



A Triple Uniqueness of the Maximum Entropy Approach

Jürgen Landes^(✉) 

Munich Center for Mathematical Philosophy, LMU Munich, Munich, Germany
juergen_landes@yahoo.de

To be uncertain is to be
uncomfortable, but to be certain
is to be ridiculous.

Chinese proverb

Abstract. Inductive logic is concerned with assigning probabilities to sentences given probabilistic constraints. The Maximum Entropy Approach to inductive logic I here consider assigns probabilities to all sentences of a first order predicate logic. This assignment is built on an application of the Maximum Entropy Principle, which requires that probabilities for uncertain inference have maximal Shannon Entropy. This paper puts forward two different modified applications of this principle to first order predicate logic and shows that the original and the two modified applications agree in many cases. A third promising modification is studied and rejected.

Keywords: Inductive logic · Maximum entropy · First order predicate language · Uncertain inference · Objective Bayesianism

1 Introduction

Inductive logic is a formal approach to model rational uncertain inferences. It seeks to analyse the degree to which premisses entail putative conclusions. Given uncertain premisses $\varphi_1, \dots, \varphi_k$ with attached uncertainties X_1, \dots, X_k an inductive logic provides means to attach uncertainty Y to a conclusion ψ , where the X_i and Y are non-empty subsets of the unit interval. An inductive logic can be represented as

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \psi^Y, \quad (1)$$

where \approx denotes an inductive entailment relation [11]. Much work has gone into the development and exploration of inductive logics, see, e.g., [4, 9, 12, 16, 29–31, 41].

I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 432308570 and 405961989.

© Springer Nature Switzerland AG 2021

J. Vejnárová and N. Wilson (Eds.): ECSQARU 2021, LNAI 12897, pp. 644–656, 2021.

https://doi.org/10.1007/978-3-030-86772-0_46

The main early proponent of inductive logic was Rudolf Carnap [2, 3]. Nowadays, the spirit of his approach today continues in the Pure Inductive Logic approach [14, 17, 21–24, 40]. In this paper, I however consider uncertain inference within the maximum entropy framework, which goes back to Edwin Jaynes [15], who put forward a Maximum Entropy Principle governing rational uncertain inference.

Maximum Entropy Principle. *Rational agents ought to use a probability function consistent with the evidence for drawing uncertain inferences. In case there is more than one such probability function, a rational agent ought to use probability functions with maximal entropy.*

In case only a single probability function is consistent with the evidence, the Maximum Entropy Principle is uncontroversial. Its strength (and sometimes controversial nature) is rooted in applications with multiple prima facie reasonable probability functions for probabilistic inference. This principle is at the heart of objective Bayesianism.

If the underlying domain is finite, then applying the Maximum Entropy Principle for inductive entailment is straight-forward and well-understood due to the seminal work of Alena Vencovská & Jeff Paris [32–34, 36–39]. Matters change dramatically for infinite domains. Naively replacing the sum by an integral in the definition of Shannon Entropy produces a great number of probability functions with infinite entropy. But then there is no way to pick a probability function with maximal entropy out of a set in which all functions have infinite entropy.

There are two different suggestions for inductive logic on an infinite first order predicate logic explicating the Maximum Entropy Principle. The Entropy Limit Approach [1, 35] and the Maximum Entropy Approach [28, 45–48]. It has been conjectured, that both approaches agree in cases in which the former approach is well defined [48, p. 191]. This conjecture has been shown to hold in a number of cases of evidence bases with relatively low quantifier-complexity [19, 25, 44].

This paper introduces modifications of the Maximum Entropy Approach and studies their relationships. I next properly introduce this approach along with some notation and the modifications. I then proceed to investigate their relationships. My main result is Theorem 1; it proves that the two suggested modifications agree with the original Maximum Entropy Approach expounded by Jon Williamson for convex optimisation problems, if at least one of these three approaches yields a unique probability function for inference on the underlying first order predicate language.

In Sect. 4, I study a third modification of Williamson’s Maximum Entropy Approach, which I reject due to the absurd probabilities it delivers for inductive inference, see Proposition 6. In Sect. 5, I put forward some concluding remarks and consider avenues for future research.

2 The Maximum Entropy Approach and Two Modifications

The formal framework and notation is adapted from [25].

A fixed first order predicate language \mathcal{L} is given. It consists of countably many constant symbols t_1, t_2, \dots exhausting the universe (for every element in the universe there is at least one constant symbol representing it) and finitely many relation symbols, U_1, \dots, U_n . In particular, note that the language does not contain a symbol for equality nor does it contain function symbols. The atomic sentences are sentences of the form $U_i t_{i_1} \dots t_{i_k}$, where k is the arity of the relation U_i ; the atomic sentences are denoted by a_1, a_2, \dots . They are by construction ordered in such a way that atomic sentences involving only constants among t_1, \dots, t_n occur before those atomic sentences that also involve t_{n+1} . The set of sentences of \mathcal{L} is denoted by $S\mathcal{L}$.

The finite sublanguages \mathcal{L}_n of \mathcal{L} are those languages, which only contain the first n constant symbols t_1, \dots, t_n and the same relation symbols as \mathcal{L} . Denote the sentences of \mathcal{L}_n by $S\mathcal{L}_n$.

The contingent conjunctions of maximal length of the form $\pm a_1 \wedge \dots \wedge \pm a_{r_n} \in S\mathcal{L}_n$ are called the n -states. Let Ω_n be the set of n -states for each $n \in \mathbb{N}$ with $|\Omega_n| = 2^{r_n}$.

Definition 1 (Probabilities on Predicate Languages). A probability function P on \mathcal{L} is a function $P : S\mathcal{L} \rightarrow \mathbb{R}_{\geq 0}$ such that:

- P1:** If τ is a tautology, i.e., $\models \tau$, then $P(\tau) = 1$.
- P2:** If θ and φ are mutually exclusive, i.e., $\models \neg(\theta \wedge \varphi)$, then $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$.
- P3:** $P(\exists x \theta(x)) = \sup_m P(\bigvee_{i=1}^m \theta(t_i))$.

A probability function on \mathcal{L}_n is defined similarly (the supremum in P3 is dropped and m is equal to n).

\mathbb{P} denotes the set of all probability functions on \mathcal{L} .

A probability function $P \in \mathbb{P}$ is determined by the values it gives to the quantifier-free sentences, this result is known as *Gaifman's Theorem* [8]. P3 is sometimes called *Gaifman Condition* [40, p. 11]. The Gaifman Condition is justified by the assumption that the constants exhaust the universe. Consequently, a probability function is determined by the values it gives to the n -states, for each n [33, p. 171].

It is thus sensible to measure the entropy of a probability function $P \in \mathbb{P}$ via n -states with varying n .

Definition 2 (n -entropy). The n -entropy of a probability function $P \in \mathbb{P}$ is defined as:

$$H_n(P) := - \sum_{\omega \in \Omega_n} P(\omega) \log P(\omega) .$$

The usual conventions are $0 \log 0 := 0$ and \log denoting the natural logarithm. The second convention is inconsequential for current purposes.

$H_n(\cdot)$ is a strictly concave function.

The key idea is to combine the n -entropies defined on finite sublanguages \mathcal{L}_n into an overall notion of comparative entropy comparing probability functions P and Q defined on the entire first order language.

So far, the literature has only studied such inductive logics with respect to the first binary relation in the following definition.

Definition 3 (Comparative Notions of Entropy). *That a probability function $P \in \mathbb{P}$ has greater (or equal) entropy than a probability function $Q \in \mathbb{P}$ could be defined in the following three ways.*

1. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$.*
2. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) \geq H_n(Q)$ and there are infinitely many n such that $H_n(P) > H_n(Q)$, denoted by $P \rceil Q$.*
3. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) \geq H_n(Q)$, denoted by $P \rceil Q$.*

The lower two definitions are alternative ways in which one could explicate the intuitive idea of comparative entropy, which have never been studied before. Prima facie, all three definitions appear reasonable.

Before I can define notions of maximal entropy with respect to the given premisses, I need to specify the set of probability functions over which entropy is to be maximised.

Definition 4 (Region of Feasible Probability Functions). *The set of probability functions consistent with all premisses, $P(\varphi_i) \in X_i$ for all i , is denoted by \mathbb{E} and defined as*

$$\mathbb{E} := \{P \in \mathbb{P} : P(\varphi_i) \in X_i \text{ for all } 1 \leq i \leq k\} .$$

In order to simplify the notation, I do not display the dependence of \mathbb{E} on the premisses.

In this paper, I only consider a fixed set of premisses, $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$. There is hence no need to complicate notation by writing $\mathbb{E}_{\varphi_1^{X_1}, \dots, \varphi_k^{X_k}}$ or similar.

Definition 5 (Set of Maximum Entropy Functions). *The set of probability functions on \mathcal{L} with maximal entropy in \mathbb{E} relative to a notion of comparative entropy $>$ defined on $\mathbb{P} \times \mathbb{P}$ can then be defined as*

$$\text{maxent}_{>} \mathbb{E} := \{P \in \mathbb{E} : \text{there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q > P\} . \quad (2)$$

The Maximum Entropy Principle now compels agents to use probabilities in $\text{maxent}_{>} \mathbb{E}$ for drawing uncertain inferences as described by the scheme for inductive logic in (1). The induced inductive logics are described in the following definition.

Definition 6 (Maximum Entropy Inductive Logics). An inductive logic with respect to $>$ is induced by attaching uncertainty $Y_{>}(\psi) \subseteq [0, 1]$ to the sentences ψ of \mathcal{L} via

$$Y_{>}(\psi) := \{r \in [0, 1] \mid \text{there exists } P \in \text{maxent}_{>} \mathbb{E} \text{ with } P(\psi) = r\} .$$

In case there are two or more different probability functions in $\text{maxent}_{>} \mathbb{E}$, there are some sentences of ψ of \mathcal{L} to which multiple different probabilities attach.

The Maximum Entropy Approach arises using $>$ for comparing entropies of probability functions in \mathbb{P} [28, 45–48].

Remark 1 (Comparisons to the Entropy Limit Approach). It is known that the Entropy Limit Approach is invariant under arbitrary permutations of the constant symbols [25, Footnote 2]. The Entropy Limit Approach however suffers from a finite model problem, a premiss sentence formalising the existence of an infinite order does not have a finite model [42, Sect. 4.1] and hence the entropy limit is undefined.

The three Maximum Entropy Approaches defined in Definition 6 are invariant under *finite* permutations of constant symbols, the three notions of comparative entropy only depend on the limiting behaviour of H_n and n -entropy is invariant under permutation of the first n constant symbols (Definition 3). Whether these Maximum Entropy Approaches are invariant under infinite permutations is still to be determined. Since these approaches do not make use of finite models, they are immune to the finite model problem. See [19, 25] for more detailed comparisons and further background.

Remark 2 (Maximum Entropy Functions). Determining maximum entropy functions is an often difficult endeavour. In concrete applications, it is often easier to determine the entropy limit first and then show that the entropy limit also has maximal entropy. See [25] for an overview of the cases in which a maximum entropy function exists and is unique.

Trivially, if the equivocator function, $P_{=} \in \mathbb{P}$, which for all n assigns all n -states the same probability of $1/|\Omega_n|$,¹ is in \mathbb{E} , then $\{P_{=}\} = \text{maxent}_{>} \mathbb{E}$.

In a forthcoming paper [26], we show that if there is only a single premiss φ such that $0 < P_{=}(\varphi) < 1$, then the maximum entropy function is obtained from (Jeffrey) updating the equivocator function. For the premiss φ^c with $0 \leq c \leq 1$ it holds that $\{cP_{=}(\cdot|\varphi_N) + (1-c)P_{=}(\cdot|\neg\varphi_N)\} = \text{maxent}_{>} \mathbb{E}$, where N is maximal such that $t_N \in \varphi$ and φ_N is defined as the disjunction of N -states ω_N such that $P_{=}(\varphi \wedge \omega_N) > 0$.

Cases with multiple uncertain premisses are, in general, still poorly understood.

¹ Note that the equivocator function is the unique probability function in \mathbb{P} which is uniform over all Ω_n , $P_{=}(\omega_n) = \frac{1}{|\Omega_n|}$ for all n and all $\omega_n \in \Omega_n$. The name for this function is derived from the fact that it is maximally equivocal. The function has also been given other names. In Pure Inductive Logic it is known as the *completely independent probability function* and is often denoted by c_{∞} in reference to the role it plays in Carnap’s famous continuum of inductive methods [3].

In the next section, I study (the relationships of) these binary relations and the arising inductive logics. Particular attention is paid to the case of a unique probability function for inference, $|\max_{>} \mathbb{E}| = 1$. These cases are of particular interest, since they deliver well-defined (unique) probabilities for inductive inference.

3 Maximal (Modified) Entropy

I first consider two notions of refinement relating these three binary relations.

Definition 7 (Strong Refinement). $>$ is called a strong refinement of \gg , if and only if the following hold

- $>$ is a refinement of \gg , for all $P, Q \in \mathbb{P}$ it holds that $P \gg Q$ entails $P > Q$,
- for all $R, P, Q \in \mathbb{P}$ it holds that, if $R \gg P$ and $P > Q$ are both true, then $R \gg Q$ and $R \neq Q$.

Definition 8 (Centric Refinement). A refinement $>$ of \gg is called centric, if and only if for all different $R, P \in \mathbb{P}$ with $R > P$ it holds that $(R+P)/2 \gg P$.

The name centric has been chosen to emphasise that the centre between R and P is greater than P .

Clearly, not all binary relations possess strong refinements; not all binary relations possess centric refinements.

Proposition 1 (Strong and Centric Refinements). $] \text{ is a strong and centric refinement of } \succ.) \text{ is a strong and centric refinement of }] \text{ and of } \succ.$

Proof. For ease of comparison, I now display the three notions of comparative entropy line by line. The first line defines $P \succ Q$, the second line $P]Q$ and the third line $P)Q$. The second conjunct in the first definition is superfluous as is the second conjunct in the third definition:

$$\begin{aligned} H_n(P) \leq H_n(Q) \text{ not infinitely often} \ \& \ H_n(P) > H_n(Q) \text{ infinitely often} \\ H_n(P) < H_n(Q) \text{ not infinitely often} \ \& \ H_n(P) > H_n(Q) \text{ infinitely often} \\ H_n(P) < H_n(Q) \text{ not infinitely often} \ \& \ H_n(P) \geq H_n(Q) \text{ infinitely often} . \end{aligned}$$

By thusly spelling out both comparative notions of entropy one easily observes that $P \succ Q$ entails $P]Q$, and that $P]Q$ entails $P)Q$. This establishes the refinement relationships.

Strong Refinements. Next note that, if $R \succ Q$ or if $R]Q$, then $R \neq Q$.

$] \text{ is a strong refinement of } \succ$: Let $R \succ P$ and $P]Q$. Then $R \neq Q$. Furthermore, $H_n(R) \leq H_n(Q)$ is true for at most finitely many n , since from some N onwards P has always greater or equal n -entropy than Q . So, $R \succ Q$.

$) \text{ is a strong refinement of }]$: Let $R]P$ and $P)Q$. Then $R \neq Q$. From some N onwards P has always greater or equal n -entropy than Q . There are also infinitely many $n \in \mathbb{N}$ such that $H_n(R) > H_n(P)$. So, $R]Q$.

) is a strong refinement of \succ : Let $R \gg P$ and $P)Q$. Then $R \neq Q$. From some N onwards P has always greater or equal n -entropy than Q . From some N' onwards R has always greater n -entropy than P . Hence, $H_n(R) \leq H_n(Q)$ can only be the case for finitely many $n \in \mathbb{N}$. So, $R \succ Q$.

Centric Refinement. First, note that different probability functions disagree on some quantifier free sentence $\varphi \in \mathcal{L}_N$ (*Gaifman's Theorem* [8]). Since $\varphi \in \mathcal{L}_{n+N}$ for all $n \geq 1$, these probability functions also disagree on all more expressive sub-languages \mathcal{L}_{n+N} .

] is a centric refinement of \succ : Fix arbitrary probability functions R, P defined on \mathcal{L} with $R]P$. $R \neq P$. From the concavity of the function H_n it follows that $H_n(\frac{R+P}{2}) > H_n(P)$, whenever $H_n(R) \geq H_n(P)$. By definition of], there are only finitely many n for which $H_n(R) \geq H_n(P)$ fails to hold. Hence, $\frac{R+P}{2} \succ P$ by definition of \succ .

) is a centric refinement of \succ : Fix arbitrary probability functions R, P defined on \mathcal{L} with $R)P$. Note that R may be equal to P . From the concavity of the function H_n it follows that $H_n(\frac{R+P}{2}) > H_n(P)$, whenever $H_n(R) \geq H_n(P)$. By definition of), there are only finitely many n for which $H_n(R) \geq H_n(P)$ fails to hold. Hence, $\frac{R+P}{2} \succ P$ by definition of \succ .

) is a centric refinement of]: Fix arbitrary probability functions R, P defined on \mathcal{L} with $R)P$. Note that R may be equal to P . Since $\frac{R+P}{2} \succ P$ (see above case) and since] is a refinement of \succ , it holds that $\frac{R+P}{2}]P$. □

Remark 3 (Properties of Comparative Entropies). If $H_n(P) = H_n(Q)$ for all even n and $H_n(P) > H_n(Q)$ for all odd n , then $P]Q$ and $P \not\succeq Q$. Hence,] is a proper refinement of \succ .

For $P = Q$ it holds that $P)Q$ and $Q)P$. Hence,) is a proper refinement of] and thus a proper refinement of \succ .

] is transitive, irreflexive, acyclic and asymmetric.) is transitive, reflexive and has non-trivial cycles, e.g., for all probability functions P, Q with zero-entropy, $H_n(P) = 0$ for all $n \in \mathbb{N}$, it holds that $P)Q$.

I now turn to entropy maximisation and the induced inductive logics.

Proposition 2 (Downwards Uniqueness). *Let $>$ be a strong refinement of \gg . If $\text{maxent}_{\gg} \mathbb{E} = \{Q\}$, then $\{Q\} = \text{maxent}_{\gg} \mathbb{E} = \text{maxent}_{>} \mathbb{E}$.*

In case the inductive logic induced by \gg as a notion of a comparative entropy provides a unique probability function for rational inference, so does the inductive logic induced by $>$.

Proof. Note at first that since $>$ is a refinement of \gg it holds that

$$\text{maxent}_{>} \mathbb{E} \subseteq \text{maxent}_{\gg} \mathbb{E} . \tag{3}$$

Maximal elements according to \gg may not be maximal according to $>$ and all maximal elements according to $>$ are also maximal according to \gg .

Assume for the purpose of deriving a contradiction that $Q \notin \text{maxent}_{>} \mathbb{E}$. Then, there has to exist a $P \in \mathbb{E} \setminus \{Q\}$ such that $P > Q$ but $P \not\gg Q$ fails to hold ($\{Q\} = \text{maxent}_{\gg} \mathbb{E}$ holds by assumption).

However, since $\{Q\} = \text{maxent}_{\gg} \mathbb{E}$ and $Q \notin \text{maxent}_{>} \mathbb{E}$ hold, there has to exist some $R \in \mathbb{E} \setminus \{P\}$ such that $R \gg P$, P cannot have maximal \gg -entropy. We hence have $R \gg P$ and $P > Q$. Since $>$ is a strong refinement of \gg , we obtain $R \gg Q$ and $R \neq Q$. Since $R \in \mathbb{E}$ it follows from the definition of $\text{maxent}_{\gg} \mathbb{E}$ that $Q \notin \text{maxent}_{\gg} \mathbb{E}$. Contradiction. So, $Q \in \text{maxent}_{>} \mathbb{E}$.

Since $\{Q\} = \text{maxent}_{\gg} \mathbb{E} \stackrel{(3)}{\supseteq} \text{maxent}_{>} \mathbb{E} \ni Q$, it follows that $\text{maxent}_{>} \mathbb{E} = \{Q\}$. □

The converse is also true for convex \mathbb{E} and centric refinements.

Proposition 3 (Upwards Uniqueness). *If \mathbb{E} is convex, $>$ is a centric refinement of \gg and $\text{maxent}_{>} \mathbb{E} = \{Q\}$, then $\{Q\} = \text{maxent}_{\gg} \mathbb{E} = \text{maxent}_{>} \mathbb{E}$.*

Proof. Assume for contradiction that there exists a feasible probability function $P \in \mathbb{E} \setminus \{Q\}$ such that P is not \gg -dominated by the probability functions in \mathbb{E} but $>$ -dominated by some $R \in \mathbb{E} \setminus \{P\}$, $R > P$. Now define $S = \frac{1}{2}(P + R)$ and note that $S \in \mathbb{E}$ (convexity) and that S, P, R are pairwise different, $|\{S, P, R\}| = 3$.

Since $>$ is a centric refinement of \gg , we conclude that $S \gg P$, which contradicts that $P \in \text{maxent}_{\gg} \mathbb{E}$ and $P \neq Q$. So, only Q can be in $\text{maxent}_{\gg} \mathbb{E}$.

Since $Q \in \text{maxent}_{>} \mathbb{E}$ and $\text{maxent}_{>} \mathbb{E} \stackrel{(3)}{\subseteq} \text{maxent}_{\gg} \mathbb{E}$ it follows that $\{Q\} = \text{maxent}_{\gg} \mathbb{E}$. □

Theorem 1 (Triple Uniqueness). *If \mathbb{E} is convex and at least one of $\text{maxent}_{\downarrow} \mathbb{E}$, $\text{maxent}_{\uparrow} \mathbb{E}$ or $\text{maxent}_{>} \mathbb{E}$ is a singleton, then*

$$\text{maxent}_{\downarrow} \mathbb{E} = \text{maxent}_{\uparrow} \mathbb{E} = \text{maxent}_{>} \mathbb{E} .$$

Proof. Simply apply the above three propositions. □

It is known that $\text{maxent}_{>} \mathbb{E}$ is a singleton, in case of a certain Σ_1 premiss [44], a class of Π_1 premisses [25] and a class of constraints for unary languages [42]. As remarked above, we show in a forthcoming paper [26], that for a single premiss φ^c with $0 < P_{=}(\varphi) < 1$ and $0 \leq c \leq 1$ there exists a unique maximum entropy function in $\text{maxent}_{>} \mathbb{E}$, which is obtained from suitably (Jeffrey) updating the equivocator.

Premiss sentences φ with $P_{=}(\varphi) = 0$ and cases with multiple uncertain premisses – on the other hand – are still poorly understood.

4 Modification Number 3

The Maximum Entropy Approach, in its original formulation, fails to provide probabilities for uncertain inference for certain evidence bases of quantifier complexity Σ_2 [43, § 2.2]. For example, for the single certain premiss $\varphi := \exists x \forall y Uxy$

every probability function P consistent with the evidence must assign φ probability one, $P \in \mathbb{E}$ entails $P(\varphi) = 1$. There must hence be, at least, one constant t_k witnessing the existence of all these y , $P(\forall y Ut_k y) > 0$. Ceteris paribus, the $k - 1$ -entropy of such functions increases the greater the k such that t_k is a witness. Now suppose for contradiction that there exists a maximum entropy function $P \in \mathbb{E}$ with t_k as the first witness. Now construct a probability Q by postponing the witness by one, t_{k+1} is the first witness of the premiss according to Q . Q can be constructed such that $H_n(Q) \geq H_n(P)$ for all n . One can then show that $R := \frac{Q+P}{2}$ has strictly greater n -entropy than P for all $r \geq n + 1$ (concavity of H_n). It is hence the case that for all $P \in \mathbb{E}$ there exists a $R \in \mathbb{E}$ such that $R \succ P$. $\text{maxent}_{\succ} \mathbb{E}$ is hence empty, see the forthcoming [26] for more details and a proof.

Theorem 1 shows that inductive logics induced by $\}$ and $\}$ also do not produce a unique probability for uncertain inference for the certain premiss $\varphi = \exists x \forall y Uxy$.

From the perspective of this paper, we see that the inductive logic of the standard Maximum Entropy Approach fails to deliver well-defined probabilities for inference for the certain premiss $\varphi = \exists x \forall y Uxy$, because the relation \succ holds for too many pairs of probability functions. Proceeding in the spirit of this paper, it seems sensible to define a modified inductive logic induced by a notion of comparative entropy, $\}$, which holds for fewer pairs of probability functions. That is, $\}$ is refined by \succ .

Closest to the spirit of Definition 3 is to define $P\}Q$ as follows.

Definition 9 (Modification Number 3). $P\}Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.

Clearly, the other three notions of comparative entropy are refinements of $\}$.

Proposition 4 (Comparison of $\}$ vs. $\succ, \}$). Neither of the three binary relations $\succ, \}$ is a strong refinement and neither is a centric refinement of $\}$.

Proof. Consider three pairwise different probability functions P, Q, R with i) $H_n(P) > H_n(Q)$ for all n , ii) $\frac{H_n(P)}{H_n(Q)} \approx 1$, iii) $H_1(Q) = H_1(R) - \delta$ for large $\delta > 0$ and iv) $H_n(Q) > H_n(R)$ for all $n \geq 1$.

Then $P\}Q$ and $Q \succ R, Q\}R, Q)R$ all hold. Now note that $H_1(P) < H_1(R)$ and thus $P\}R$ fails to hold. Hence, none of $\succ, \}$ is a strong refinement of $\}$. Finally, observe that $\frac{Q+R}{2}\}R$ fails to hold. Hence, none of $\succ, \}$ is a centric refinement of $\}$. □

The aim here is to define a different inductive logic producing well-defined probabilities for the premiss sentence $\varphi = \exists x \forall y Rxy$ (and other premiss sentences). Theorem 1 shows that a different logic can only arise, if none of the other three notions of comparative entropy is a strong and centric refinement of $\}$. Proposition 4 shows that neither of these notions is a strong and centric refinement. It is hence in principle possible that $\}$ defines a novel inductive logic.

The following proposition shows that this not only possible in principle by providing a case in which the induced inductive logics do come apart.

Proposition 5 (Inductive Logics). *The binary relation $\} induces a different inductive logic than $\succ,],)$.$*

Proof. Let U be the only and unary relation symbol of \mathcal{L} . Suppose there is no evidence, then all probability functions are consistent with the empty set of premisses, $\mathbb{E} = \mathbb{P}$. Then every $P \in \mathbb{P}$ with $P(Ut_1) = P(\neg Ut_1) = 0.5$ has maximal 1-entropy. Hence, all such P are members of $\text{maxent}_{\} \mathbb{E}$. For $\square \in \{\succ,],)\}$ it holds that $\text{maxent}_{\square} \mathbb{E} = \{P_{=}\}$. So, $\text{maxent}_{\square} \mathbb{E}$ is a proper subset of $\text{maxent}_{\} \mathbb{E}$:

$$\text{maxent}_{\square} \mathbb{E} = \{P_{=}\} \subset \{P \in \mathbb{P} : P(Ut_1) = P(\neg Ut_1)\} \subset \text{maxent}_{\} \mathbb{E} .$$

□

This proof leads to the following more general observation:

Proposition 6 (Finite Sublanguages). *If there exists an $n \in \mathbb{N}$ and a $P \in \mathbb{E}$ such that $H_n(P) = \max\{H_n(Q) : Q \in \mathbb{E}\}$, then $P \in \text{maxent}_{\} \mathbb{E}$.*

This strong focus on single sublanguages \mathcal{L}_n makes $\text{maxent}_{\}$ unsuitable as an inductive logic for infinite predicate languages, as the following example demonstrates.

Example 1 (Absurdity of Modification 3). Consider the case in which the premisses jointly determine the probabilities on \mathcal{L}_1 . For example, the given language \mathcal{L} contains two relation symbols: a unary relation symbol U_1 and a binary relation symbol U_2 . The premisses are $U_2t_1t_1$ holds with certainty and $P(U_1t_1) = 10\%$. Then every probability function $P \in \mathbb{P}$ that satisfies these two premisses has a 1-entropy of $H_1(P) = -0.9 \cdot \log(0.9) - 0.1 \cdot \log(0.1)$. So, $H_1(P) = \max\{H_1(Q) : Q \in \mathbb{E}\}$. This means that every feasible probability function (of which there are many) is a maximum entropy function – regardless of how entropic (or not) probabilities are assigned to the other sublanguages \mathcal{L}_n for $n \geq 2$:

$$\text{maxent}_{\} \mathbb{E} = \mathbb{E} .$$

5 Conclusions

Maximum entropy inductive logic on infinite domains lacks a paradigm approach. The Entropy Limit Approach, the Maximum Entropy Approach as well as the here studied modified Maximum Entropy Approaches induce the same unique inductive logic in a number of natural cases (Theorem 1 and [25,42,44]). This points towards a unified picture of maximum entropy inductive logics – in spite of the number of possible ways to define such inductive logics.

This uniqueness is particularly noteworthy in light of a string of results suggesting and comparing different notions of entropy (maximisation), which lead to different maximum entropy functions [5–7, 10, 13, 18, 20, 27].

The Maximum Entropy Approach fails to provide probabilities for uncertain inference for some evidence bases of quantifier complexity Σ_2 [43, § 2.2]. In these cases, for all $P \in \mathbb{E}$ there exists a $Q \in \mathbb{E}$ such that $Q \succ P$ and maxent \mathbb{E} is hence empty [26]. One way to sensibly define an inductive logic could be to consider a binary relation which is refined by \succ . Unfortunately, the most obvious way to define such an inductive logic produces absurd results (Proposition 6). Finding a way to sensibly define a (maximum entropy) inductive logic properly dealing with such cases must be left to further study.

Further avenues for future research suggest themselves. Firstly, the question arises whether the first two here suggested modifications and the original Maximum Entropy Approach agree more widely or whether they come apart in important cases. If they do provide different probabilities for inductive inference, which of them is to be preferred and why? Secondly, are there further *prima facie* plausible ways to modify the Maximum Entropy Approach? Thirdly, are there other modifications of the Entropy Limit Approach? If so, how do they look like and what are the implications for the induced inductive logics? Fourthly, what is the status of the entropy limit conjecture [48, p. 191], the conjecture that the Entropy Limit Approach and the Maximum Entropy Approach agree under the assumption that the former is well-defined, in light of these modifications? Finally, cases with multiple uncertain premisses remain poorly understood and pose a challenge to be tackled.

Acknowledgements. Many thanks to Jeff Paris, Soroush Rafiee Rad, Alena Venovská and Jon Williamson for continued collaboration on maximum entropy inference. I'm also indebted to anonymous referees who helped me improve this paper.

References

1. Barnett, O., Paris, J.B.: Maximum entropy inference with quantified knowledge. *Logic J. IGPL* **16**(1), 85–98 (2008). <https://doi.org/10.1093/jigpal/jzm028>
2. Carnap, R.: The two concepts of probability: the problem of probability. *Philos. Phenomenological Res.* **5**(4), 513–532 (1945). <https://doi.org/10.2307/2102817>
3. Carnap, R.: *The Continuum of Inductive Methods*. Chicago University of Chicago Press, Chicago (1952)
4. Crupi, V.: Inductive logic. *J. Philos. Logic* **44**(6), 641–650 (2015). <https://doi.org/10.1007/s10992-015-9348-8>
5. Crupi, V., Nelson, J., Meder, B., Cevolani, G., Tentori, K.: Generalized information theory meets human cognition: introducing a unified framework to model uncertainty and information search. *Cogn. Sci.* **42**, 1410–1456 (2018). <https://doi.org/10.1111/cogs.12613>
6. Csiszár, I.: Axiomatic characterizations of information measures. *Entropy* **10**(3), 261–273 (2008). <https://doi.org/10.3390/e10030261>
7. Cui, H., Liu, Q., Zhang, J., Kang, B.: An improved deng entropy and its application in pattern recognition. *IEEE Access* **7**, 18284–18292 (2019). <https://doi.org/10.1109/access.2019.2896286>

8. Gaifman, H.: Concerning measures in first order calculi. *Isr. J. Math.* **2**(1), 1–18 (1964). <https://doi.org/10.1007/BF02759729>
9. Groves, T.: Lakatos’s criticism of Carnapian inductive logic was mistaken. *J. Appl. Logic* **14**, 3–21 (2016). <https://doi.org/10.1016/j.jal.2015.09.014>
10. Grünwald, P.D., Dawid, A.P.: Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Ann. Stat.* **32**(4), 1367–1433 (2004). <https://doi.org/10.1214/009053604000000553>
11. Haenni, R., Romeijn, J.W., Wheeler, G., Williamson, J.: *Probabilistic Argumentation*, Synthese Library, vol. 350. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-0008-6_3
12. Halpern, J.Y., Koller, D.: Representation dependence in probabilistic inference. *J. Artif. Intell. Res.* **21**, 319–356 (2004). <https://doi.org/10.1613/jair.1292>
13. Hanel, R., Thurner, S., Gell-Mann, M.: Generalized entropies and the transformation group of superstatistics. *Proc. Nat. Acad. Sci.* **108**(16), 6390–6394 (2011). <https://doi.org/10.1073/pnas.1103539108>
14. Howarth, E., Paris, J.B.: Pure inductive logic with functions. *J. Symbolic Logic*, 1–22 (2019). <https://doi.org/10.1017/jsl.2017.49>
15. Jaynes, E.T.: *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge (2003)
16. Kließ, M.S., Paris, J.B.: Second order inductive logic and Wilmers’ principle. *J. Appl. Logic* **12**(4), 462–476 (2014). <https://doi.org/10.1016/j.jal.2014.07.002>
17. Landes, J.: *The Principle of Spectrum Exchangeability within Inductive Logic*. Ph.D. thesis, Manchester Institute for Mathematical Sciences (2009). <https://jlandes.files.wordpress.com/2015/10/phdthesis.pdf>
18. Landes, J.: Probabilism, entropies and strictly proper scoring rules. *Int. J. Approximate Reason.* **63**, 1–21 (2015). <https://doi.org/10.1016/j.ijar.2015.05.007>
19. Landes, J.: The entropy-limit (Conjecture) for Σ_2 -premisses. *Stud. Logica.* **109**, 423–442 (2021). <https://doi.org/10.1007/s11225-020-09912-3>
20. Landes, J., Masterton, G.: Invariant equivocation. *Erkenntnis* **82**, 141–167 (2017). <https://doi.org/10.1007/s10670-016-9810-1>
21. Landes, J., Paris, J., Vencovská, A.: Language invariance and spectrum exchangeability in inductive logic. In: Mellouli, K. (ed.) *ECSQARU 2007. LNCS (LNAI)*, vol. 4724, pp. 151–160. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75256-1_16
22. Landes, J., Paris, J.B., Vencovská, A.: Some aspects of polyadic inductive logic. *Stud. Logica* **90**(1), 3–16 (2008). <https://doi.org/10.1007/s11225-008-9140-7>
23. Landes, J., Paris, J.B., Vencovská, A.: Representation theorems for probability functions satisfying spectrum exchangeability in inductive logic. *Int. J. Approximate Reason.* **51**(1), 35–55 (2009). <https://doi.org/10.1016/j.ijar.2009.07.001>
24. Landes, J., Paris, J.B., Vencovská, A.: A survey of some recent results on spectrum exchangeability in polyadic inductive logic. *Synthese* **181**, 19–47 (2011). <https://doi.org/10.1007/s11229-009-9711-9>
25. Landes, J., Rafiee Rad, S., Williamson, J.: Towards the entropy-limit conjecture. *Ann. Pure Appl. Logic* **172**, 102870 (2021). <https://doi.org/10.1016/j.apal.2020.102870>
26. Landes, J., Rafiee Rad, S., Williamson, J.: Determining maximal entropy functions for objective Bayesian inductive logic (2022). Manuscript
27. Landes, J., Williamson, J.: Objective Bayesianism and the maximum entropy principle. *Entropy* **15**(9), 3528–3591 (2013). <https://doi.org/10.3390/e15093528>
28. Landes, J., Williamson, J.: Justifying objective Bayesianism on predicate languages. *Entropy* **17**(4), 2459–2543 (2015). <https://doi.org/10.3390/e17042459>

29. Niiniluoto, I.: The development of the Hintikka program. In: Gabbay, D.M., Hartmann, S., Woods, J. (eds.) *Handbook of the History of Logic*, pp. 311–356. Elsevier, Kidlington (2011). <https://doi.org/10.1016/b978-0-444-52936-7.50009-4>
30. Ognjanović, Z., Rašković, M., Marković, Z.: *Probability Logics*. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-47012-2>
31. Ortner, R., Leitgeb, H.: Mechanizing induction. In: Gabbay, D.M., Hartmann, S., Woods, J. (eds.) *Handbook of the History of Logic*, pp. 719–772. Elsevier (2011). <https://doi.org/10.1016/b978-0-444-52936-7.50018-5>
32. Paris, J.B.: Common sense and maximum entropy. *Synthese* **117**, 75–93 (1998). <https://doi.org/10.1023/A:1005081609010>
33. Paris, J.B.: *The Uncertain Reasoner's Companion: A Mathematical Perspective*, 2 edn. Cambridge Tracts in Theoretical Computer Science, vol. 39. Cambridge University Press, Cambridge (2006)
34. Paris, J.B.: What you see is what you get. *Entropy* **16**(11), 6186–6194 (2014). <https://doi.org/10.3390/e16116186>
35. Paris, J.B., Rad, S.R.: A note on the least informative model of a theory. In: Ferreira, F., Löwe, B., Mayordomo, E., Mendes Gomes, L. (eds.) *CiE 2010. LNCS*, vol. 6158, pp. 342–351. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13962-8_38
36. Paris, J.B., Vencovská, A.: On the applicability of maximum entropy to inexact reasoning. *Int. J. Approximate Reason.* **3**(1), 1–34 (1989). [https://doi.org/10.1016/0888-613X\(89\)90012-1](https://doi.org/10.1016/0888-613X(89)90012-1)
37. Paris, J.B., Vencovská, A.: A note on the inevitability of maximum entropy. *Int. J. Approximate Reason.* **4**(3), 183–223 (1990). [https://doi.org/10.1016/0888-613X\(90\)90020-3](https://doi.org/10.1016/0888-613X(90)90020-3)
38. Paris, J.B., Vencovská, A.: In defense of the maximum entropy inference process. *Int. J. Approximate Reason.* **17**(1), 77–103 (1997). [https://doi.org/10.1016/S0888-613X\(97\)00014-5](https://doi.org/10.1016/S0888-613X(97)00014-5)
39. Paris, J.B., Vencovská, A.: Common sense and stochastic independence. In: Corfield, D., Williamson, J. (eds.) *Foundations of Bayesianism*, pp. 203–240. Kluwer, Dordrecht (2001)
40. Paris, J.B., Vencovská, A.: *Pure Inductive Logic*. Cambridge University Press, Cambridge (2015)
41. Paris, J.B., Vencovská, A.: Six problems in pure inductive logic. *J. Philos. Logic* (2019). <https://doi.org/10.1007/s10992-018-9492-z>
42. Rafiee Rad, S.: *Inference processes for first order probabilistic languages*. Ph.D. thesis, Manchester Institute for Mathematical Sciences (2009). <http://www.rafeerad.org/manthe.pdf>
43. Rafiee Rad, S.: Equivocation axiom on first order languages. *Stud. Logica.* **105**(1), 121–152 (2017). <https://doi.org/10.1007/s11225-016-9684-x>
44. Rafiee Rad, S.: Maximum entropy models for Σ_1 sentences. *J. Logics Appl.* **5**(1), 287–300 (2018). <http://www.collegepublications.co.uk/admin/download.php?ID=ifcolog00021>
45. Williamson, J.: Objective Bayesian probabilistic logic. *J. Algorithms* **63**(4), 167–183 (2008). <https://doi.org/10.1016/j.jalgor.2008.07.001>
46. Williamson, J.: Objective Bayesianism with predicate languages. *Synthese* **163**(3), 341–356 (2008). <https://doi.org/10.1007/s11229-007-9298-y>
47. Williamson, J.: *In Defence of Objective Bayesianism*. Oxford University Press, Oxford (2010)
48. Williamson, J.: *Lectures on Inductive Logic*. Oxford University Press, Oxford (2017)