Discussing an Expected Utility and Weighted Entropy Framework

José Pinto Casquilho
Post-Graduate and Research Program, Universidade Nacional Timor Lorosa’e, Dili, Timor Leste (East Timor)
Email: josecasquilho@gmail.com

Received 25 December 2013; revised 25 January 2014; accepted 1 February 2014

Abstract

In this paper, it is discussed a framework combining traditional expected utility and weighted entropy (EU-WE)—also named mean contributive value index—which may be conceived as a decision aiding procedure, or a heuristic device generating compositional scenarios, based on information theory concepts, namely weighted entropy. New proofs concerning the maximum value of the index and the evaluation of optimal proportions are outlined, with emphasis on the optimal value of the Lagrange multiplier and its meaning. The rationale is a procedure of maximizing the combined value of a system expressed as a mosaic, denoted by characteristic values of the states and their proportions. Other perspectives of application of this EU-WE framework are suggested.

Keywords

Mosaic Composition, Expected Utility, Weighted Entropy, Mean Contributive Value

1. Introduction

Shannon entropy has been widely used in ecological studies as a measure of diversity at different scales in space, from local community level to landscapes and regions. Guiasu and Guiasu [1] [2] provide an extensive literature review covering the properties and applications of these measures of diversity in general, and, concerning landscape metrics, there are also several reviews of the theme, (e.g., [3]-[5]), Shannon entropy being recommended for landscape management as it is considered an index sensitive to the presence of rare habitats [6].

The core of the rationale that relates utility and information theory concepts can be summarized as it was stated by Bernardo [7], recognizing the decision problem underlying a problem of statistical inference: expected information may be conceived as expected utility.

1.1. Statistical Entropy and Diversity

Under the scope of a general theory of communication, Shannon [8] defined entropy as a measure of how much
choice is involved in the selection of an event, or of how uncertain we are of the outcome, and settled the formula \( H = -K \sum_{i=1}^{n} p_i \log p_i \) where the constant \( K > 0 \) merely amounts to a choice of a unit of measure. He retrieved previous work of Hartley [9] outlining a quantitative measure whereby the capacities of various systems to transmit information might be compared, taking a measure of information as the logarithm of the number of possible symbol sequences. Jaynes [10] wrote that entropy as a concept may be regarded as a measure of our degree of ignorance as to the state of a system and that quantity \( H \) measures in a unique way the amount of uncertainty represented by a probability distribution. Also, entropy is claimed to be a measure of the average randomness of a stochastic system [11] and it is referred to be the only meaningful functional for measuring uncertainty and information in probability theory [12]. The information value of an event is defined as \(-\log p_i\), so formula \( H \) denotes the mean information value of a sample space, related to the coexistence of a multi-state system or a mosaic.

Rényi [13] generalized Shannon entropy as a 1-parameter functional family \( H_q = \log \left( \sum_{i=1}^{n} p_i^q \right) / (1-q) \), under the scope of measures of entropy and information defined on the set of generalized probability distributions, entailing that \( \lim_{q \to 0} H_q = H \). Rényi generalized entropy function has been recently referred to as a continuum of diversity measures [14]. Hill [15] proved that the exponential form of Rényi entropies \( N_q = \exp H_q \)—he called diversity numbers—has an immediate connection with diversity indices used in Ecology: the richness in species, the exponential form of Shannon entropy and the inverse of Simpson’s index.

1.2. Weighted Entropy and Utility

Weighted entropy was first proposed by Bellis and Guiasu [16] taking into account the two basic concepts of objective probability and subjective utility, thus defining the information supplied by the event \( E_i \) with a probability \( p_i \) and an utility \( u_i \)—the last meaning the value of an outcome relative to a specified goal—with the formula \( I(u_i, p_i) = -k u_i \log p_i \), and \( k > 0 \). Guiasu [17] derived the principle of maximum information obtaining the probability distribution maximizing weighted entropy—he later called useful entropy [18]—and Aggarwal and Picard [19] settled a general overview of information measures with preference, the preference of an event being defined as the product of its probability and utility. Several applications with weighted entropy were performed in the middle eighties: for instance, Batty [20] used weighted entropy to discuss the spatial pattern of aggregation in cities, while Nawrocki and Harding [21] used state-value weighted entropy as a measure of investment risk; Taneja and Tuteja [22] extended the concept to derive the characterization of a quantitative-qualitative measure of inaccuracy. Later, Guiasu and Guiasu [23] revised the theme under ecology analysis, noting that, whenever measuring the diversity of ecosystems, additional information—such as absolute abundance, economic significance or ecological importance of species—has to be taken into account, reflected in the weights, a concept that was further extended to joint weighted entropy related to the joint probability distribution assigned to pairs of species [24].

Casquilho et al. [25] derived independently the main results concerning weighted entropy, under a 1-parameter generalization of Shannon formula focused on an ecological and economic application at the landscape level, from which followed the EU-WE framework here discussed—weighted entropy was then named mean informative value index and EU-WE framework was defined as mean contributive value index. These results were applied to discuss compositional scenarios of forest ecomosaics [26]-[28], with a non-linear utility scope where the concept of contributive value plays a central role: contributive value is a relational form of value, it is the value that some part confers on the whole of which it is a part, because this contribution is conditioned by the presence and extent of other parts [29], so emphasizing that the contributive value of a part should not be confused with the value that this part has on its own, independently from the context [30]. The value that a part has on its own was referred to as a characteristic, or intrinsic, value.

Ricotta [31] mentioned weighted entropy as a contribute towards bridging the gap between ecological diversity indices and measures of biodiversity and Allen et al. [32] used a related, unconstrained, form of weighted entropy under the scope of phylogenetic measures.

The work presented here has some similarity with a decision aiding procedure based on expected utility and Shannon entropy [33], though here we use weighted entropy. The objectives of this paper include proving and discussing the mathematical properties of the optimal solution and providing a critical analysis of an expected utility and weighted entropy framework (EU-WE) as a conceptual device generating relative compositional sce-
scenarios of mosaics based on optimality criteria.

2. Methodology

In what follows a new set of the results and proofs are presented, equivalent, but different, from those presented before, e.g., [25], now indexed to the optimal value of the Lagrange multiplier, which allows for an insightful interpretation, anchoring the optimal solution of this EU-WE framework with weighted entropy.

2.1. Index $K_U$

We will be dealing with proportions, defining a normalized measure space. Proportions are relative extension measures—as well as relative frequencies and probabilities—the difference is that proportions reflect the extension of actual, or presumably effective, states of a system, and probabilities are possibility measures of events compatible with Kolmogorov’s axiomatic definition. Nevertheless, the two concepts are intimately linked under the scope of objective or physical probabilities, which often uses probability practically as a synonym for proportion [34]. Anscombe and Aumann [35] pointed out that physical probabilities can be determined empirically by noting the proportion of successes in some trials. Either as proportions, relative frequencies, or probabilities, these real numbers denote the same mathematical object, a simplex, with different connotations or semantic injunctions, depending on the context.

Assume that a system is characterized by a scenario of the world defined as the set of $n$ elementary states, or sample space: $S = \{s_i\}_{i=1}^n$; also assume that $w_i > 0$ is a real positive number denoting a characteristic value of the state of the system applying the injective function $W(s_i) = w_i$; the power set $\Omega = \wp(S)$ is the set of events; members of the collection generated by $S$; elementary states occur linked to the discrete distribution denoted as $p = \{p_i\}_{i=1}^n$ with $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$ defining a $n-1$ simplex. The structure $(S, \Omega, p)$ is a normalized measure space, with the condition $p_j = 0$ meaning the absence of the indexed $j$ state.

Next we outline an information based family of utility functions:

$$U_i = u_i \left(1 - \log p_i\right) \quad \text{for} \quad i = 1, \ldots, n; \quad \text{symbol} \ U \ \text{defines a monotone increasing transformation of the original characteristic values such as} \ u_i = U\left(w_i\right) \ \text{with} \ u_i > 0. \ \text{We can assign proportions or probabilities} \ \Pr[Y = y_i] = p_i \ \text{for} \ i = 1, \ldots, n \ \text{and the utility functions here defined are the product of utilities and context values}, \ y_i = u_i c_i \ \text{where the context value is expressed by} \ c_i = 1 - \log p_i \ \text{with range} \ 1 \leq c_i < +\infty, \ \text{and the term} \ -\log p_i \ \text{is the information value relative to elementary state} \ i. \ \text{Context values} \ c_i \ \text{increase with correspondent information values, meaning that scarcity is considered as if it entails more relevance, which is the case, for instance, with the methodology outlined by Haddock et al. [36], where the use of scarcity weights to assess impacts in landscape changes reflect the analogy that “endangered habitats” have a similar status of endangered species.}

The EU-WE framework here to be discussed, denoted index $K_U$, is then defined using the expected value operator $E[\cdot]$, evaluating the weighted average of the values $y_i$, with the suitable decomposition

$$E[Y] = \sum_{i=1}^n u_i p_i - \sum_{i=1}^n u_i \log p_i,$$

rewritten in Equation (1) as

$$K_U = \bar{U} + H_U \quad (1)$$

where $H_U$ means the weighted entropy of the utilities $u_i$ and $\bar{U}$ is traditional expected utility. Obviously, denoting $h_i = p_i \left(1 - \log p_i\right)$ we may rewrite (1) as $K_U = \sum_{i=1}^n u_i h_i$ which may be interpreted as a nonlinear-expected utility because the terms $h_i$ aren’t additive proportions or probabilities: although they verify the unitary hypercube condition, as we have that $\lim_{p_i \to 0^+} h_i = 0$ thus allowing for the extension $0 \leq h_i \leq 1$, we also see that the sum $\sum_{i=1}^n h_i$ equates $1 + H$, with $H$ denoting Shannon entropy.

2.2. Optimal Proportions and Maximum Value of Index $K_U$

Building auxiliary Lagrange function defined as $L = \sum_{i=1}^n u_i p_i \left(1 - \log p_i\right) - \alpha \left(\sum_{i=1}^n p_i - 1\right)$ we can find the partial derivatives evaluated as $\partial L / \partial p_i = -u_i \log p_i - \alpha$ for $i = 1, \ldots, n$, from what follows, solving the equations...
\[ \frac{\partial L}{\partial p_i} = 0, \] 
the results: \( \alpha = -u_i \log p_i > 0 \) if \( p_i \in ]0,1[ \) and \( p_i = \exp(-\alpha/u_i) \) with the closure condition \( \frac{\partial L}{\partial \alpha} = 0 \) re-expressed as \( \sum_{i=1}^{n} \exp(-\alpha/u_i) = 1 \).

The numeric solution of this equation will be denoted \( \alpha^{*} \) and we can prove it is unique, following a corollary of Bolzano theorem. If we define \( f(\alpha) = \sum_{i=1}^{n} \exp(-\alpha/u_i) - 1 \) we observe that \( f(\alpha) \) has a strictly negative derivative computed as \( f'(\alpha) = -\sum_{i=1}^{n} \left( \frac{1}{u_i} \exp(-\alpha/u_i) \right) < 0 \) hence \( f(\alpha) \) is strictly decreasing; the following calculus of limits confirm the existence and the uniqueness of the solution: \( \lim_{\alpha \to 0} f(\alpha) = n-1 > 0 \) if \( n \geq 2 \) and \( \lim_{\alpha \to \infty} f(\alpha) = -1 \), so there is only one real value such that \( f(\alpha) = 0 \).

Thus, the critical point of index \( K_U \) has coordinates defined by (2):

\[ p_i^* = \exp(-\alpha^*/u_i) \quad \text{for} \quad i = 1, \ldots, n \] (2)

As expression (2) depends on the value of \( \alpha^* \) first we have to solve numerically the equation that defines the closure condition in the simplex, what implies solving Equation (3) for \( \alpha^* \)

\[ \alpha^* : \sum_{i=1}^{n} \exp(-\alpha/u_i) - 1 = 0 \] (3)

It can also be proven that each optimal coordinate \( p_i^* \) increases with the correspondent characteristic value, or utility, and decreases when the other utilities increase [26] [27], as it should be expectable, lying within the open interval \( ]0,1[ \).

Next, let us prove that the critical point is a maximum in analogy with Guiasu procedure [17]. We build the auxiliary function \( A = K_U - \alpha = \sum_{i=1}^{n} u_i p_i \left( 1 - \log p_i \right) - \alpha \left( \sum_{i=1}^{n} p_i \right) \) and rearrange the terms obtaining the sequence of equalities:

\[ \sum_{i=1}^{n} p_i \left( u_i - u_i \log p_i - \alpha \right) = \sum_{i=1}^{n} p_i \left( 1 - \log p_i - \alpha/u_i \right) = \sum_{i=1}^{n} u_i p_i \left( 1 - \log \left( p_i \exp(\alpha/u_i) \right) \right) \]

and eventually get the equivalent mathematical expression for the auxiliary function

\[ A = \sum_{i=1}^{n} u_i \exp(-\alpha/u_i) p_i \exp(\alpha/u_i) \left( 1 - \log \left( p_i \exp(\alpha/u_i) \right) \right). \]

Using the auxiliary result \( f(x_i) = x_i \left( 1 - \log x_i \right) \leq 1 \) if \( x_i > 0 \) we have that the maximum point is located at \( x_i = 1 \) with the maximum value becoming \( f(1) = 1 \); then, replacing \( x_i = p_i \exp(\alpha/u_i) \) we conclude that inequality \( A \leq \sum_{i=1}^{n} u_i \exp(-\alpha/u_i) \) holds, and the maximum value is \( A^* = \sum_{i=1}^{n} u_i \exp(-\alpha^*/u_i) \) which is reached if and only if we have the replacement \( p_i^* = \exp(-\alpha^*/u_i) \), thus verifying the result (2) as the maximum point coordinates.

### 2.3. The Optimal Value of the Lagrange Multiplier of Index \( K_U \)

Retrieving auxiliary function \( A \) we can rewrite: \( A' = K_U - \alpha' = \sum_{i=1}^{n} u_i p_i^* \) and the optimal value of the Lagrange multiplier evaluates as \( \alpha^* = K_U - \sum_{i=1}^{n} u_i p_i^* = -\sum_{i=1}^{n} u_i p_i^* \log p_i^* \) entailing the result:

\[ \alpha^* = H_U \left( \alpha^* \right) \] (4)

Formula (4) is the weighted entropy of the optimal solution of index \( K_U \) defined as \( \alpha^* = \left( p_1^*, \ldots, p_n^* \right) \) evaluated by Formula (2) after numerical evaluation of Equation (3); note that this is not to be confused with the maximum point of weighted entropy itself.

Optimal proportions are indifferent to a linear positive transformation in the utilities, such as a change of scale or units of measure. In particular, if we replace the utilities \( u_i \) by its normalized (and dimensionless) form \( v_i = u_i / \sum u_i \) —acknowledging that the \( v_i \)’s are intrinsically merged within another simplex—the optimal solutions \( p_i \) are the same as those evaluated with the original utilities \( u_i \), and the optimal Lagrange multiplier value becomes also normalized as \( \alpha^* = \alpha^*/\sum u_i \).
2.4. Range of Index $K_U$

The minimum value of index $K_U$ is $\min K_U = \min \{u_i\}_{i=1,...,n} = u_m$ which is a straightforward result because weighted entropy verifies $H_U \geq 0$; as $K_U = \bar{U} + H_U$ defined in (1) the second term of the sum vanishes at the vertexes of the simplex and the first term takes the minimum value when $\bar{U} = u_m$ so the result holds. The maximum value of the index $\max K_U$ may be calculated with optimal proportions by direct substitution as $K_U^* = \sum_{i=1}^{n} u_i p_i^* - \sum_{i=1}^{n} u_i \log p_i^*$ or, equivalently, indexed to the optimal value of the Lagrange multiplier evaluated in (3), obtaining the expression

$$K_U^* = \sum_{i=1}^{n} u_i \exp\left(-\alpha^*/u_i\right)\left(1+\alpha^*/u_i\right)$$

(5)

Thus, since $K_U$ is a continuous function in the proportions $\{p_i\}_{i=1,...,n}$, defined in a compact set—the simplex—we have the general result, following Bolzano-Weierstrass theorem, concerning the range of the EU-WE here discussed: $u_m \leq K_U \leq K_U^*$, the last term of the double inequality referred to in Equation (5).

2.5. Exemplification

As an example, we retrieve characteristic economic values of forest habitats from [27] and exemplify the use of the formulas presented in this paper. In the case, we have $W = \{112, 618, 91, 136, 191\}$ and with neutral utilities of the form $U(w_i) = w_i$ for $i = 1, \cdots, 5$ we obtain for Equation (3) the following expression:

$$\alpha^* \quad e^{\alpha^*} + e^{\frac{\alpha^*}{618}} + e^{\frac{\alpha^*}{91}} + e^{\frac{\alpha^*}{136}} + e^{\frac{\alpha^*}{191}} = 1,$$

which solved numerically for $\alpha^*$ gives the value $\alpha^* = 310.31$. Then, applying Formula (2), rounding up to three decimals, we get: $p_1^* = 0.063$; $p_2^* = 0.605$; $p_3^* = 0.033$; $p_4^* = 0.102$ and $p_5^* = 0.197$.

Using these optimal values and evaluating Formula (4) we obtain 310.43 which is quite similar to the value of $\alpha^*$ evaluated numerically, thus confirming that the optimal value of the Lagrange multiplier is the weighted entropy of the optimal point of index $K_U$. Finally, computing (5) we get the maximum value of the index: $K_U^* = 745.88$.

Other utilities could have been used besides the neutral, either convex, risk-taking utilities, as it would be the case with $U(w_i) = w_i^2$ or concave, risk-averting utilities, such as $U(w_i) = \sqrt{w_i}$; the first type would promote the proportions related to higher characteristic values penalizing the remnant, and the second type, on the contrary, would enhance a more balanced optimal composition solution.

3. Discussion

The EU-WE framework here discussed emphasizes the notion of contributive value of each component of a mosaic—or stable state of a simultaneous multi-state system—depending both on context and utility values. Others seem to identify contributive value with utility itself (see [37]), which is not the case here, where traditional expected utility is balanced by the weighted entropy of the utilities. We remark that weighted entropy—or related forms of Shannon entropy—though it has already a long history of almost half a century since its first reference [16], is yet being referred to as a risk measure in portfolio selection strategies (e.g. [38] [39]).

The static nonlinear optimization procedure presented here may be applied with focus on compositional scenarios generated under active adaptive ecosystem management paradigm sensu Gunderson et al. [40], a reinforcement process, or as a dynamic system: if the characteristic values, or the correspondent utilities, change in time, so will change the optimal proportions, addressing the issue of whether the mosaic will after some time converge to a steady state or be continuously changing (see [41]). Clark et al. [42] point out that utility analysis has been used in environmental policy design studies to help articulate conflicting experiences and simplify comparisons of policies, though it is also useful remember that the precautionary principle advises that resilience of an ecosystem may be lost because of activities that focus on an optimal control strategy of a single target variable [43]. Last, let us recall that decision analysis concerns with the balancing of factors that influence a decision, in a procedure that incorporates uncertainties, values, and preferences, in a basic structure that models the decision [44], and recently it was proposed a semiotic interpretation of weighted entropy as mean contextual re-
levance of events indexed to a sample space and a context of utility [45].

References


