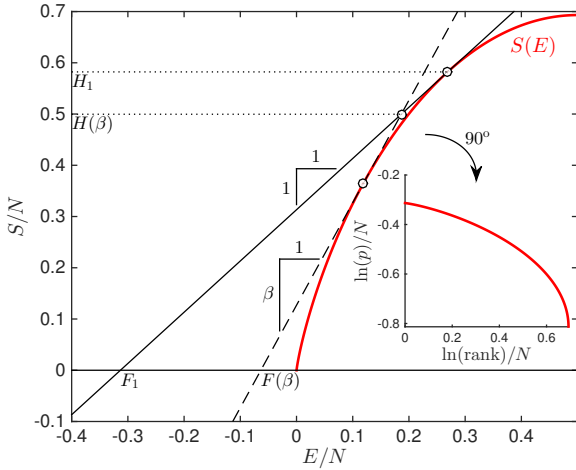$$\lim_{\beta \to 1} H(\beta) = \left. \frac{dF}{d\beta} \right|_{\beta=1} = H_1. \qquad (23)$$

FIG. 1: Geometric construction of the Rényi entropy from the density of states. In the classical Legendre construction, the free energy $F(\beta)$ is obtained as the intersection of the tangent to the micro-canonical entropy curve $S(E)$ (in red) of slope β, where β is the inverse temperature, and the abscissa. The Rényi entropy of order β, $H(\beta)$, is obtained as the intersection between the tangents of slope 1 and β, projected onto the ordinate. Inset: the micro-canonical entropy curve is equivalent, up to a 90° rotation, to the rank-frequency curve represented on a logarithmic scale.

This limit can also be undertood geometrically. When $\beta \to 1$, the intersection between the two tangents tends to the point of tangency of slope 1, $dS/dE|_{E^*} = 1$, where $S(E^*)$ is equal to the Shannon entropy H_1 (Eq. 18).

From the abundance distribution to Rényi entropy: a geometric approach

Now that we have derived analogous relations to standard thermodynamics, we can use the geometric representation of the Legendre transform to read off diversity measures from data. The Legendre construction can be transposed into the language of the abundance distribution, provided that this distribution is appropriately represented as a rank-frequency curve. Recall that $S(E) = \ln C_E(E)$, where C_E is the rank of states of energy E, ordered from the lowest to the highest energy, *i.e.* from the most frequent to the least frequent state. On the other hand, $E = -\ln p + F_1$, where p is the frequency. Thus, the micro-canonical entropy function, S *vs.* E, and the rank-frequency relation in logarithmic scale, $\ln(p)$ *vs.* $\ln(\text{rank})$, are exactly equivalent up to a 90° rotation, as illustrated in the inset of Fig. 1.

Thanks to this equivalence, the Legendre construction described above can be applied directly to the rank-frequency curve plotted on a log-log scale. We illustrate such a consctruction with the distribution of generation probabilities of T-cell receptor beta chains [19]. DNA sequences s coding for the beta chain of T-cell recep-

tors, which are involved in recognizing pathogens, are generated according to the probability distribution $p(s)$, which was inferred from the data. For each generated sequence, its probability of generation p, and thus its energy $E = -\ln p$, is also output by the model. This allows us to compute empirically the probability distribution of E under the model, which is proportional to the number of states with a certain energy multiplied by their probability $p = e^{-E}$, $P(E) \propto \rho(E)e^{-E}$, from which $\rho(E)$ and then $C_E(E) = \int_0^E dE' \rho(E')$ are obtained. Note that this distribution is synthetically created from the model of generation by drawing random, independent sequences. It is distinct from clone-size distributions usually found in the literature [11, 20, 21], which have a clonal structure and are not made of independent samples. Also note that this ensemble has no natural thermodynamic limit, because sequences have a finite length. It makes for a good test case for our method in a real-world example.

The rank-frequency plot is represented in Fig. 2 in a double logarithmic scale. Following the previous arguments, in this representation the diversity index $D(\beta) = e^{H(\beta)}$ can be approximated by:

1. drawing the tangent of slope -1 (black solid line) to the rank-frequency curve;

2. drawing the tangent of slope $-\beta^{-1}$ (dashed line) to the same curve;

3. projecting the intersection point between these two lines onto the rank axis.

The tangency point of the tangent of slope 1 gives the true diversity index $D(1) = \prod_s p(s)^{-p(s)}$, *i.e.* the exponential of Shannon's entropy. In Fig. 2 we illustrate the example of $\beta = 2$, which allows us to read of Simpson's inverse index $D(2) = 1/\sum_s p(s)^2$. The exact values for these two quantities, $D(1) = 4.9 \cdot 10^{13}$ and $D(2) = 3.4 \cdot 10^9$, are roughly approximated, although underestimated, by the construction.

In the true thermodynamic limit, which is not strictly realized but approached in this example, the diversity measure $D(\beta)$ is effectively dominated by just a fraction of sequences whose rank is close to $D(\beta)$ (on a logarithmic scale), according to Laplace's approximation. A consequence of this concentration is that different diversity measures, such as Shannon's entropy or Simpson's index, may in fact be determined by entirely distinct sequences.

The construction allows for a quick assessment of whether the sampling depth can support the estimation of the Rényi entropy $H(\beta)$, and its associated diversity $D(\beta)$, for a given index β. When the tangent of slope $-\beta^{-1}$ touches the curve towards its end, where states are becoming rare and may have been observed only once, it is probably safe to assume that the Rényi entropy cannot be reliably computed from the data, because it is determined by states which have not been sampled well. This limitation applies to the Legendre construction as well as to any other estimate of Renyi's entropy. Such
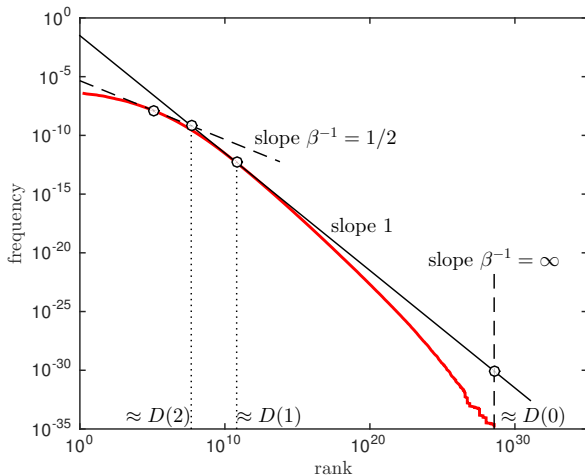
FIG. 2: Illustration of the Legendre construction of the Rényi entropy on the rank-frequency curve of randomly generated T-cell receptor beta chains [19]. The construction is identical to that of Fig. 1, with slope β replaced by slope $-\beta^{-1}$ because of the rotation. The projection onto the rank axis gives an approximation to the diversity index of order β, $D(\beta) = \exp[H(\beta)]$.

a diagnosis indicates which diversity measure might be appropriate to use in a given context, depending on the shape of the rank-frequency curve.

An extreme case is $\beta = 0$: the tangent of slope $-\beta^{-1} = -\infty$ intersects with the rank-frequency curve at the maximal possible rank, which is also the total number of sampled types. In most cases (as in this one) this maximal rank does not represent well the true total diversity, $D(0)$, which should also include unseen types. A similar underestimation is expected to happen for finite values of β for which the tangent is ill-defined.

Singular cases

It is interesting to consider what happens to the Rényi entropy when the rank-frequency relation is locally a power law. In the micro-canonical framework, a power law in the cumulative density of abundances [16],

$$C(p) \propto \frac{1}{p^{\alpha}}, \qquad (24)$$

translates into a linear density of states, $S(E) = S_0 + \alpha(E - E_0)$. This behaviour, as long as it spans an extensive range of energies, leads to a discontinuity in the derivative of $F(\beta)$ at $\beta = \alpha$. For $\alpha \neq 1$, Eq. 5 implies that the derivative of $H(\beta)$ exhibits the same kind of discontinuity at $\beta = \alpha$.

For $\alpha = 1$ the discontinuity is of a different nature. Eq. 5 can be expanded around $\beta = 1$ as:

$$H(\beta) \approx H_1 + \frac{\beta - 1}{2} \left. \frac{d^2 F}{d\beta^2} \right|_{\beta=1}, \qquad (25)$$

hence:

$$\left. \frac{dH}{d\beta} \right|_{\beta=1} = \frac{1}{2} \left. \frac{d^2 F}{d\beta^2} \right|_{\beta=1}. \qquad (26)$$

Therefore, the discontinuity in $dF(\beta)/d\beta$ at $\beta = 1$ translates into a discontinuity in $H(\beta)$ itself. Again, this discontinuity can be seen geometrically. Let us assume that $S(E) = S_0 + (E - S_0)$ over a range (E_1, E_2). The tangent of slope 1 coincides with $S(E)$ throughout this range. As a result, the intersection between the tangent of slope β jumps from E_1 to E_2 as β crosses 1, causing $H(\beta)$ to jump from $S(E_1)$ to $S(E_2)$.

This kind of singularity not only implies discontinuities in the Rényi entropy or its derivatives, but also suggest that the entropy may ill-defined or hard to estimate when $\alpha = 1$. In that case, a whole range of (S, E) pairs, instead of a single point, are candidates for the tangency point between the line of slope 1 and the micro-canonical entropy $S(E)$. The entropy is ultimately determined by corrections that are ignored in the thermodynamic limit.

The micro-canonical entropy need not be strictly linear over a portion of energies for a discontuinuity to occur. In fact, any convexity in $S(E)$ is predicted to produce the same effect [17].

For this reason, caution should be used when dealing with distributions that look like a power law, or are not concave in logarithmic scale. Not only may the Legendre construction be unreliable, but so may other more direct estimates of the Rényi entropy, since the system lacks a characteristic energy scale. Interestingly, several abundance distributions in biology have been reported to follow power laws with exponent $\alpha = 1$ [9, 10, 16], for which the Rényi entropy is expected to have a discontinuity.

Discussion

In this paper we have made an explicit link between classical representations of diversity in ecology and other fields, and the framework of classical statistical mechanics. This mapping allows one to bring many quantities coming under many different names – species abundance distribution, clone-size distribution, frequency spectrum, Shannon entropy, Rényi entropy, Simpson's index, *etc.* – within a common framework. It provides an quick an easy way to simply read off diversity directly from rank-frequency plots.

Our geometric construction assumes the thermodynamic limit, which may not be satisfied or even well defined. For instance, the distribution of abundances prediced by a neutral model, or Fisher's log-series $P(n) \propto \alpha^n/n$ [22], does not admit a natural definition of system's size, and thus has no well defined thermodynamic limit. The same goes for Pareto distributions $P(n) \propto n^{-a}$. In these cases where no thermodynamic limit exists, the Legendre construction is no substitute for a direct estimate, but may still give a reasonable guess. It can also

hint whether such a direct estimate is possible at all, by identifying the range of frequencies or abundances that are expected to dominate the diversity measure.

Depending on how well sampled the distribution is, different orders β of the Rényi entropy may be appropriate. The proposed framework can aid in choosing the right measure depending on the data. In general, the less well sampled the data is, the higher the order should be chosen. For instance, Simpson's index is less sensitive to poor sampling than the Shannon entropy, which itself is easier to estimate from the data than the total number of states. Ultimately, the particular form of the abundance distribution should be examined to decide which measure can or should be used.

[1] Zipf GK (1949) *Human behavior and the principle of least effort* (Addison-Wesley Press).

[2] Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* 46:323–351.

[3] Schwab DJ, Nemenman I, Mehta P (2014) Zipf's law and criticality in multivariate data without fine-tuning. *Phys. Rev. Lett.* 113:068102.

[4] Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037.

[5] Chisholm RA, Pacala SW (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc. Natl. Acad. Sci.* 107:15821–15825.

[6] Yule G (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series BContaining Papers of a Biological Character* 213:21–87.

[7] Simon H (1955) On a class of skew distribution functions. *Biometrika* 42:425–440.

[8] Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science (80-.).* 286:11.

[9] Tkačik G, et al. (2015) Thermodynamics and signatures of criticality in a network of neurons. *Proc. Natl. Acad. Sci.* 112:11508–13.

[10] Stephens GJ, Mora T, Bialek W, Tkačik G, Bialek W (2013) Thermodynamics of natural images. *Phys. Rev. Lett.* 110:018701.

[11] Desponds J, Mora T, Aleksandra W (2016) Fluctuating fitness shapes the clone size distribution of immune repertoires. *Proc Natl Acad Sci USA* 113:274.

[12] Shannon CE (1948) A mathematical Theory of Communication. *Bell Sys Tech J* 27:379–423 & 623–656.

[13] Hill AMO (1973) Diversity and Evenness : A Unifying Notation and Its Consequences. *Ecology* 54:427–432.

[14] Rényi A (1961) On measures of entropy and information. *Entropy* 547:547–561.

[15] Hastings MB, González I, Kallin AB, Melko RG (2010) Measuring Renyi entanglement entropy in quantum Monte Carlo simulations. *Phys. Rev. Lett.* 104:2–5.

[16] Mora T, Bialek W (2011) Are Biological Systems Poised at Criticality? *J. Stat. Phys.* 144:268–302.

[17] Touchette H, Ellis RS, Turkington B (2004) An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A* 340:138–146.

[18] Baez JC (2011) Renyi Entropy and Free Energy. *arXiv* p 1102.2098.

[19] Murugan A, Mora T, Walczak AM, Callan CG (2012) Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc. Natl. Acad. Sci.* 109:16161–16166.

[20] Mora T, Walczak AM, Bialek W, Callan CG (2010) Maximum entropy models for antibody diversity. *Proc. Natl. Acad. Sci. U. S. A.* 107:5405–5410.

[21] Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front. Immunol.* 4:485.

[22] Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12:42–58.