

Article

Remarks on the Maximum Entropy Principle with Application to the Maximum Entropy Theory of Ecology

Marco Favretti 

Dipartimento di Matematica “Tullio Levi-Civita”, Università degli Studi di Padova, 35122 Padova, Italy; favretti@math.unipd.it; Tel.: +39-049-827-1408

Received: 1 November 2017; Accepted: 21 December 2017; Published: 27 December 2017

Abstract: In the first part of the paper we work out the consequences of the fact that Jaynes’ Maximum Entropy Principle, when translated in mathematical terms, is a *constrained extremum problem* for an entropy function $H(p)$ expressing the uncertainty associated with the probability distribution p . Consequently, if two observers use different independent variables p or $g(p)$, the associated entropy functions have to be defined accordingly and they are different in the general case. In the second part we apply our findings to an analysis of the foundations of the Maximum Entropy Theory of Ecology (M.E.T.E.) a purely statistical model of an ecological community. Since the theory has received considerable attention by the scientific community, we hope to give a useful contribution to the same community by showing that the procedure of application of MEP, in the light of the theory developed in the first part, suffers from some incongruences. We exhibit an alternative formulation which is free from these limitations and that gives different results.

Keywords: Maximum Entropy principle; Maximum Entropy Theory of Ecology; Shannon entropy; Boltzmann counting

1. Introduction

Maximum Entropy Principle MEP (E.T. Jaynes, 1957, see [1,2]) is a powerful inference principle which allows to determine the probability distribution that describes a system on the basis of the information available, usually in the form of averages of observables (random variables) of interest for the system. It is based on: (i) the enumeration of the system states $i = 1, \dots, N$; (ii) the introduction of one or more functions that translate the information available of the system in the form of constraints on the probability i.e., as $f(p) = c$ where c is a vector of average values; and (iii) a function measuring the *uncertainty* associated to a probability distribution candidate to describe the system. We consider here only systems with a finite number of states since the extra generality of considering an infinite number of states is not necessary for the aims of this work. The sought distribution is the one for which uncertainty is maximal on the set of distributions for which $f(p) = c$. This distribution thus represents the least biased estimate on the basis of the given information whether or not its prescriptions are empirically confirmed by experiments. Each the above outlined Steps (i)–(iii) has a profound meaning and impact on the implementation of the MEP procedure. For example, if the states of the system are a priori equally probable, the uncertainty function to use is Shannon entropy $H(p)$, while if they are statistically described by a prior distribution q the uncertainty function is (minus) the relative entropy $D(p|q)$. If the systems states are the result of a coarse-graining procedure from a bigger set of a priori equiprobable states, then a prior distribution q which counts the number of coarse-grained states has to be used for $D(p|q)$ in order to render the MEP procedure invariant with respect to the coarse graining as explained in [3]. See also [4] where a number of different probability distribution used in ecology are derived for a system of individuals arranged in different cells by simply making different assumptions

on the individuals (indistinguishable or not), on the constraints and on the coarse graining. The variety of MEP responses simply reflects the facts that different pieces of information are considered; if the previsions fail to be confirmed by the experiment, this is the signal that relevant information for the statistical description on the system under study has been neglected.

A basic tenet of the Maximum Entropy Principle is that two observer which are given the same information expressed by Steps (i) and (ii) above must obtain the same result upon application of the MEP, i.e., the solution to the constrained extremum problem has to be unique. Notice that the MEP procedure does not specify the probability $p = (p_1, \dots, p_N)$ to use; thus if one observer uses p and another uses p' which is related to p by a one-to-one transformation $p = g(p')$, provided that they set up the constraints in the form $f(p) = c$ and $f(g(p')) = c$ respectively, they are translating in mathematical terms the same information on the system. This represents the same degree of freedom of using different (but related by an invertible transformation) systems of coordinates to describe the position of a point in physics. Therefore, to ensure that the results of the application of the MEP procedure are independent of the choice of the p used the uncertainty function has to be defined accordingly. Proposition 1 in Section 2 below is the main tool to determine the form of the uncertainty function using g -related distributions.

The analysis of this subtle point of the application of MEP is the main aim of this paper. We argue that since there seems not to be a preferred choice for p they have to be considered all equivalent; therefore the related uncertainty functions are also equivalent and can be derived by a single given one. The question is: which is the translation in mathematical terms of the initial information which is entitled to use as uncertainty function the Shannon entropy? We give an answer by resorting to the Boltzmann statistical derivation of the entropy function, which has the same form of Shannon one. This statistical procedure, based on independent tosses of indistinguishable particles in bins, can be applied to different models and allows determining from a sort of “first principle” the form of the entropy function to use. We think that there is room for further investigation of this subtle issue of the application of MEP.

In the second part of the paper we apply the above findings to a specific application of MEP to ecology called Maximum Entropy Theory of Ecology (M.E.T.E.) as exposed in [5]. M.E.T.E. is a purely statistical model of an ecological community. Using the Maximum Entropy Principle the theory aims at inferring the form of some of the most used distributions in ecology from the knowledge of macroscopic information on the system: number of individuals, number of species and total metabolic requirement. MEP has been applied in the field of statistical ecology mainly to derive one of the most important probability distributions, the Species Abundance Distribution (SAD) but not exclusively (see [6] or [7] where it is used for modeling species geographic distributions with presence-only data). Several papers ([3,8,9]) have revised the application of the MEP inference principle to ecological problems, stressing the importance of the choice of the system' coordinate [10–12], of the prior distribution [4,8] and of the entropy function [3,6]. The interest of [5] is that it considers simultaneously information on the distribution of individuals among species (SAD) and on the distribution of metabolic energy rate among individuals. In this richer scenario a number of probability distributions for different observables of interest are derived. The output of the work has grown to become a comprehensive theory of application of MEP to ecological problems, called Maximum Entropy Theory of Ecology (see the book [13,14]) which has been extensively tested with real ecosystem data in [15–17]. In Section 4.1 we review the assumptions of the model in [5] and the related MEP solution; in Section 4.2 we propose an alternative application of the MEP procedure starting from the same initial information but giving a different result. In Section 5 we discuss the non equivalence of the two procedures using the Proposition 1 below and motivate why the MEP solution in [5] is flawed by some inconsistencies.

2. MEP Formulation Using Different Variables

The Maximum Entropy Principle is generally expressed as: p is the distribution that maximizes Shannon entropy $H(p)$ over the set of distributions allowed by the constraints. Let p' defined by

$p' = g^{-1}(p)$ with g invertible be a different choice of the variable to be used to express the probability distribution and the constraints and suppose that for the solution of this constrained extremum problem we use Lagrange multipliers method. Then, the following Proposition holds (see Appendix A for a proof).

Proposition 1. \hat{p} is the constrained extremum of $H(p)$ over the set $f(p) = c$ if and only if $\hat{p}' = g^{-1}(\hat{p})$ is the constrained extremum of $H(g(p'))$ over the set $f(g(p')) = c$.

From the above Proposition 1 we learn that if we are given the entropy function and the constraints as $(p, f(p), H(p))$ and we want to use another variable $p' = g^{-1}(p)$, to find the same MEP solution we have to transform the constraints *and* the entropy function accordingly using $(p', f \circ g(p'), H \circ g(p'))$. In this way, changing variable and finding the maximum are operations whose order can be interchanged. A natural question arises in this setting: which one is the couple $(p, f(p))$ translating in mathematical terms the initial information on the system that “is justified” in using the entropy function in Shannon form? We will give a partial answer to this question in Section 3 below.

Note that the invariance requirement with respect to the choice of the independent variable is different from the requirement, introduced in the axiomatic derivation of MEP in [18], of invariance *in the form of* $H(p)$ with respect to a transformation of the type $y = \Gamma(x)$, where $x \in D$ are the (possibly infinite) states of a system and $p = p(x)$. For a system with finite states the transformation Γ (called coordinate transformation in [18]) reduces to a relabeling of the states. Shannon entropy $H(p) = -\sum_{i=1}^N p_i \ln p_i$ being additive is clearly invariant in form, expressing the fact that different observers may use different labels for the N system states. This relabeling transformation

$$p'_k = g_k(p_1, \dots, p_N) = p_{\varphi_k(1, \dots, N)} \quad k = 1, \dots, N$$

where φ_k is a permutation of the N state labels is thus a particular case of the more general g considered in the above Proposition 1. It is clearly an invertible transformation in which $H(p)$ and $H(g(p'))$ do coincide, but this is no longer assured for more general transformations g ; see Section 3 below.

3. A Simple, Ecologically Oriented Example

To throw light on the questions considered above we take a very simple system used in a number of physical models that for our purposes can be considered as a minimal example of an ecological community. Consider a set \mathcal{I} of N_0 identical individuals (balls) which belong to S species (indistinguishable urns). Suppose that each species contains at least one individual so that the maximal number of individuals allowed to one species is $N = N_0 - (S - 1)$. Suppose that the species labels can be interchanged so that the observable quantity is not the number of individuals n_α of a given species α but the number S_n of species which have exactly n individuals, for $n = 1, \dots, N$.

Suppose that the set of species \mathcal{S} is equipped with the uniform probability $1/S$ and consider the random variable $v : \mathcal{S} \rightarrow \mathcal{N}, \alpha \mapsto v(\alpha) = n_\alpha, \mathcal{N} = \{1, \dots, N\}$ that assigns to a species its abundance. Let $\phi(n)$ be the density associated to the random variable v (in the sequel P denotes probability)

$$\phi(n) = P(v = n) = P(\alpha \in v^{-1}(n)) = \frac{|v^{-1}(n)|}{S} = \frac{S_n}{S}, \quad n \in \mathcal{N} \tag{1}$$

where $|A|$ is the number of elements of A . Set for later use $\mathcal{S}_n = v^{-1}(n) \subset \mathcal{S}$, therefore $|\mathcal{S}_n| = S_n$.

Let the set of individuals \mathcal{I} be equipped with the uniform probability $1/N_0$, let $\alpha : \mathcal{I} \rightarrow \mathcal{S}, i \mapsto \alpha(i)$ be the assignation of each individual to a species and consider the random variable $v \circ \alpha : \mathcal{I} \rightarrow \mathcal{N}, i \mapsto v(\alpha(i))$. Moreover, consider the following subset \mathcal{I}_n of \mathcal{I} obtained by grouping together all the individuals belonging to the species having n individuals (we call this subset the “multispecies” n)

$$\mathcal{I}_n = \{i \in \mathcal{I} : v(\alpha(i)) = n\} \subset \mathcal{I}, \quad I_n = |\mathcal{I}_n|. \tag{2}$$

As a straightforward consequence, we have that $I_n = nS_n$. Let $\phi_I(n)$ be the density associated to the random variable $\nu \circ \alpha$ that is

$$\phi_I(n) = P(\nu \circ \alpha = n) = P(i \in \mathcal{I}_n) = \frac{I_n}{N_0}, \quad n \in \mathcal{N}. \tag{3}$$

Both ϕ and ϕ_I are probability distributions on \mathcal{N} ; it is easy to see that since $I_n = nS_n$, they are related by an invertible transformation $\phi_I = g(\phi)$ defined as

$$\phi_I(n) = g_n(\phi) = \frac{nS}{N_0}\phi(n) \quad n = 1, \dots, N. \tag{4}$$

In ecology the distribution ϕ is called Species Abundance Distribution (SAD) and is probably the most important (and investigated) indicator of how an ecological community is structured. While $\phi(n)$ gives the probability of extracting a species of abundance n from the set of species, $\phi_I(n)$ gives the probability that an individual extracted from the set \mathcal{I} belongs to a species of abundance n . Real communities are composed of many species with few individuals (rare species) and few ones which contain the majority of the individuals (common species). As a consequence, the two probability distribution are very different.

The information on the system resumed by the numbers N_0, S sets the following equalities

$$N_0 = \sum_{n=1}^N nS_n = \sum_{n=1}^N I_n \quad \text{and} \quad S = \sum_{n=1}^N S_n = \sum_{n=1}^N \frac{1}{n} I_n. \tag{5}$$

If we divide the above two equalities by S or N_0 respectively, we obtain the following equations to be satisfied by ϕ

$$\frac{N_0}{S} = \sum_{n=1}^N n\phi(n), \quad 1 = \sum_{n=1}^N \phi(n) \tag{6}$$

and ϕ_I respectively

$$1 = \sum_{n=1}^N \phi_I(n), \quad \frac{S}{N_0} = \sum_{n=1}^N \frac{1}{n} \phi_I(n). \tag{7}$$

Constraints (6) and (7) are g -related, that is ϕ satisfies Equation (6) if and only if $g(\phi)$ satisfies Equation (7). Note that in first case we are constraining the weighted *arithmetic* average of $1, \dots, N$ while in second one we are constraining the weighted *harmonic* average. Note also that the actual information being used is the *ratio* N_0/S or equivalently S/N_0 , so that the MEP description of the system is independent of the size N_0, S of the system provided that the ratio N_0/S is kept constant.

Suppose that two different observers translate in mathematical terms the information contained in the data N_0, S using the two above introduced variables ϕ and ϕ_I and that both invoke the MEP to find the least informative probability distribution allowed by the constraints. The application of MEP with entropy function $H(\phi) = -\sum_{n=1}^N \phi(n) \ln \phi(n)$ and Constraint (6) gives, upon application of the Lagrange multipliers methods, the geometric probability distribution

$$\hat{\phi}(n) = \frac{e^{-\lambda n}}{Z(\lambda)}, \quad Z(\lambda) = \sum_{n=1}^N e^{-\lambda n} \tag{8}$$

while the MEP procedure with *the same* entropy function $H(\phi_I) = -\sum_{n=1}^N \phi_I(n) \ln \phi_I(n)$ and Constraint (7) gives the solution

$$\hat{\phi}_I(n) = \frac{e^{-\gamma \frac{1}{n}}}{Z(\gamma)}, \quad Z(\gamma) = \sum_{n=1}^N e^{-\gamma \frac{1}{n}} \tag{9}$$

whose corresponding SAD is

$$\tilde{\phi}(n) = g^{-1}(\hat{\phi}_I)(n) = \frac{N_0 e^{-\gamma \frac{1}{n}}}{Sn Z(\gamma)}. \quad (10)$$

It is easy to see that $\hat{\phi} \neq g^{-1}(\hat{\phi}_I)$, therefore with this example we have shown that changing independent variables and using the MEP with the Shannon entropy are *non commuting operations* in the general case. Had we used Constraint (7) with the transformed entropy

$$H'(\phi_I) = H(g^{-1}(\phi_I)) = - \sum_{n=1}^N \phi_I(n) \frac{N_0}{nS} \ln[\phi_I(n) \frac{N_0}{nS}] \quad (11)$$

we would have obtained the solution $g(\hat{\phi})$ with $\hat{\phi}$ in Equation (8). Note that the transformed entropy in Equation (11) does not have the form of a relative entropy function therefore it can not be obtained by introducing a suitable prior probability distribution. Note also that the non commutativity considered above is different from the non commutativity of MEP procedure with respect to *coarse graining* of the system state considered in [3,4].

Which One is the Correct Application of MEP ?

Both $\phi(n)$ in Equation (8) and $\tilde{\phi}(n)$ in Equation (10) result from the application of the MEP starting from the same initial information N_0, S on the system and provide non equivalent answers to the question of how the abundance of individuals is distributed among the species. Since the set of distributions allowed by the constraints (feasible set) in the two formulations Equations (6) and (7) are related by an invertible transformation $D_I = g(D)$, the non equivalence must reside in the choice of the form of the entropy functions which are not g -related according to Proposition 1. On the basis of the same Proposition, once the form of the entropy function is established in a given formulation of MEP ($p, f(p), H(p)$), it is also determined for all g -related formulations. It seems thus necessary to invoke a sort of first principle to establish the correct form of the entropy function in the initial MEP formulation.

A possible approach could be to compare the form of the resulting probability distribution against field data. This is not entirely satisfactory since the discrepancy between the MEP solution and the empirical curve could be due to the fact that we have neglected (or we are not aware of) some information on the system which is relevant for the question addressed. In this case the discrepancy between the model and the empirical pattern is a signal that other factors shaping the probability distribution are acting on the system. By the way, for the problem considered above, it is generally acknowledged in the ecological literature (see e.g., [5]) that the geometric SAD in Equation (8) has a poor fit with the empirical SAD, while the one introduced in (10) could provide a better fit since it is more flexible (it can be monotonically decreasing or display an interior mode depending on the value of the Lagrange multiplier).

At least for systems with a discrete number of states another criterion can be introduced, which uses the combinatorial argument providing one of the justifications for the form of Shannon entropy historically preceding the axiomatic derivation in [18].

This is the celebrated Boltzmann problem of statistical mechanics [19]. It is useful to briefly review its solution here because it tells us how to determine the correct form of the entropy function from first principles in cases where its determination is not obvious. It deals with the distribution of K indistinguishable particles (individuals) among M equally spaced energy levels $\epsilon = 1, \dots, M$. Boltzmann original formulation dealt with arbitrarily spaced energy levels but this extra generality is not needed here. If n_ϵ is the number of particles having energy ϵ and $\mathbf{n} = (n_1, \dots, n_M)$ is the vector of occupation numbers of the energy levels, the logarithm of the probability of \mathbf{n} is

$$\ln W(\mathbf{n}) = \ln \frac{1}{M^K} \frac{K!}{n_1! \dots n_M!} \quad (12)$$

which using Stirling approximation $\ln n! \approx n \ln n$ can be rewritten as

$$\ln W(\mathbf{n}) = K \ln K - \sum_{\epsilon} n_{\epsilon} \ln n_{\epsilon}. \tag{13}$$

The maximization of $\ln W(\mathbf{n})$ under the constraints $\sum_{\epsilon} n_{\epsilon} = K$ and $\sum_{\epsilon} \epsilon n_{\epsilon} = E_0$ gives the most common, i.e., least informative vector of occupation numbers. The solution of the constrained maximization problem is the celebrated Boltzmann-Gibbs distribution

$$n_{\epsilon} = \frac{K e^{-\beta \epsilon}}{\sum_{\epsilon=1}^M e^{-\beta \epsilon}}, \quad \psi(\epsilon) = \frac{n_{\epsilon}}{K} = \frac{e^{-\beta \epsilon}}{\sum_{\epsilon=1}^M e^{-\beta \epsilon}} \tag{14}$$

Note that writing $n_{\epsilon} = \psi(\epsilon)K$, $\ln W$ in Equation (13) coincides up to a constant with

$$H(\psi) = - \sum_{\epsilon} \psi(\epsilon) \ln \psi(\epsilon). \tag{15}$$

In the same manner, if we have S indistinguishable species to be assigned in the N levels (abundances) of \mathcal{N} let $\mathbf{s} = (S_1, \dots, S_N)$ be the occupation vector with constraints $\sum_n S_n = S$ and $\sum_n n S_n = N_0$. As before, setting $S\phi(n) = S_n$ the least informative probability distribution is

$$\phi(n) = \frac{e^{-\lambda n}}{\sum_n e^{-\lambda n}} = \frac{e^{-\lambda n}}{Z(\lambda)} \tag{16}$$

and the associated entropy is

$$H(\phi) = - \sum_n \phi(n) \ln \phi(n) \tag{17}$$

Therefore the form of the entropy function for the formulation given by Equation (6) of the MEP problem using the variable ϕ can be derived by Boltzmann combinatorial argument.

Let us see which is the resulting entropy function if we apply Boltzmann argument to the formulation given by Equation (7) using ϕ_I . Remember that

$$(A) \quad \phi(n) = \text{Prob}(\alpha \in S_n), \quad (B) \quad \phi_I(n) = \text{Prob}(i \in \mathcal{I}_n) \tag{18}$$

so basically in (A) we are “tossing” S species in the N bins S_n while in (B) we are “tossing” N_0 individuals in the N multispecies I_n . In the Boltzmann argument the *tosses are supposed to be independent* and this is true for (A) but not for (B). Indeed a multispecies I_n is the grouping of S_n species each having n individuals so the number of individuals in \mathcal{I}_n can be updated only by multiples of n , i.e., adding or subtracting a species of n individuals. Hence the tosses of single individuals does not represent independent tosses. If we want to support the choice of the entropy function by Boltzmann argument we have to consider *tosses of groups* of n individuals. Let i_n be the number of individuals tossed in \mathcal{I}_n . Therefore the related occupation vector has the form

$$\mathbf{i} = \left(\frac{i_1}{1}, \dots, \frac{i_n}{n}, \dots, \frac{i_N}{N} \right), \quad W(\mathbf{i}) = \frac{1}{N^{N_0}} \frac{N_0!}{i_1! \dots i_N!} \tag{19}$$

which leads to an entropy (recall that $i_n = N_0 \phi_I(n)$) of the form

$$H(\phi_I) = - \sum_n \frac{N_0}{n} \phi_I(n) \ln \left[\frac{N_0}{n} \phi_I(n) \right]$$

and under the Constraints (7) above to the solution

$$\phi_I(n) = n \frac{S}{N_0} \frac{e^{-\lambda n}}{Z(\lambda)} = g(\hat{\phi}(n)). \tag{20}$$

We have shown that in the case where the probability distribution being used can be linked at least conceptually to independent repetitions of an experiment, we have a criterion to derive the correct form of the entropy function from a first principle. Of course to build a general theory this simple example deserves further generalization.

4. Maximum Entropy Theory of Ecology

4.1. Review of M.E.T.E. Assumptions

Let us review the assumptions of M.E.T.E. as exposed in [5]. In [5], various metrics (probability distributions) are derived but here we will content ourselves with the analysis of the derivation of two of them, namely the Species Abundance Distribution $\phi(n)$ and the distribution of metabolic energy $\psi(\epsilon)$ between the individuals.

Following [5], we consider a community made of N_0 individuals living on an area A_0 , subdivided into S species and having total metabolic energy rate E_0 . Assume that each species has at least one individual and that the individual metabolic rate is a discrete quantity which ranges from 1 to a maximum M in a suitable energy unit. With these assumptions, the abundance of a species α is $\nu(\alpha) = 1, \dots, N$ with $N = N_0 - S + 1$ and the individual metabolic rate is $m(i) \in \mathcal{M} = \{1, \dots, M\}$, with $M = E_0 - S_0 + 1$. Each individual is labelled by a double label (α, ϵ) indicating the species $\alpha = 1, \dots, S$ and the metabolic rate $\epsilon = 1, \dots, M$. Here we depart from [5] by assuming for the sake of simplicity that the individual metabolic rate is a discrete quantity while in [5] it is assumed to be a continuous one. This change does not affects our argument since for dealing with the continuous case it is enough to substitute the sums with integrals.

The information on the system is limited to the values N_0, S, E_0 , and A_0 , but for our aims the knowledge of A_0 is not relevant. Therefore we are dealing with a non spatial model of a community of S non interacting species. On the basis of this sole knowledge we would like to answer to two questions: (Q₁) How is the energy distributed among the individuals? (Q₂) How are the individuals distributed among the species? Note that, since the species' label can be exchanged, the correct formulation of (Q₂) is how many of the species have exactly n individuals for $n = 1, \dots, N$, which is the information contained in $\phi(n)$.

Both questions can be answered in a non deterministic manner; for (Q₂), we use the density ϕ introduced in Equation (1), while, for (Q₁), we introduce the map $m : \mathcal{I} \rightarrow \mathcal{M}, i \mapsto m(i)$ that we will consider as a random variable and its density $\psi(\epsilon)$ on \mathcal{M}

$$\psi(\epsilon) = P(m = \epsilon) = P(i \in m^{-1}(\epsilon)) = \frac{I_\epsilon}{N_0}, \quad \forall \epsilon \in \mathcal{M} \tag{21}$$

where $\mathcal{I}_\epsilon = m^{-1}(\epsilon) \subset \mathcal{I}$ and $I_\epsilon = |\mathcal{I}_\epsilon|$. For later use, consider also the joint probability distribution

$$P(n, \epsilon) = Prob(i \in \mathcal{I}_n, i \in \mathcal{I}_\epsilon) = \frac{1}{N_0} |\mathcal{I}_n \cap \mathcal{I}_\epsilon| \tag{22}$$

having marginals, respectively, $\phi_{\mathcal{I}}(n)$ and $\psi(\epsilon)$, since the space \mathcal{I} is doubly partitioned in multispecies and energy classes.

4.2. Our Solution of M.E.T.E. Problem

The following constraints contain all the information (N_0, S, E_0) available on the system

$$\sum_{\epsilon} \psi(\epsilon) = 1, \quad \sum_{\epsilon} \epsilon \psi(\epsilon) = \frac{E_0}{N_0} \tag{23}$$

$$\sum_n \phi(n) = 1, \quad \sum_n n \phi(n) = \frac{N_0}{S} \tag{24}$$

and a straightforward application of MEP with the constraints in Equation (23) and the entropy in Equation (15) [with the constraints in Equation (24) and the entropy in Equation (17) respectively] gives the answers $\hat{\psi}$ in Equation (14) [$\hat{\phi}$ in Equation (16)] to the above questions (Q₁) and (Q₂), i.e.,

$$\hat{\psi}(\epsilon) = \frac{e^{-\beta\epsilon}}{Z(\beta)} = P(i \in \mathcal{I}_\epsilon), \quad \hat{\phi}(n) = \frac{e^{-\lambda n}}{Z(\lambda)} = P(\alpha \in \mathcal{S}_n).$$

Basically, we are solving twice Boltzmann problem of statistical mechanics. Now, if one is interested in the joint probability

$$Q(n, \epsilon) = P(\alpha \in \mathcal{S}_n, i \in \mathcal{I}_\epsilon) \tag{25}$$

we proceed as before, introducing the vector $\mathbf{c} = (C_{1,1}, \dots, C_{N,M})$ where $C_{n,\epsilon}$ counts the number of times the object (α, i) is assigned to the bin $\mathcal{S}_n \times \mathcal{I}_\epsilon$. Hence

$$W(\mathbf{c}) = \frac{1}{(NM)^{N_0S}} \frac{(N_0S)!}{C_{1,1}! \dots C_{N,M}!} \tag{26}$$

with

$$\ln W(\mathbf{c}) = -N_0S \ln N_0S - \sum_{n,\epsilon} C_{n,\epsilon} \ln C_{n,\epsilon} \tag{27}$$

to be maximized under the *decoupled* constraints ($\sum_\epsilon C_{n,\epsilon} = S_n, \sum_n C_{n,\epsilon} = n_\epsilon$)

$$\sum_{n,\epsilon} C_{n,\epsilon} = N_0S, \quad \sum_{n,\epsilon} n C_{n,\epsilon} = N_0, \quad \sum_{n,\epsilon} \epsilon C_{n,\epsilon} = E_0. \tag{28}$$

The related entropy is, setting $C_{n,\epsilon} = Q(n, \epsilon)N_0S$

$$H_{\mathcal{N} \times \mathcal{M}}(Q) = - \sum_{n,\epsilon} Q(n, \epsilon) \ln Q(n, \epsilon) \tag{29}$$

and the maximum entropy distribution with Constraint (28) can be shown to be

$$Q(n, \epsilon) = \phi(n)\psi(\epsilon) = \frac{e^{-\lambda n}}{Z(\lambda)} \frac{e^{-\beta\epsilon}}{Z(\beta)}. \tag{30}$$

We claim that Equation (30) is the correct solution of the MEP problem dealt with in [5,13] in the sense that the choice of the entropy function adopted is supported by the Boltzmann argument. We do not claim that this solution has a better fit with empirically derived patterns with respect to others but only that is derived in a consistent way. In this solution the two random variables ν and m are uncorrelated, therefore the energy and species constraints can be dealt with *separately* giving Equations (14) and (16) or *at the same time*. This is a well known property of MEP: if the constraints concern only the marginals of a joint probability, then the MEP solution is the product of two unidimensional distributions.

5. Analysis of the Application of MEP in M.E.T.E.

The starting point of M.E.T.E. theory is the following probability distribution in Equation (31) on $\mathcal{N} \times \mathcal{M}$ introduced in [5], formula (4c). Quoting [5], page 2702 above formula (1a) : “ $R(n, \epsilon)$ is the probability that if a species is picked at random from the species list, then it has abundance n , and if an individual is picked at random from *that species* with abundance n , then its metabolic rate is ϵ ” (is between $[\epsilon, \epsilon + d\epsilon]$ in the original continuum energy formulation, italics is ours)

$$R(n, \epsilon) = \phi(n)\Theta(\epsilon|n). \tag{31}$$

This a delicate point: we think that using “that species” makes the above definition logically inconsistent because if there are more than one species with abundance n it is impossible to know

which species is being picked and so the probability of picking an individual with a given metabolic rate is not uniquely defined. One can readily convince oneself of this by considering a toy model of the systems along the lines of example in Box 7.1 in [13], page. 144, but with at least two species having equal abundance and containing individuals with the same metabolic rate. The only way for us to be logically consistent is to substitute “that species” with “a species”. This is equivalent to picking an individual of given metabolic rate from the multispecies \mathcal{I}_n . In this paper we have thus intended the definition of $R(n, \epsilon)$ as amended in this way.

Note however that in the sequel of [5], below formula (4c), it is (correctly) written “a species”. Note also that the same ambiguity is present in the book [13], Chapter 7, pages 142 and 143. In our framework, we thus rephrase the definition as

$$R(n, \epsilon) = \phi(n)P(i \in \mathcal{I}_\epsilon | i \in \mathcal{I}_n) \tag{32}$$

and using the definition of conditional probability $P(A, B) = P(A)P(B|A)$ and (4), (22) also as

$$R(n, \epsilon) = \phi(n) \frac{P(i \in \mathcal{I}_\epsilon, i \in \mathcal{I}_n)}{P(i \in \mathcal{I}_n)} = \phi(n) \frac{P(n, \epsilon)}{\phi_I(n)} = \frac{N_0}{Sn} P(n, \epsilon) \tag{33}$$

Therefore, the probability distributions $P(n, \epsilon)$ and $R(n, \epsilon)$ are related by the invertible transformation g (we use the same symbol for the sake of simplicity)

$$P(n, \epsilon) = g_n(R(n, \epsilon)) = \frac{nS}{N_0} R(n, \epsilon).$$

Moreover, the information N_0, S, E_0 sets the following constraints on $R(n, \epsilon)$, see (1a)–(1c) in [5]

$$\sum_{n, \epsilon} nR(n, \epsilon) = \frac{N_0}{S} \tag{34}$$

$$\sum_{n, \epsilon} \epsilon nR(n, \epsilon) = \frac{E_0}{S} \tag{35}$$

$$\sum_{n, \epsilon} R(n, \epsilon) = 1 \tag{36}$$

hence also $R(n, \epsilon)$ is properly normalized. In [5], the MEP procedure is applied as follows: the probability distribution $R(n, \epsilon)$ is the one that maximize the Shannon entropy

$$H(R) = - \sum_{n, \epsilon} R(n, \epsilon) \ln R(n, \epsilon) \tag{37}$$

on the set defined by Constraints (34)–(36) above. It is therefore the least informative probability distribution on the basis of the sole knowledge of the macroscopic ratios $E_0/S, N_0/S$. Note that, as observed in [5], if the previous ratios are known, E_0/N_0 is also known.

Standard application of the MEP gives the probability distribution depending on the unknown multipliers λ and β (see [5], formula (6))

$$R(n, \epsilon) = \frac{e^{-\lambda n - \beta \epsilon n}}{Z(\lambda, \beta)}. \tag{38}$$

The Lagrange multipliers λ and β have to be determined by inserting Equation (38) into the constraints in Equations (34) and (35) but an analytical solution of the resulting equations does not exist. Adopting some approximations in [5] we are lead to the following explicit formulae for the marginal (the SAD distribution)

$$\Phi(n) = \sum_{\epsilon} R(n, \epsilon) \approx \frac{1}{\ln \lambda^{-1}} \frac{e^{-\lambda n}}{n} \tag{39}$$

which is the celebrated Fisher log-series ([20]) and to

$$\Psi(\epsilon) = \frac{S}{N_0} \sum_n R(n, \epsilon) \approx \lambda \beta \frac{e^{-\lambda - \beta \epsilon}}{(1 - e^{-\lambda - \beta \epsilon})^2} \tag{40}$$

for the energy marginal. Moreover, the conditional probability is

$$\Theta(\epsilon|n) \approx \beta n e^{-\beta n \epsilon}. \tag{41}$$

What is striking in the above result, is that it prescribes a non zero correlation (Equation (41)) between the distribution of energy among individuals and the distribution of abundance among species, even if in the initial information nothing seems to hint at a possible correlation. This is for us the signal of a flaw in the above application of MEP.

The Correct Form of Entropy Function of M.E.T.E. Problem

Our aim now is to derive the form of the entropy function for the MEP problem in the $R(n, \epsilon)$ variable with constraints on energy and species in Equations (34)–(36) using the Boltzmann argument. Since $R(n, \epsilon)$ is not a joint probability distribution, we derive the entropy function for the distribution $P(n, \epsilon) = P(i \in \mathcal{I}_n, i \in \mathcal{I}_{\epsilon})$ in (22) which is related to $R(n, \epsilon)$ by the change of variable (33) and use the Proposition 1. The g -related constraints of Equations (34)–(36) for $P(n, \epsilon)$ are

$$\sum_{n, \epsilon} P(n, \epsilon) = 1, \quad \sum_{n, \epsilon} \frac{1}{n} P(n, \epsilon) = \frac{S}{N_0}, \quad \sum_{n, \epsilon} \epsilon P(n, \epsilon) = \frac{E_0}{N_0}.$$

Note that using $P(n, \epsilon)$ the constraints are only on the marginals of $P(n, \epsilon)$ hence unlike for $R(n, \epsilon)$ the MEP solution is the product of the marginals, see (48) below. Given the occupation vector $\mathbf{i} = (i_1, \dots, i_N)$ of the N_0 individuals in the N multispecies, for every n we have to distribute i_n individuals among the M energy groups \mathcal{I}_{ϵ} and each arrangement $\mathbf{i}_n = (i_{n,1}, \dots, i_{n,M})$ with $i_n = \sum_{\epsilon} i_{n,\epsilon}$ has probability

$$W_n(\mathbf{i}_n) = \frac{1}{M^{i_n}} \frac{i_n!}{i_{n,1}! \dots i_{n,M}!} \tag{42}$$

Therefore, taking into account that we have i_n/n independent tosses in the bin \mathcal{I}_n , we have to use the multiplicity factor $W(\mathbf{i})$ in Equation (19)

$$W = W(\mathbf{i}) \prod_{n=1}^N W_n(\mathbf{i}_n) \tag{43}$$

hence

$$\ln W = \ln W(\mathbf{i}) + \ln \prod_{n=1}^N W_n(\mathbf{i}_n). \tag{44}$$

The maximum of $\ln W$ is reached when both terms in the right hand side are maximized, but the first term $\ln W(\mathbf{i})$ is independent of the arrangement in the energy levels. Therefore we can maximize the second term in the right hand side of Equation (44) for fixed values of \mathbf{i} and then maximize the first term with respect to \mathbf{i} . Now

$$\ln \prod_{n=1}^N W_n(\mathbf{i}_n) = \sum_n \ln W_n(\mathbf{i}_n) = - \sum_{n, \epsilon} i_{n,\epsilon} \ln i_{n,\epsilon} + f(\mathbf{i}) \tag{45}$$

is to be maximized under the $N + 1$ constraints (note that the constraint $\sum_{n,\epsilon} i_{n,\epsilon} = N_0$ is already enforced since $\sum_n i_n = N_0$)

$$\sum_{n,\epsilon} \epsilon i_{n,\epsilon} = E_0, \quad \text{and} \quad \sum_{\epsilon} i_{n,\epsilon} = i_n, \quad n = 1, \dots, N. \quad (46)$$

The unconstrained extremum problem for the Lagrange function

$$G = - \sum_{n,\epsilon} i_{n,\epsilon} \ln i_{n,\epsilon} - \beta (\sum_{n,\epsilon} \epsilon i_{n,\epsilon} - E_0) - \sum_n \mu_n (\sum_{\epsilon} i_{n,\epsilon} - i_n) \quad (47)$$

gives the solution

$$i_{n,\epsilon} = i_n \frac{e^{-\beta\epsilon}}{\sum_n e^{-\beta\epsilon}} = i_n \frac{e^{-\beta\epsilon}}{Z(\beta)}.$$

Since $i_{n,\epsilon} = P(n, \epsilon)N_0$ and $i_n = \phi_{\mathcal{I}}(n)N_0$ we have

$$P(n, \epsilon) = \phi_{\mathcal{I}}(n) \frac{e^{-\beta\epsilon}}{Z(\beta)} \quad (48)$$

and to get the complete solution we have to maximize $W(\mathbf{i})$, which has already been dealt with in the previous section, Equations (19) and (20). Therefore, the MEP procedure for the distribution $P(n, \epsilon)$ with the constraints gives the solution

$$P(n, \epsilon) = \frac{nS}{N_0} \frac{e^{-\lambda n}}{Z(\lambda)} \frac{e^{-\beta\epsilon}}{Z(\beta)}. \quad (49)$$

Using the change of variables in Equation (33) and Proposition 1, we find the solution of the MEP problem in [5] with respect to the variables $R(n, \epsilon)$

$$R(n, \epsilon) = \frac{N_0}{Sn} P(n, \epsilon) = \frac{e^{-\lambda n}}{Z(\lambda)} \frac{e^{-\beta\epsilon}}{Z(\beta)} \quad (50)$$

which coincides with $Q(n, \epsilon)$ in (30).

6. Discussion

A non trivial task in the application of the MEP is the translation in mathematical terms (i.e., as constraints on the sought probability distribution) of the information available on the system. In some cases one may hesitate between mathematically equivalent formalizations. What is not so well known is that the use of Shannon entropy $H(p)$ may not be justified in all cases. We have provided a criterion based on Boltzmann method of the most probable occupation vector to derive the correct form of the entropy function in a given formulation. Note that the problem addressed here is different from the search of a suitable prior distribution for the relative entropy $D(p|q)$. In the second part of the paper we have applied this analysis to critically examine the Maximum Entropy Theory of Ecology (M.E.T.E.). Even if the application of MEP contained in M.E.T.E. seems flawed, the resulting SAD proposed by M.E.T.E. is the Fisher log-series, which is widely accepted in the ecological community and considered as giving a best fit for many ecological communities (although the use of log-series SAD has been questioned recently, see [21]). Therefore, the previsions of species abundance based on M.E.T.E. may well be in good agreement with field data. What seems to be hardly giving a sound base in M.E.T.E. is the claim that the information contained in N_0, S, E_0 produces a coupling between the distribution of abundance between species and the distribution of metabolic energy between individuals.

Acknowledgments: We thank A. Maritan, S. Suweis, A. Tovo, M. Formentin and M. Pavon for fruitful discussions on the subject of this paper.

Conflicts of Interest: The authors declare no conflict of interest

Appendix A

Proof of Proposition 1. To seek for the constrained extrema of $H(p)$ over the set $f(p) - c = 0$, $c \in \mathbb{R}^k$, we introduce the Lagrange function

$$G(p, \lambda) = H(p) + \sum_{\alpha=1}^k \lambda_{\alpha} (f_{\alpha}(p) - c_{\alpha})$$

\hat{p} is a constrained extremum if and only if for all i

$$(\nabla_p G(\hat{p}, \lambda))_i = (\nabla H(\hat{p}))_i + \sum_{\alpha=1}^k \lambda_{\alpha} \frac{\partial f_{\alpha}(\hat{p})}{\partial p_i} = (\nabla H(\hat{p}) + df^T(\hat{p})\lambda)_i = 0.$$

Let $H'(p') = H(g(p'))$ and $f'(p') = f(g(p'))$ where $p = g(p')$ is invertible and

$$G'(p', \mu) = H'(p') + \sum_{\alpha=1}^k \mu_{\alpha} (f'_{\alpha}(p') - c_{\alpha}).$$

p' is a constrained extremum of G' over the set $f'(p') = c$ if and only if for all i

$$(\nabla_{p'} G(p', \mu))_i = (\nabla H'(p'))_i + \sum_{\alpha=1}^k \mu_{\alpha} \frac{\partial f'_{\alpha}(p')}{\partial p'_i} = 0$$

that is

$$\sum_j \frac{\partial H(g(p'))}{\partial p_j} \frac{\partial g_j(p')}{\partial p'_i} + \sum_{j,\alpha} \mu_{\alpha} \frac{\partial f_{\alpha}(g(p'))}{\partial p_j} \frac{\partial g_j(p')}{\partial p'_i} = 0$$

or, in compact notation

$$dg^T(p') (\nabla_p H(g(p')) + df^T(g(p'))\mu) = 0$$

Since g is invertible, $\det g \neq 0$, and $\nabla_p H(g(p')) + df^T(g(p'))\mu = 0$ if and only if $\hat{p} = g(p')$ is a constrained extremum for $H(p)$ and $\lambda = \mu$. \square

References

1. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620.
2. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
3. Banavar, J.R.; Maritan, A.; Volkov, I. Applications of the principle of maximum entropy: From physics to ecology. *J. Phys.* **2010**, *22*, 063101.
4. Haegeman, B.; Rampal, S.E. Entropy maximization and the spatial distribution of species. *Am. Nat.* **2010**, *175*, E74–E90.
5. Harte, J.; Zillio, T.; Conlinsk, E.; Smith, A.B. Maximum entropy and the state-variable approach to macroecology. *Ecology* **2008**, *89*, 2700–2711.
6. Dewar, R.C.; Porte, A. Statistical mechanics unifies different ecological patterns. *J. Theor. Biol.* **2008**, *251*, 389–403.
7. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259.
8. Pueyo, S.; He, F.; Zillio, T. The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol. Lett.* **2007**, *10*, 1017–1028.
9. Frank, S.A. Measurement scale in maximum entropy models of species abundance. *J. Evolut. Biol.* **2011**, *24*, 485–496.
10. Haegeman, B.; Loreau, M. Limitations of entropy maximization in ecology. *Oikos* **2008**, *117*, 1700–1710.

11. Shipley, B. Limitations of entropy maximization in ecology: A reply to Haegeman and Loreau. *Oikos* **2009**, *118*, 152–159.
12. Haegeman, B.; Loreau, M. Trivial and non trivial applications of entropy maximization in ecology: A reply to Shipley. *Oikos* **2009**, *118*, 1270–1278.
13. Harte, J. *Maximum Entropy and Ecology. A Theory of Abundance, Distribution, and Energetics*; Oxford University Press: Oxford, UK, 2011.
14. Harte, J.; Newman, A. Maximum information entropy: A foundation for ecological theory. *Trends Ecol. Evol.* **2014**, *29*, 384–389.
15. White, P.E.; Thibault, K.M.; Xiao, X. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* **2012**, *93*, 1772–1778.
16. McGlinn, D.J.; Xiao, X.; Kitzes, J.; White, E.P. Exploring the spatially explicit predictions of the Maximum Entropy Theory of Ecology. *Glob. Ecol. Biogeogr.* **2015**, *24*, 675–684.
17. Xiao, X.; McGlinn, D.J.; White, E.P. A strong test of the maximum entropy theory of ecology. *Am. Nat.* **2015**, *185*, E70–E80.
18. Shore, J.; Johnson, R. Axiomatic derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37.
19. Schrodinger, E. *Statistical Thermodynamics*; Courier Corporation: North Chelmsford, MA, USA, 1989.
20. Fisher, R.A.; Corbet, A.S.; Williams, C.B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **1943**, *12*, 42–58.
21. Tovo, A.; Suweis, S.; Formentin, M.; Favretti, M.; Volkov, I.; Banavar, J.R.; Maritan, A. Upscaling species richness and abundances in tropical forests. *Sci. Adv.* **2017**, *3*, e1701438.



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).