



# Maximum entropy, logistic regression, and species abundance

Fangliang He

*F. He (fhe@ualberta.ca), Dept of Renewable Resources, Univ. of Alberta, Edmonton, T6G 2H1, Canada.*

There is considerable debate about the utility of statistical mechanics in predicting diversity patterns in terms of life history traits. Here, I reflect on this debate and show that a community is controlled by the balance of two opposite forces: the entropic part (the natural tendency of the system to be in the configuration with the highest possible entropy) and environmental, ecological and evolutionary constraints maintaining order (reducing entropy). The Boltzmann distribution law that can be derived from the maximum entropy formalism provides a fundamental model for linking species abundance to life history traits and environmental constraining factors. This model predicts a global pattern of diversity evenness along a latitudinal gradient. Although the Boltzmann distribution and the logistic regression models represent two fundamentally different approaches, the two models have an identical mathematical form. Their identical formalisms facilitate the interpretation of logistic regression models with statistical mechanics, and reveal several limitations of the maximum entropy formalism. I argued that although maximum entropy formalism is a promising tool for modeling species abundances and for linking microscopic quantities of individual life history traits to macroscopic patterns of diversity, it is necessary to revise the Boltzmann distribution law for successful prediction of species abundance.

Why are the species in a community not equally abundant? Why are many species rare and a few species abundant? What makes the sharing of niche space so unfair? Is this pattern driven by the luck of dice rolling or by environmental, ecological and evolutionary forces? The search for an answer to these questions has been the major undertaking and challenge for community ecologists. The recent application of statistical mechanics to these questions has attracted much attention (McGill 2006, Shipley et al. 2006, Banavar and Maritan 2007, Pueyo et al. 2007, Whitfield 2007, Dewar and Porté 2008, Haegeman and Loreau 2008, Harte et al. 2008, Shipley 2008). Using a major statistical mechanical tool, called maximum entropy (MaxEnt) formalism, Shipley et al. (2006) have shown that the abundances of 30 herbaceous species in a vineyard chronosequence in southern France can be reasonably predicted by eight life history traits of perenniality, leaf characters, seed production and stem mass, etc. The proposal to use life history traits to study the properties of ecological communities is not new but a current major research focus in ecology (McGill et al. 2006). The use of MaxEnt to link microscopic quantities of individual life history traits to macroscopic patterns of diversity is innovative. Debate has, however, arisen. The research is criticized to be circular, non-parsimonious and even useless (Marks and Muller-Landau 2007, Roxburgh and Mokany 2007, Shipley et al. 2007, 2008, Haegeman and Loreau 2008). Here I reflect on this debate and discuss the promise and limitations of MaxEnt formalism for understanding the organization of community assemblages from the perspective of statistical mechanics.

At present, there are two major lines of application of the MaxEnt formalism to ecology. The first one is to infer abundances of individual species in terms of traits or other biotic and environmental characteristics, as represented by Alexeyev and Levich (1997), Levich (2000) and Shipley et al. (2006). The second is to infer macroecological patterns (e.g. relative species–abundance distributions, species–area curves, biodiversity–productivity relationships, etc.), as represented by Banavar and Maritan (2007), Pueyo et al. (2007), Dewar and Porté (2008) and Harte et al. (2008). The differences of these two applications have been nicely explained in Appendix A of Pueyo et al. (2007). The focus of this paper is on the first application, i.e. to infer abundances of individual species in terms of life history traits or any other predictive variables.

## The inference of maximum entropy formalism

Thermodynamics has two fundamental laws. The First law is simply the principle of energy conservation, which is a book-keeping law. The Second law predicts that isolated systems always tend toward disorder – the entropy of a closed system will never diminish as it evolves. Here entropy is a measure of disorder (or uncertainty): high entropy means high disorder and vice versa. ‘Disorder’ is a sloppy but useful metaphor to understand the Second law. As Gibbs (the collected work of J. Willard Gibbs, Yale Univ. Press, New Haven 1948) and Jaynes (1957) showed, the Second law can be derived by maximizing the information entropy associated with a given system, compatible with the constraints the system is

subjected to. What Jaynes (1957) discovered is that the Max-Ent principle was powerful enough to be used as a statistical tool in its own respect, independent of thermodynamics. In most real systems, the information entropy is not free to be at its maximum unconstrained value. Instead, it is checked by many constraints (e.g. conservation of energy in thermodynamics or the variation in life history traits or environmental variables). The state of a system is thus determined by the balance of these two opposite forces: the entropic part (the natural tendency of the system to be in the configuration with the highest possible entropy) and the constraints that the system has to obey (Fig. 1). The operation of these two forces can be mathematically demonstrated, and both are taken into account in the MaxEnt formalism. As a statistical tool, what MaxEnt does is simply to reconstruct the most probable underlying probability distribution when only some average values are known.

To illustrate, let's consider a hypothetical community of six species. The multiplicity of the system is the total number of ways the community can be partitioned into a particular set  $\{n_1, n_2, n_3, n_4, n_5, n_6\}$ , where  $n_i$  is the abundance of the  $i$ th species and  $N = \sum n_i$ , given by

$$W = \frac{N!}{n_1!n_2!n_3!n_4!n_5!n_6!} \quad (1)$$

For example, if the community has  $N=12$ , the number of ways to partition the community into the composition  $\{0, 1, 1, 1, 1, 8\}$  is  $W = \frac{12!}{0!1!1!1!1!8!} = 11\,880$ , which is much

smaller than  $W=7\,484\,400$  for the even partition  $\{2, 2, 2, 2, 2\}$ . This means the latter partition is much more plausible than the former with the standard assumption of the equiprobability of microstates (Dill and Bromberg 2002). Therefore, for a community free of any types of constraints, the most plausible expectation should be equal abundance. The inclusion of any constraints will break down the equability and lead to the construction of the most conservative, non-committal distribution consistent

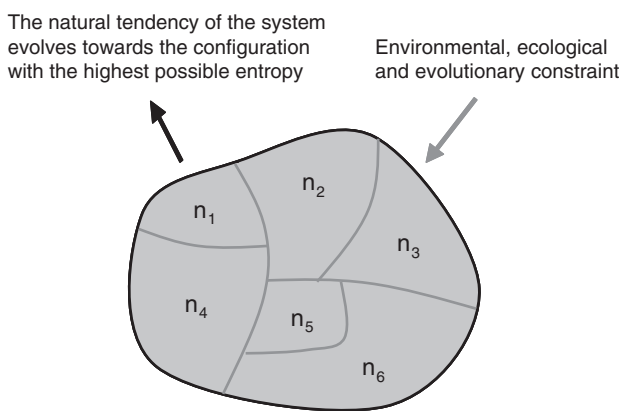


Figure 1. Hypothetical community consists of six species with abundances:  $n_1, \dots, n_6$ . The assemblage of the community is determined by the balance of the tendency of the system to be in the configuration with the highest possible entropy that drives the system toward disorder and environmental, ecological and evolutionary constraints that reduce entropy and maintain order.

with data (Jaynes 2003). This non-uniform distribution can be readily found from the MaxEnt formalism as shown below.

The first step of MaxEnt formalism is to reexpress the multiplicity (Eq. 1) by the form of entropy (i.e. the Shannon information index). This reexpression is easy to derive by using Stirling's approximation (for large  $x$ ):  $\log(x!) = x \log x - x$  (Jaynes 1957, also see Shipley et al. 2006, Pueyo et al. 2007):

$$H = \frac{\log W}{N} = -\sum_{i=1}^6 p_i \log(p_i) \quad (2)$$

where  $p_i$  is the relative abundance of species  $i$ . Mathematically,  $p_i$  must satisfy condition:

$$\sum_{i=1}^6 p_i = 1 \quad (3)$$

According to the Second law of thermodynamics, a system at equilibrium is in the state compatible with its constraints with the maximum possible entropy. Free of any constraints, this is a state where  $p_i$ 's are all equal,  $p_i = 1/6$  ( $i=1, \dots, 6$ ), i.e. maximum entropy would predict flat distributions in the absence of constraints. This is consistent with Bayesian statistics – without any prior knowledge, a uniform distribution is the most plausible model.

Now suppose we have partial knowledge (constraints) about the system and we want to incorporate this knowledge to infer  $p_i$ . Let's assume we know that the six species vary in some life history trait (e.g. leaf thickness), denoted by  $x$ . The average of the thickness across the six species (called the "community-aggregated trait" by Shipley et al. 2006) is:

$$\sum_{i=1}^6 x_i p_i = \bar{x} \quad (4)$$

The objective function for maximizing entropy (Eq. 2), subject to the conditions of Eq. 3 and 4, can be expressed by Lagrange multipliers:

$$f(p_i) = -\sum_{i=1}^6 p_i \log(p_i) + \alpha \left( 1 - \sum_{i=1}^6 p_i \right) + \beta \left( x - \sum_{i=1}^6 x_i p_i \right) \quad (5)$$

where  $\alpha$  and  $\beta$  are parameters. This objective function can be solved by setting the derivative of Eq. 5 with respect to  $p_i$  to zero. With some simple calculus, it can be shown that  $f(p_i)$  reaches maximum when

$$p_i = \frac{e^{-\beta x_i}}{\sum_{j=1}^6 e^{-\beta x_j}} \quad (6)$$

This is an exponential distribution, the celebrated Boltzmann distribution law (Dill and Bromberg 2002). The denominator is called the partition function which links microscopic properties of life-history traits to the macroscopic property of abundance.

The Boltzmann distribution law (Eq. 6) is a powerful model for inferring macroscopic patterns of diversity. Let's use the classical example of dice rolling (Jaynes 1957, Dill and Bromberg 2002) to illustrate how to use this law to predict abundance. We can imagine the six species community as a die with die numbers  $x = \{1, 2, 3, 4, 5, 6\}$  corresponding to trait values. Let's further assume that before rolling we

know nothing about the die: fair or biased. After a certain number of trials, we observe the average score of the trait  $\bar{x} = 3.5$ . Substituting  $x$ ,  $\bar{x}$  and Eq. 6 into 4, we have

$$\frac{\sum_{i=1}^6 x_i e^{-\beta x_i}}{\sum_{i=1}^6 e^{-\beta x_i}} = \frac{e^{-\beta} + 2e^{-2\beta} + \dots + 6e^{-6\beta}}{e^{-\beta} + e^{-2\beta} + \dots + e^{-6\beta}} = 3.5$$

It is easy to find out that  $\beta = 0$  satisfies this equation, leading to  $p = 1/6$  (Fig. 2a). This is not surprising because 3.5 is the average of  $x = \{1, 2, 3, 4, 5, 6\}$  for a fair die. In a different experiment, if the average score is  $\bar{x} = 2.5$ , we would expect that the chance for the smaller face values to appear is higher. In this case,  $\beta = 0.371$  and the distribution predicted by maximum entropy is right skewed (Fig. 2b). In contrast, if the average score is  $\bar{x} = 4.5$ , the larger faces are more likely to appear and  $\beta = -0.371$ . The distribution is left skewed (Fig. 2c).

The implication of this example is remarkable. First, it shows that the prediction of species abundance by the maximum entropy does not require abundance to be measured a priori but is independently made from life history traits. Therefore, MaxEnt formalism itself is not a method that involves circular argument. In practice, empirical  $p_i$ 's are often used to estimate community-aggregated traits ( $\bar{x}$ ), which in turn are used to estimate the same  $p_i$ 's as in the application of Shipley et al. (2006), but this is not an inherent procedure of the MaxEnt formalism. If just the average value of a trait (e.g. leaf thickness) is known (without involving  $p_i$ ) for species in a community, we can then use the method to infer abundances of the species. An example where we may directly measure an average value of a trait is given by remote sensing (Belluco et al. 2006), in which the resulting data are average values of light reflection by plants in a certain zone of the spectrum. By applying the MaxEnt formalism, it is possible to predict the distribution of plant species without knowing empirical  $p_i$ 's.

Second, the MaxEnt formalism is a parsimonious predictive approach. To find out if a die is biased, we could roll it, say 1000 times, and count the number of times each face appears. If the probability distribution is flat, then the die would be considered unbiased. In reality, however, we are seldom able to count all the possibilities if the total number of individuals and the number of species (or molecules or atoms in statistical mechanics) are large, and they usually are. In this situation, the MaxEnt formalism says that the

probability distribution can be inferred from the Boltzmann distribution law if just the average (or total) score is known. Note, however,  $\bar{x} = 3.5$  is evidence for a fair die (but it is not a proof).

As the die example shows, maximum entropy provides a tool to predict the least biased distribution that is consistent with the observed score. It is worth to note, however, this prediction is built on the independent assumption that is needed for counting the multiplicity of Eq. 1. In ecology species often interact in many ways and are not independent. Although it is extremely difficult to count for microscopic interactions of individuals, it is possible to incorporate species interactions in MaxEnt prediction through some specific constraints (Dewar and Porté 2008, Haegeman and Loreau 2008).

### A global pattern of diversity evenness predicted by the MaxEnt formalism

The objective function (Eq. 5) is composed of three components: the entropy, a mathematical constraint, and life history constraints. The distribution that maximizes entropy will represent the 'broadest' (i.e. most even) distribution that agrees with the constraints we impose. Incidentally, this is also the distribution with the highest multiplicity which makes the distribution of species abundances in a community become more evenly distributed. On the other hand, the trait constraints represent the opposite force that drives the system towards skewed distribution. The dynamics of a community is thus controlled by the balance of these two forces.

The balance of entropy maximization and trait constraints can predict, for example, changes in the evenness of tree species abundance along a global scale latitudinal gradient. This prediction can be deduced as follow. The number of life history traits is geographically invariant: whether in a tropical community or a boreal forest, trees all have the same set of traits (e.g. photosynthetic rates, specific leaf area, shade (in) tolerance, size, mass, etc), although their values can vary considerably. But the number of tree species varies by hundred fold from tropical (several hundred species in a hectare plot) to boreal forests (less than ten in the same size of area). Therefore, the entropy (or the multiplicity  $W$  of Eq. 1) of tropical forests is almost inevitably larger than that of temperate forests and, in turn, larger than that of boreal forests, while the force of constraints is relatively constant across the forests (because they have the same number of constraints). This global pattern is consistent with the widely observed latitudinal gra-

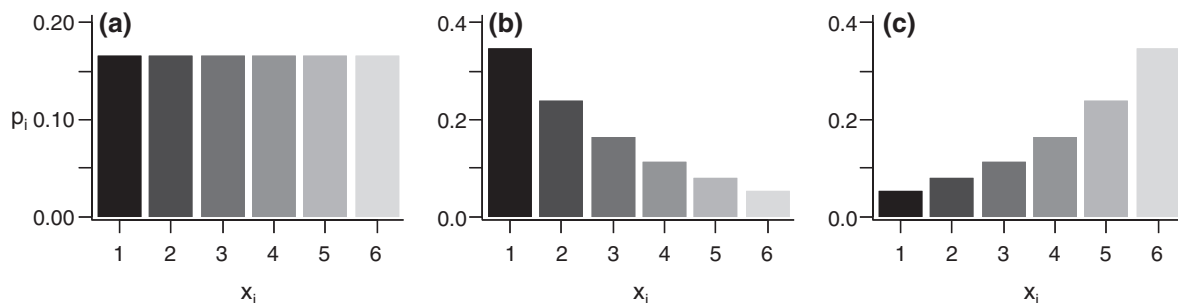


Figure 2. The probability distribution of dice outcomes predicted by the Boltzmann distribution law by knowing only average scores ( $\bar{x}$ ): (a).  $\bar{x} = 3.5$  and  $\beta = 0$ ; (b)  $\bar{x} = 2.5$  and  $\beta = 0.371$ ; and (c)  $\bar{x} = 4.5$  and  $\beta = -0.371$ .

dient in tree species-abundance distributions in which there are dominant species in temperate forests but not in tropical forests (Hubbell 2001). This prediction is also consistent with the observation that species-abundance distributions become more even along a successional gradient (Bazzaz 1975).

Marks and Muller-Landau (2007) did an analysis comparing the importance of entropy in explaining species abundances. They found the importance of entropy decreased with the increase of the number of traits, and species abundances were more successfully explained by the number of traits than entropy. This result is consistent with what is hypothesized here: there is a tradeoff between entropy and constraints.

## The Boltzmann distribution law and logistic regression

The MaxEnt formalism and logistic regression are two approaches that differ in two fundamental ways: (1) the MaxEnt formalism is Bayesian inference, while logistic regression is frequentist statistics, and (2) the MaxEnt formalism is a predictive tool (putting the circular data treatment aside), while logistic regression is a data fitting method. Despite of these differences, it is interesting to observe the connection between the Boltzmann distribution and the statistical model of logistic regression (Blower 2004).

The logistic regression is sometimes called maximum-entropy classifier in the literature and has been used in machine learning (Berger et al. 1996) and modeling species distribution (Phillips et al. 2006). The Boltzmann distribution law (Eq. 6) for two species in essence has the same mathematical form of logistic regression model. For more than two species, it is multicategorical (or called multinomial) logistic regression (Agresti 2002). This equivalence is not surprising because both maximum entropy and the multinomial logistic regression stem from the same multinomial distribution, which, for the case of dice rolling, is

$$W = \frac{N!}{n_1!n_2!n_3!n_4!n_5!n_6!} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} p_5^{n_5} p_6^{n_6} \quad (7)$$

where  $N = \sum_{i=1}^6 n_i$  and  $\sum_{i=1}^6 p_i = 1$ . For the logistic regression of two species, the multinomial distribution is reduced to a binomial distribution.

The establishment of this connection is practically very useful because species abundance can now be easily modeled using logistic regression and it can take as many explanatory variables (constraints) as possible, whether they are life history traits or environmental factors. As an example, let's consider a community of three species with two explanatory variables, the logistic models are:

$$p_i = \frac{e^{-(\beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i})}}{1 + \sum_{j=1}^2 e^{-(\beta_{0j} + \beta_{1j}x_{1j} + \beta_{2j}x_{2j})}} \quad i = 1, 2 \quad (8)$$

where  $x_{1j}$  and  $x_{2j}$  are traits measured for the two species, and  $\beta_{ij}$  are regression coefficients. There are only two equations ( $i = 1, 2$ ) for three species because one of them is redundant due to  $p_1 + p_2 + p_3 = 1$ . The occurrence of 1 (and  $\beta_{0i}$ ) in the denominator of Eq. 8 arises from the fact that for the third species,  $x_{1i} = x_{2i} = 0$ . Note this 1 is seemingly missing from Eq.

6; it is easy to show Eq. 8 and 6 are identical if one of the  $p_i$ 's in Eq. 6 is treated as a reference baseline (because  $\sum p_i = 1$ , one of the  $p_i$ 's is completely determined by the rest). Equation. 8 can be easily parameterized using the maximum likelihood method and the significance of the explanatory variables ( $x_1$  and  $x_2$ ) can be tested using the Wald statistic (Agresti 2002).

Although the Boltzmann law provides a statistical mechanical interpretation for logistic regression models (Blower 2004), the law does not have a mechanism for variable selection. One can dump whatever traits in hands into Eq. 6 even though the variables may not contribute to abundance at all. Different from the MaxEnt formalism, logistic regression and the maximum likelihood method provide statistical guidelines for variable selection. Roxburgh and Mokany (2007) have done an interesting experiment by randomly assigning traits to each species. They showed that the randomly assigned traits still explained a considerable amount of variation of species abundance. This puzzle can easily be explained by the logistic regression (Eq. 8) in which none of the traits, except the intercept term ( $\beta_0$ ), is significant. Random assignment of traits does not affect  $\beta_0$  and consequently the explained variation remains the same. This example reiterates the importance of conducting model selection. Therefore, when it comes to modelling abundance-trait relationship, I would suggest using logistic regression by taking advantage of the well established statistical theory of the method and the model selection procedure. This suggestion may diminish the significance of the MaxEnt formalism as a predictive tool. In any case, in real applications rarely can we obtain independent average trait values – average traits almost always have to be estimated from empirical  $p_i$ 's as did in Shipley et al. (2006). The circular use of the MaxEnt formalism due to data treatment does not render the method any advantage over logistic regression.

Cross-validation has shown that the abundance variation explained by a hypothetical set of eight traits is much less than the original prediction of Shipley et al. (0.32 vs 0.94 in  $r^2$ ) (Marks and Muller-Landau 2007). Rare species are particularly poorly predicted. This low explanation power is neither surprising nor a failure of the method, however. Given that species abundance can be influenced by almost any factors and none of them may be dominating, eight traits should not be considered as a large set of predictors. Moreover, some of the traits may even be found to contribute very little to the variation of abundance if model selection is performed. In this context, two situations may arise from the inference of MaxEnt. One is that if a constraint is a linear combination of other constraints (multicollinearity), then it will have little influence on abundance. The other is that if the Lagrange multiplier (an approximate measure of the importance of a trait) is zero (or very small), then the constraint will have no (or little) influence.

In addition to the lack of model selection procedure, the MaxEnt formalism, comparing to logistic regression, suffers three other limitations. (1) The current use of the MaxEnt formalism in ecology is restricted to the Boltzmann distribution which only accounts for the linear effect of traits on abundance (Shipley et al. 2006). This is not a problem for modeling gas and fluid because higher order terms are not needed there. But in ecology it is important to generalize Boltzmann distribution to include nonlinear effects (e.g.  $x^2$  or interactive terms) because the effect of traits may well be nonlin-

ear. Nonlinear terms can and are used in MaxEnt procedure (Gull and Daniell 1978, Jaynes 1979). (2) To find appropriate constraints is critical for the success of the MaxEnt prediction. As pointed out by Haegeman and Loreau (2008), the MaxEnt formalism is not able to identify critical predictors that may be missing from the data. This inadequacy however can be readily rescued by the goodness-of-fit assessment of logistic regression. A systematic departure of residuals in regression suggests certain key explanatory variables may be missing. Currently, the goodness-of-prediction procedure of MaxEnt has not yet been fully developed. Although cross-validation may be used to identify the importance of variables under evaluation (Marks and Muller-Landau 2007), it cannot say anything about what may be missing. (3) The condition  $\sum p_i = 1$  requires that the species of the community be completely surveyed. This condition will not be met if some species are missing from the survey. Although the missing species can be dealt with by renormalization, it is currently not clear how this would affect the result of the MaxEnt inference. Given that few field surveys are true census, it is essential to evaluate the consequence of missing species before the MaxEnt formalism can be fully appreciated. In applications when absolute abundance (not the probability  $p_i$ ) is the focus of interest, the MaxEnt formalism (and logistic regression) is not appropriate but Poisson regression should be used instead.

### Concluding remarks

Statistical mechanics has provided an elegant tool for predicting species abundance. But it is not a panacea. Given the complexity of ecosystems, one should not naively expect that a dozen of traits could satisfactorily predict even the primary properties of ecosystems. It is still a long way to go before we can confidently answer the question of why there are so few or so many of species. One of the challenges would be to think how we may add more predictive power while keeping the elegance of the statistical mechanical approach. It is perhaps time to revise the Boltzmann distribution law (e.g. by including nonlinear and interactive terms of predictive variables) for predicting species abundance.

*Acknowledgements* – I thank Tommaso Zillio and Salvador Pueyo for discussions that helped me understand the maximum entropy formalism. Owen Petchey provided constructive comments that significantly improved the study. Tommaso helped revision of an earlier draft. This work was supported by the National Excellent Centres for the Sustainable Forest Management and the GEOIDE of Canada, and the Natural Sciences and Engineering Research Council of Canada.

### References

Agresti, A. 2002. *Categorical data analysis* (2nd ed.). – Wiley.  
 Alexeyev, V. L. and Levich, P. 1997. A search for maximum species abundances in ecological communities under conditional diversity optimization. – *Bull. Math. Biol.* 59: 649–677.

Banavar, J. R. and Maritan, A. 2007. The maximum relative entropy principle. – arXiv:cond-mat/0703622v1.  
 Bazzaz, F. A. 1975. Plant species diversity in old-field successional ecosystems in southern Illinois. – *Ecology* 56: 485–488.  
 Berger, A. L. et al. 1996. A maximum entropy approach to natural language processing. – *Comput. Linguistics* 22: 39–71.  
 Blower, D. J. 2004. An easy derivation of logistic regression from the Bayesian and maximum entropy perspective. – *Bayesian inference and maximum entropy methods in science and engineering*. 23rd Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conf. Proc. 707: 30–43.  
 Belluco, E. et al. 2006. Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. – *Remote Sens. Environ.* 105: 54–67.  
 Dewar, R. C. and Porté, A. 2008. Statistical mechanics unifies different ecological patterns. – *J. Theor. Biol.* 251: 389–403.  
 Dill, K. A. and Bromberg, S. 2002. Molecular driving forces: statistical thermodynamics in chemistry and biology. – Garland Science.  
 Gull, S. F. and Daniell, G. J. 1978. Image reconstruction from incomplete and noisy data. – *Nature* 272: 686–690.  
 Haegeman, B. and Loreau, M. 2008. Limitations of entropy maximization in ecology. – *Oikos* 117: 1700–1710.  
 Harte, J. et al. 2008. Maximum entropy and the state-variable approach to macroecology. – *Ecology* 89: 2700–2711.  
 Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. – Princeton Univ. Press.  
 Jaynes, E. T. 1957. Information theory and statistical mechanics. – *Phys. Rev.* 106: 620–630.  
 Jaynes, E. T. 1979. Where do we stand on maximum entropy? – In: Levine, R. D. and Tribus, M. (eds), *The maximum entropy formalism*. MIT Press, pp. 15–118.  
 Jaynes, E. T. 2003. *Probability theory: the logic of science*. – Cambridge Univ. Press.  
 Levich, A. P. 2000. Variational modelling theorems and algoecoenoses functioning principles. – *Ecol. Modell.* 131: 207–227.  
 McGill, B. J. 2006. A renaissance in the study of abundance. – *Science* 314: 770–772.  
 McGill, B. J. et al. 2006. Rebuilding community ecology from functional traits. – *Trends Ecol. Evol.* 21: 178–185.  
 Marks, C. O. and Muller-Landau, H. C. 2007. Comment on “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. – *Science* 316: 1425c.  
 Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographical distributions. – *Ecol. Modell.* 190: 231–259.  
 Pueyo, S. et al. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. – *Ecol. Lett.* 10: 1017–1028.  
 Roxburgh, S. H. and Mokany, K. 2007. Comment on “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. – *Science* 316: 1425b.  
 Shipley, B. et al. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. – *Science* 314: 812–814.  
 Shipley, B. et al. 2007. Response to comments on “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. – *Science* 316: 1425d.  
 Shipley, B. 2008. Limitations of entropy maximization in ecology: a reply to Haegeman and Loreau. – *Oikos* 118: 152–159.  
 Whitefield, J. 2007. Survival of the likeliest? – *PLoS Biol.* 5: e142.