Determining maximal entropy functions for objective Bayesian inductive logic

Draft of March 28, 2022

Abstract

According to the objective Bayesian approach to inductive logic, premisses inductively entail a conclusion just when every probability function with maximal entropy, from all those that satisfy the premisses, satisfies the conclusion. However, when premisses and conclusion are constraints on probabilities of sentences of a first-order predicate language, it is by no means obvious how to determine these maximal entropy functions. This paper makes progress on the problem in the following ways. Firstly, we introduce the concept of an entropy limit point and show that, if the set of probability functions satisfying the premisses contains an entropy limit point, then this limit point is unique and is the maximal entropy probability function. Next, we turn to the special case in which the premisses are simply sentences of the logical language. We show that if the uniform probability function gives the premisses positive probability, then the maximal entropy function can be found by simply conditionalising this uniform prior on the premisses. We generalise our results to demonstrate agreement between the maximal entropy approach and Jeffrey conditionalisation in the case in which there is a single premiss that specifies the probability of a sentence of the language. We show that, after learning such a premiss, certain inferences are preserved, namely inferences to inductive tautologies. Finally, we consider potential pathologies of the approach: we explore the extent to which the maximal entropy approach is invariant under permutations of the constants of the language, and we discuss some cases in which there is no maximal entropy probability function.

1 Introduction

Inference under uncertainty remains one of the challenges of our time. While there is widespread agreement that probabilities are well suited to capture uncertainty and that Bayesian and Jeffrey conditionalisation are key principles of rationality, there is significant disagreement about the proper choice of probabilities and their use. One prominent approach to uncertain inference appeals to the Maximum Entropy Principle of Jaynes (1957). This selects a probability function, from all those that agree with the available evidence, that is as equivocal as possible in the sense that it has maximum Shannon entropy (Shannon, 1948). The Maximum Entropy Principle is ofte employed as part of an objective Bayesian approach to inference (Jaynes, 2003; Williamson, 2010).

The use of the Maximum Entropy Principle on finite domains is well-understood. A number of axiomatic characterisations highlight some of its most important properties, such as irrelevance of extraneous information, independence in the absence of evidence of dependence, and invariance under uniform refinements of the underlying finite domain (Paris and Vencovská, 1990, 1997; Paris, 1994, 1998). Furthermore, MaxEnt inference is known to agree on finite domains with what one might call 'baseline rationality': Bayesian and Jeffrey conditionalisation turn out to be special cases of MaxEnt inference (Williams, 1980). While Jeffrey conditionalisation can only deal with a single uncertain premiss at a time, of the form $P(F) = c$, MaxEnt inference can handle multiple uncertain premisses of more complex forms simultaneously. Given a fixed finite domain and premisses of a suitable form, MaxEnt inference introduces an objective relation between premisses and conclusions, independent of the inferring agent. This objectivity facilitates the implementation of MaxEnt inferences in algorithms and automated systems.[1]

The application of MaxEnt to infinite domains is much less well understood. Firstly, axiomatic characterisations have yet to be put forward. Second, MaxEnt inference is only known to agree with Jeffrey conditionalisation on certain infinite domains that lack a logical structure (Caticha and Giffin, 2006). The focus of this paper is to shed some light on the application of MaxEnt to infinite domains—in particular, to its use as semantics for objective Bayesian inductive logic on infinite predicate languages.

There are two different explications of MaxEnt on infinite predicate languages. One, due to Jeff Paris and his co-workers, takes limits of maximum entropy functions on finite sublanguages (Barnett and Paris, 2008; Rafiee Rad, 2009; Paris and Rafiee Rad, 2010; Rafiee Rad, 2018, 2021). The second explication considers maximal entropy probability functions defined on the infinite language as a whole (Williamson, 2008; Landes and Williamson, 2015; Williamson, 2017; Rafiee Rad, 2017; Landes, 2021a). The limit approach provides a means to determine the probabilities for MaxEnt inference. However, this construction has problems: in some cases, it does not yield an answer at all (Rafiee Rad, 2009; Paris and Rafiee Rad, 2010); in other cases the constructed probabilities fail to satisfy the given premisses (Landes, 2021b). The maximal entropy approach can be used in a wider range of situations (Rafiee Rad, 2009, 2017), but the approach is less constructive and it is less clear how to determine maximal entropy probability functions. It has however been conjectured that both approaches agree where the limit approach is well defined (Williamson, 2017; Landes et al., 2021).

In this paper we study the second of these two approaches: the maximal entropy approach. We first give a method for determining the maximal entropy probability function in many general scenarios, by introducing the concept of an entropy limit point (Theorem 12). Then we show that the approach generalises both Bayesian conditionalisation (Theorem 30) and Jeffrey conditionalisation (Theorem 37). This not only clarifies which probabilities the maximal entropy

[1] Note however that inference using MaxEnt can be computationally complex in the worst cas—see Paris (1994, Chapter 10) and Pearl (1988, p. 463), and also Goldman (e.g., 1987); Goldman and Rivest (e.g., 1988); Ormoneit and White (e.g., 1999); Balestrino et al. (e.g., 2006); Chen et al. (e.g., 2010); Landes and Williamson (e.g., 2016). We will not be concerned with computational complexity in this paper.

approach picks out, but also gives a simple way to determine these probabilities and shows that the maximal entropy approach agrees with baseline rationality.

We then turn to general features of the maximal entropy approach. We see that certain inferences drawn in the absence of any premisses—inferences to inductive tautologies—are preserved when a premiss is added (§7). We show that while the notion of comparative entropy used to define the maximal entropy probability functions can depend on the order of the constant symbols (Proposition 44), this order is rendered irrelevant in all cases in which the maximal entropy approach simplifies to Bayesian or Jeffrey conditionalisation (Theorem 45, Corollary 46). Finally, it becomes clear why the maximal entropy approach fails to provide probabilities in some cases. These cases are those where the premiss has zero prior probability. Updating on events of zero prior probability is notoriously problematic. We investigate the extent of these failures in §9, show that they arise in all levels of the arithmetic hierarchy including and above Σ_2 (Theorem 48), and provide a refinement of the approach to handle these problematic cases.

It is worth noting the relation between this approach and perhaps the most well-known approach to inductive logic, namely that of Rudolf Carnap (see, e.g., Carnap, 1952). In common with Carnap's approach, we consider the problem of developing an inductive logic involving sentences of a first-order predicate language. However, the maximal entropy approach differs in two key respects. Firstly, our setting is more general, as it considers premiss statements which attach probabilities or sets of probabilities to sentences of the logical language, while Carnap considered only the sentences themselves. Second, our approach is based on the idea of entropy maximisation, while Carnap's approach appeals to Bayesian conditionalisation involving exchangeable prior probability functions. The latter approach is susceptible to serious objections (Williamson, 2017, Chapter 4).

2 Objective Bayesian Inductive Logic

An important class of probabilistic logics consider entailment relationships of the following form (Haenni et al., 2011):

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y \quad .$$

Here, $\varphi_1, \dots, \varphi_k, \psi$ are sentences of a logical language $\mathcal{L}$ and $X_1, \dots, X_k, Y$ are sets of probabilities. This entailment relationship should be interpreted as saying: $\varphi_1, \dots, \varphi_k$ having probabilities in $X_1, \dots, X_k$ respectively inductively entails that ψ has probability in Y.

The objective Bayesian approach to inductive logic interprets probabilities as rational degrees of belief. It takes the premisses on the left-hand side of the entailment relationship to capture all the constraints on rational degrees of belief that are inferred from evidence, and it uses Jaynes' Maximum Entropy Principle to determine a rational belief function with which to calculate the probability of a conclusion statement ψ. Thus if $\mathcal{L}$ is a finite propositional language, $X_1, \dots, X_k$ are closed convex sets of probabilities (i.e. closed intervals), and the premisses are consistent, an entailment relationship holds just when the probability function with maximum entropy, amongst all those that satisfy the premisses, gives a probability in Y to ψ (Williamson, 2010, Chapter 7).

This approach has been extended to the case in which $\mathcal{L}$ is a first-order predicate language in the following way. Suppose $\mathcal{L}$ has countably many constant symbols $t_1, t_2, \dots$ and finitely many relation symbols $U_1, \dots, U_l$. Let $a_1, a_2, \dots$ run through the atomic sentences of the form $U_i t_{i_1} \dots t_{i_k}$ in such a way that those atomic sentences involving only $t_1, \dots, t_n$ occur before those involving t_{n+1}, for each n. Consider the finite sub-languages $\mathcal{L}_n$, containing only constant symbols $t_1, \dots, t_n$.

Definition 1 (n-states)**.** Ω_n is the set of *n-states* of $\mathcal{L}$, i.e., sentences of the form $\pm a_1 \wedge \dots \wedge \pm a_{r_n}$ involving the atomic sentences $a_1, \dots, a_{r_n}$ of $\mathcal{L}_n$, which only feature the constants $t_1, \dots, t_n$.[2] The n-states for $\mathcal{L}$ are thus the sentences

$$\bigwedge_{\substack{1 \le i \le l \\ 1 \le t_{j_1}, \dots, j_{k_i} \le n}} U_i^{\epsilon_{t_{j_1}, \dots, t_{j_k}}} t_{j_1} \dots t_{j_{k_i}}$$

where k_i is the arity of U_i, $\epsilon_{t_{j_1}, \dots, t_{j_k}} \in \{0, 1\}$ and $U_i^1 t_{j_1} \dots t_{j_{k_i}} = U_i t_{j_1} \dots t_{j_{k_i}}$ and $U_i^0 t_{j_1} \dots t_{j_{k_i}} = \neg U_i t_{j_1} \dots t_{j_{k_i}}$.

Let $S\mathcal{L}, S\mathcal{L}_n$ be the sets of sentences of $\mathcal{L}, \mathcal{L}_n$ respectively.

Definition 2 (N_φ)**.** For a single given sentence φ we use N_φ to denote the greatest index of the constants appearing in φ, i.e., the greatest number n such that t_n occurs in φ. If φ has no constants, we adopt the convention that $N_\varphi = 1$.

Definition 3 (Probability)**.** A *probability function* P on $\mathcal{L}$ is a function $P : S\mathcal{L} \longrightarrow \mathbb{R}_{\ge 0}$ 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: $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^m \theta(t_i)\right)$.

A probability function is determined by the values it gives to the n-states—see, e.g., Williamson (2017, §2.6.3) and Gaifman (1964). We denote the set of probability functions by $\mathbb{P}$.

Of particular importance will be the *equivocator* function, $P_=$, which gives the same probability to each n-state, for each n.

Definition 4 (Measure)**.** The *measure* of a sentence θ is the probability given to it by the equivocator function. In particular, θ has *positive measure* if and only if $P_=(\theta) > 0$.

Definition 5 (Feasible Region)**.** We use $\mathbb{E}$ to refer to the set of probability functions that satisfy the premisses $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$, i.e.,

$$\mathbb{E} := \{P \in \mathbb{P} : P(\varphi_1) \in X_1, \dots, P(\varphi_k) \in X_k\} \ .$$

Two special cases will be particularly important in this paper. To distinguish the case of a single categorical premiss, φ, we often write $\mathbb{E}_\varphi$ instead of $\mathbb{E}$.

[2]The n-states are sometimes referred to as 'state descriptions'.

In the case of a single uncertain premiss, φ^X, we write $\mathbb{E}_{\varphi^X}$. Throughout, we shall assume that the X are intervals and that the feasible region is non-empty, $\mathbb{E} \neq \emptyset$.

Definition 6 (n-entropy)**.** The *n-entropy* of a probability function P is defined as $H_n(P) \stackrel{\text{df}}{=} -\sum_{\omega\in\Omega_n} P(\omega)\log P(\omega)$.

The n-entropies, which only take into account the probabilities on finitely many n-states, are then used to define a notion of comparative entropy on the infinite language $\mathcal{L}$ as a whole:

Definition 7 (Comparative Entropy)**.** We say that the probability function $P \in \mathbb{P}$ *has greater entropy than* $Q \in \mathbb{P}$, if and only if the n-entropy of P dominates that of Q for sufficiently large n, i.e., if and only if there is an $N \in \mathbb{N}$ such that for all $n \geq N$, $H_n(P) > H_n(Q)$.

The greater entropy relation defines a partial order on the probability functions on $\mathcal{L}$. We will focus on the maximal elements in $\mathbb{E}$ of this partial ordering:

Definition 8 (Maximal Entropy Functions)**.** The set of maximal entropy functions, maxent $\mathbb{E}$, is defined as

$$\text{maxent}\,\mathbb{E} := \{P \in \mathbb{E} \ : \ \text{there is no } Q \in \mathbb{E} \text{ that has greater entropy than } P\}.$$

In the absence of any premisses, $\text{maxent}\,\mathbb{E} = \text{maxent}\,\mathbb{P} = \{P_=\}$.

In this paper, we invoke the objective Bayesian notion of inductive entailment, denoted by $\mathrel{|\!\!\!\approx}^{\circ}$ (Williamson, 2017, §5.3):

Definition 9 (Objective Bayesian Inductive Entailment)**.** The premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ inductively entail ψ^Y, denoted by $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\!\approx}^{\circ} \psi^Y$, if $P(\psi) \in Y$ for all $P \in \text{maxent}\,\mathbb{E}$.

Note that this definition applies where maxent $\mathbb{E}$ is non-empty. We consider the case in which maxent $\mathbb{E}$ is empty in §9.

We will say that ψ is an *inductive tautology* if $\mathrel{|\!\!\!\approx}^{\circ} \psi$, i.e., if it has measure 1. It is an *inductive contradiction* if $\mathrel{|\!\!\!\approx}^{\circ} \neg\psi$, i.e., if it has measure 0. It is *inductively consistent* if $\not\mathrel{|\!\!\!\approx}^{\circ} \neg\psi$, i.e., if it has positive measure.

While the objective Bayesian approach provides appropriate semantics for inductive logic, it is not obvious how to determine the maximal entropy functions in order to ascertain whether a given entailment relationship holds. This is because the definition of maxent $\mathbb{E}$ seems to require a sort through members of $\mathbb{E}$ in order to find those with maximal entropy—a process that would be unfeasible in practice. This paper aims to address this question.

§3 introduces the concept of an entropy limit point in order to characterise maxent $\mathbb{E}$ in terms of certain limits of n-entropy maximisers. This gives a constructive procedure for determining maxent $\mathbb{E}$ when it contains an entropy limit point.

In §4 and §5 we consider an important special case—that in which the premisses are categorical sentences $\varphi_1, \ldots, \varphi_k$ (without attached probabilities) and where the maximal entropy function can be obtained simply by conditionalising the equivocator function.

3 Entropy Limit Points

This section adapts the techniques of Landes et al. (2021, §5) in order to characterise maxent $\mathbb{E}$ in terms of certain limits of n-entropy maximisers. Landes et al. (2021) were concerned with a very different question: that of showing that the above objective Bayesian semantics for inductive logic in terms of maximal entropy functions yields the same inferences as those produced by Paris' limit approach discussed in §1. Nevertheless, the results of Landes et al. (2021, §5) can be straightforwardly adapted to the present problem. The proofs of the two results in this section, which are close to those of Landes et al. (2021, Proposition 36) and Landes et al. (2021, Theorem 39), have been provided in Appendix 1.

We will consider the set of n-entropy maximisers for each n:

$$\mathbb{H}_n = \{P \in \mathbb{P} : H_n(P) \text{ is maximised}\} \ .$$

We now introduce the key concept of this section:

Definition 10 (Entropy Limit Point)**.** $P \in \mathbb{P}$ is an *entropy limit point* of $\mathbb{P}_1, \mathbb{P}_2, \ldots \subseteq \mathbb{P}$, if for each n there is some $Q_n \in \mathbb{P}_n$ such that $|H_n(Q_n) - H_n(P)| \longrightarrow 0$ as $n \longrightarrow \infty$. $P \in \mathbb{P}$ will be called an *entropy limit point* of $\mathbb{E}$ if it is an entropy limit point of $\mathbb{H}_1, \mathbb{H}_2, \ldots$.

Entropy limit points of $\mathbb{E}$ are of special interest because they are also limit points in terms of the L_1 distance,

$$\|P - Q\|_n \stackrel{\text{df}}{=} \sum_{\omega \in \Omega_n} |P(\omega) - Q(\omega)| \ .$$

Proposition 11. *If P is an entropy limit point of $\mathbb{E}$, then there are functions $Q_n \in \mathbb{H}_n$, for $n \geq 1$, such that $\|Q_n - P\|_n \longrightarrow 0$ as $n \longrightarrow \infty$.*

This property enables us to characterise the set of maximal entropy functions more constructively, in terms of a limit of n-entropy maximisers:

Theorem 12 (Entropy Limit Point)**.** *If $\mathbb{E}$ contains an entropy limit point P, then*

$$\text{maxent}\, \mathbb{E} = \{P\} \ .$$

Note that there can be at most one entropy limit point P of $\mathbb{E}$. This is because $\mathbb{E}$ is convex (by the convexity of $X_1, \ldots, X_k$) and the n-entropy maximiser of a convex set is uniquely determined on $\mathcal{L}_n$. Thus, the $\mathbb{H}_n$ can have at most one L_1 limit point.

Theorem 12 provides a simple procedure for showing that a hypothesised function P is in fact a maximal entropy function: show that it is an entropy limit point of n-entropy maximisers, and show that it is in $\mathbb{E}$. (Note that this is only a sufficient condition: if $\mathbb{E}$ contains no entropy limit point, then Theorem 12 does not allow us to infer anything about maxent $\mathbb{E}$.) Landes et al. (2021, Lemmas 40, 44) provide some tools for demonstrating that a hypothesised function is an entropy limit point of $\mathbb{E}$.

Example 13. Suppose we have a single premiss $\forall x U x^{\{c\}}$ where $\mathcal{L}$ has a single unary predicate U and $c \in [0,1]$. (We will often omit the curly braces and write φ^c instead of $\varphi^{\{c\}}$ in such cases.) In this case, the number r_n of atomic sentences of $\mathcal{L}_n$ is n. Any n-entropy maximiser gives probability c to the n-state $Ut_1 \wedge \ldots \wedge Ut_n$, which we abbreviate by θ_n, and divides probability $1-c$ amongst all other n-states:

$$P^n(\omega_n) = \begin{cases} c & : \quad \omega_n = \theta_n \\ \frac{1-c}{2^n-1} & : \quad \omega_n \models \neg\theta_n \end{cases} .$$

By the argument of Landes et al. (2021, Example 42), the following probability function is an entropy limit point:

$$P(\omega_n) = \begin{cases} c + \frac{1-c}{2^n} & : \quad \omega_n = \theta_n \\ \frac{1-c}{2^n} & : \quad \omega_n \models \neg\theta_n \end{cases} .$$

$P \in \mathbb{E}$ because $P(\forall x U x) = \lim_{n\to\infty} P(\theta_n) = c$. Hence by Theorem 12, $\operatorname{maxent} \mathbb{E} = \{P\}$.

Example 14. Consider a single categorical premiss $U_1 t_1 \vee \exists x \forall y U_2 xy$. In this case, $\mathbb{H}_n = \{P \in \mathbb{E} : P_{\restriction \mathcal{L}_n} = P_{=\restriction \mathcal{L}_n}\}$ for all n. Thus the equivocator function is the unique entropy limit point of $\mathbb{E}$. However, the equivocator function is not in $\mathbb{E}$, so it cannot be the maximal entropy function. Indeed, as will become apparent later (Theorem 30), $\operatorname{maxent} \mathbb{E} = \{P_=(\cdot|U_1 t_1)\}$.

4 Categorical Premisses and Bayesian Conditionalisation

We now consider an important special case: that in which the premisses are categorical sentences $\varphi_1, \ldots, \varphi_k$ of $\mathcal{L}$, i.e., there are no attached sets of probabilities $X_1, \ldots, X_k$, or equivalently, $X_1 = \ldots = X_k = \{1\}$. Let φ be the sentence $\varphi_1 \wedge \ldots \wedge \varphi_k$. In this section and the next, we consider $\mathbb{E} = \mathbb{E}_\varphi \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi) = 1\}$ and we show that there are several cases in which $\operatorname{maxent} \mathbb{E}$ can be found simply by conditionalising the equivocator function on φ.

Our first result directly applies Theorem 12:

Corollary 15. *If $P_=(\cdot|\varphi)$ is an entropy limit point of $\mathbb{E}_\varphi$, then*

$$\operatorname{maxent} \mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} .$$

Proof: $P_=(\cdot|\varphi)$ is contained in $\mathbb{E}_\varphi$ because $P_=(\varphi_i|\varphi) = 1$ for each $i = 1, \ldots, k$. Hence, Theorem 12 applies. ■

Note that the condition that $P_=(\cdot|\varphi)$ is an entropy limit point of $\mathbb{E}_\varphi$ presupposes that the probability function $P_=(\cdot|\varphi)$ is well defined, i.e., that φ has positive measure, $P_=(\varphi) > 0$.

Corollary 16. *If $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large n, then*

$$\operatorname{maxent} \mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} .$$

Proof: If $P_=(\cdot|\varphi) \in \mathbb{H}_n$ for sufficiently large n, then $P_=(\cdot|\varphi)$ is an entropy limit point of $\mathbb{E}_\varphi$. Hence, Corollary 15 applies. ■

Corollary 16 is useful because where it applies it provides a particularly simple procedure for determining maxent $\mathbb{E}_\varphi$. Also, it shows that the move to the infinite does not disrupt agreement between the Maximum Entropy Principle and conditionalisation: as long as conditionalising on φ maximises n-entropy for each sufficiently large n, it maximises entropy on the language as a whole. Because of its interest, we provide an alternative, more direct proof of Corollary 16 in Appendix 2.

Example 17. Suppose we have a single categorical premiss $\exists x Ux$, where $\mathcal{L}$ has a single unary predicate symbol U. $P_=(\exists x Ux) = 1$, so $P_=(\cdot|\exists x Ux) = P_=(\cdot)$. $P_= \in \mathbb{H}_1, \mathbb{H}_2, \ldots$, so Corollary 16 applies and maxent $\mathbb{E}_\varphi = \{P_=\}$.

Example 18. Suppose we have categorical premisses $Ut_2 \to Vt_3, \forall x \exists y Wxy$, where $\mathcal{L}$ has unary predicate symbols U and V and a binary relation symbol W. Now $P_=((Ut_2 \to Vt_3) = 0.75)$ and $P_=(\forall x \exists y Wxy) = 1$. So $P_=((Ut_2 \to Vt_3) \wedge \forall x \exists y Wxy) = 0.75$, and $P_=(\cdot|(Ut_2 \to Vt_3) \wedge \forall x \exists y Wxy) = P_=(\cdot|Ut_2 \to Vt_3)$. This latter function is in $\mathbb{H}_3, \mathbb{H}_4, \ldots$, so Corollary 16 applies and maxent $\mathbb{E}_\varphi = \{P_=(\cdot|Ut_2 \to Vt_3)\}$.

Finally, we note an important consequence of Corollary 16:

Corollary 19. *If φ is satisfiable and logically equivalent to a quantifier-free sentence, then*

$$\operatorname{maxent} \mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} \ .$$

Proof: Since φ is satisfiable and logically equivalent to a quantifier-free sentence, $P_=(\varphi) > 0$ (Paris, 1994, pp. 95, 102). Moreover, $P_=(\cdot|\varphi) \in \mathbb{H}_n$ for all $n \geq N_\varphi$, where N_φ is the greatest index of the constant symbols appearing in φ. ■

This result can be thought of as an analogue of Seidenfeld (1986, Result 1), which demonstrates agreement between the Maximum Entropy Principle and conditionalisation in the case of a finite domain. In the next section, we show that this result can be extended to the situation in which φ is not quantifier free.

5 An Alternative Route to Conditionalisation

This section demonstrates agreement between the maximal entropy approach and conditionalisation without appeal to entropy limit points.

As above we consider categorical sentences $\varphi_1, \ldots, \varphi_k$ and abbreviate $\varphi_1 \wedge \ldots \wedge \varphi_k$ by φ. The following definition will be central to several of the results in this section:

Definition 20 (n-support φ^n)**.** Let sentence φ^n be the disjunction of those n-states ω that are inductively consistent with φ, i.e., n-states ω such that $\not\mathrel{|\!\!\approx} \neg(\omega \wedge \varphi)$. Equivalently, these are the n-states ω such that $\varphi \wedge \omega$ has positive measure. Thus,

$$\varphi^n \stackrel{\text{df}}{=} \bigvee \{\omega \in \Omega_n : P_=(\omega \wedge \varphi) > 0\} \ .$$

If there are no n-states inductively consistent with φ, we take φ^n to be an arbitrary contradiction on $\mathcal{L}_n$.

We call φ^n the *inductive support* of φ on $\mathcal{L}_n$, or simply the *n-support* of φ. φ^{N_φ} will be referred to as the *support* of φ.[3] We use $|\varphi^n|$ to denote the number of n-states in the n-support φ^n, i.e., the number of n-states inductively consistent with φ.

Our main result of this section, Theorem 30, will show that when φ has positive measure, the maximal entropy function is the equivocator function conditional on φ, or, equivalently, the equivocator conditional on the support of φ. This provides a straightforward way of determining the maximal entropy function in that case.

We will first prove some technical lemmas to which the main result will appeal. The first lemma invokes the concept of exchangeability:

Definition 21 (Constant Exchangeability)**.** Let $\theta(x_1, x_2, \ldots, x_l)$ be a formula of $\mathcal{L}$ that does not contain constants. A probability function P on $\mathcal{SL}$ satisfies *constant exchangeability*, if and only if for all such θ and all sets of pairwise distinct constants $t_1, t_2, \ldots, t_l$, and $t'_1, t'_2, \ldots, t'_l$ it holds that

$$P(\theta(t_1, t_2, \ldots, t_l)) = P(\theta(t'_1, t'_2, \ldots, t'_l)) \ .$$

Equivalently, for all $n \in \mathbb{N}$ and all n-states $\omega_n, \nu_n \in \Omega_n$, if ω_n can be obtained from ν_n by a permutation of the first n constants then $P(\omega_n) = P(\nu_n)$.

We are obliged to Jeff Paris for pointing out the following lemma and Proposition 24 which follows from it.

Lemma 22. *Let ω_n be an n-state and suppose that the probability function P on $\mathcal{SL}$ satisfies constant exchangeability and $P(\varphi \wedge \psi|\omega_n) = P(\varphi|\omega_n) \cdot P(\psi|\omega_n)$ for all pairs of quantifier-free sentences φ, ψ with shared constants among $\{t_1, \ldots, t_l\}$, $l \leq n$. Then $P(\varphi \wedge \psi|\omega_n) = P(\varphi|\omega_n) \cdot P(\psi|\omega_n)$ for all $\varphi, \psi \in \mathcal{SL}$ whose shared constants are among $\{t_1, \ldots, t_l\}$.*

Proof: The proof follows by a straightforward adaptation of the proof of Paris and Vencovská (2015, Corollary 6.2) and proceeds by induction on the quantifier complexity of $\varphi \wedge \psi$ when written in Prenex Normal Form.

The result holds by assumption when $\varphi \wedge \psi$ is quantifier free. For the induction step it is sufficient to consider

$$\exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t}) \wedge \exists x_1, \ldots, x_s \psi(x_1, \ldots, x_s, \vec{t'})$$

where all constants appearing in both $\vec{t}$ and $\vec{t'}$ are included in $\{t_1, \ldots, t_l\}$.

Let $u_1, u_2, u_3, \ldots$ be distinct constants containing those in $\vec{t}$ and $u'_1, u'_2, u'_3, \ldots$ distinct constants containing those in $\vec{t'}$ such that $\{u_1, u_2, u_3, \ldots\}$ and $\{u'_1, u'_2, u'_3, \ldots\}$ are disjoint except for the constants shared in $\vec{t}$ and $\vec{t'}$.

By Paris and Vencovská (2015, Lemma 6.1),

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\ldots,i_r \leq n} \theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t})\right) \leftrightarrow \exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t})\,|\omega_n\right) = 1$$

[3] Recall that N_φ is the greatest index of the constants appearing in φ, or 1 if no constants appear in φ.

and

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\right) \leftrightarrow \exists x_1,\dots,x_s\eta(x_1,\dots,x_s,\vec{t'})\,|\omega_n\right) = 1.$$

Then for every $\epsilon > 0$ there is N large enough such that for all $n \geq N$

$$P\left(\left(\bigvee_{i_1,\dots,i_r\le n} \theta(u_{i_1},u_{i_2},\dots,u_{i_r},\vec{t})\right) \leftrightarrow \exists x_1,\dots,x_r\theta(x_1,\dots,x_r,\vec{t})\,|\omega_n\right) > 1-\epsilon/4$$

and

$$P\left(\left(\bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\right) \leftrightarrow \exists x_1,\dots,x_s\eta(x_1,\dots,x_t,\vec{s'})\,|\omega_n\right) > 1-\epsilon/4$$

by Paris and Vencovská (2015, Lemma 3.7),

$$P\left(\exists x_1,\dots,x_r\theta(x_1,\dots,x_r,\vec{t}) \wedge \exists x_1,\dots,x_s\eta(x_1,\dots,x_s,\vec{t'})\,|\omega_n\right) -$$

$$P\left(\bigvee_{i_1,\dots,i_r\le n} \theta(u_{i_1},u_{i_2},\dots,u_{i_r},\vec{t}) \wedge \bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\,|\omega_n\right) < \epsilon/2$$

But

$$P\left(\bigvee_{i_1,\dots,i_r\le n} \theta(u_{i_1},u_{i_2},\dots,u_{i_r},\vec{t}) \wedge \bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\,|\omega_n\right)$$

equals

$$P\left(\bigvee_{i_1,\dots,i_r\le n} \theta(u_{i_1},u_{i_2},\dots,u_{i_r},\vec{t})\,|\omega_n\right) \cdot P\left(\bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\,|\omega_n\right)$$

by induction hypothesis, and taking n large enough we have

$$P\left(\bigvee_{i_1,\dots,i_r\le n} \theta(u_{i_1},u_{i_2},\dots,u_{i_r},\vec{t})\,|\omega_n\right) \cdot P\left(\bigvee_{i_1,\dots,i_s\le n} \eta(u'_{i_1},u'_{i_2},\dots,u'_{i_s},\vec{t'})\,|\omega_n\right) -$$

$$P\left(\exists x_1,\dots,x_r\theta(x_1,\dots,x_r,\vec{t})\,|\omega_n\right) \cdot P\left(\exists x_1,\dots,x_s\eta(x_1,\dots,x_s,\vec{t'})\,|\omega_n\right) < \epsilon/2$$

and thus

$$\begin{aligned} &P\left(\exists x_1,\dots,x_r\theta(x_1,\dots,x_r,\vec{t}) \wedge \exists x_1,\dots,x_s\eta(x_1,\dots,x_s,\vec{t'})\,|\omega_n\right) \\ &\quad - P\left(\exists x_1,\dots,x_r\theta(x_1,\dots,x_r,\vec{t})\,|\omega_n\right) \cdot P\left(\exists x_1,\dots,x_s\eta(x_1,\dots,x_s,\vec{t'})\,|\omega_n\right) \\ &<\epsilon/2+\epsilon/2=\epsilon \end{aligned}$$

which gives the required result. ■

Corollary 23 (Zero-one law for constant-free sentences)**.** *Every constant-free sentence has measure 0 or 1.*

Proof: Paris and Vencovská (2015, Corollary 6.2) show the following: if probability function P on $S\mathcal{L}$ satisfies constant exchangeability and $P(\varphi\wedge\psi) = P(\varphi)\cdot P(\psi)$, whenever φ,ψ are quantifier free sentences of $\mathcal{L}$ that mention no constants in common, then $P(\varphi\wedge\psi) = P(\varphi)\cdot P(\psi)$ for any sentences φ,ψ of the language $\mathcal{L}$ which do not mention any constants in common.

Note that $P_=$ satisfies constant exchangeability and the assumption of Paris and Vencovská (2015, Corollary 6.2) is thus satisfied. Let φ be a sentence that does not mention any constant. Then φ,φ are two sentences that do not mention any constants in common. Since probability functions assign logically equivalent sentences the same probability we now easily find

$$P_=(\varphi) = P_=(\varphi\wedge\varphi) = P(\varphi)\cdot P(\varphi) \ .$$

So, $P_=(\varphi) = P_=(\varphi)^2$. This means that $P_=(\varphi)$ has to be either zero or one. ■

Hence, every inductively consistent constant-free sentence is an inductive tautology: $P_=(\varphi) > 0$ for constant-free φ implies that $P_=(\varphi) = 1$.

Proposition 24. *For all $\varphi\in S\mathcal{L}$ and all $n\geq N_\varphi$, $P_=(\varphi) = P_=(\varphi^n)$ and*

$$P_=\Big(\varphi\leftrightarrow\varphi^n\Big) = 1 \ .$$

Proof: Consider two sentences $\varphi,\psi\in S\mathcal{L}$ which mention at most the first $N := N_{\varphi\wedge\psi}$ constants. From Lemma 22 we obtain that for all $\omega_n\in\Omega_n$ it holds that

$$P_=(\varphi\wedge\psi|\omega_n) = P_=(\varphi|\omega_n)\cdot P_=(\psi|\omega_n) \ .$$

Using the trick on Paris and Vencovská (2015, P. 53) putting $\psi = \varphi$ we obtain $P_=(\varphi\wedge\varphi|\omega_n) = P_=(\varphi|\omega_n) = P_=(\varphi|\omega_n)^2$ and so

$$P_=(\varphi|\omega_n) = \begin{cases} 0 \\ 1 \ . \end{cases}$$

Using the definition of a conditional probability we find

$$P_=(\varphi\wedge\omega_n) = \begin{cases} 0, & \text{if and only if } P_=(\varphi|\omega_n) = 0 \\ P_=(\omega_n), & \text{if and only if } P_=(\varphi|\omega_n) > 0 \ . \end{cases} \tag{1}$$

So,

$$\begin{aligned} P_=(\varphi) =& P_=(\varphi\wedge\bigvee_{\omega_n\in\Omega_n}\omega_n) = P_=(\varphi\wedge\bigvee_{\substack{\omega_n\in\Omega_n\\ P_=(\varphi\wedge\omega_n)>0}}\omega_n) = P_=(\varphi\wedge\varphi^n) \\ =& \sum_{\substack{\omega_n\in\Omega_n\\ P_=(\varphi\wedge\omega_n)>0}} P_=(\varphi\wedge\omega_n) = \sum_{\substack{\omega_n\in\Omega_n\\ P_=(\varphi\wedge\omega_n)>0}} P_=(\omega_n) \overset{(1)}{=} P_=(\bigvee_{\substack{\omega_n\in\Omega_n\\ P_=(\varphi\wedge\omega_n)>0}}\omega_n) \\ =& P_=(\varphi^n) \ . \end{aligned} \tag{2}$$

So,

$$\begin{aligned} P_{=}(\neg\varphi \wedge \neg\varphi^n) =& P_{=}(\neg\varphi) + P_{=}(\neg\varphi^n) - P_{=}(\neg\varphi \vee \neg\varphi^n) \\ =& P_{=}(\neg\varphi) + 1 - P_{=}(\varphi^n) - 1 + P_{=}(\varphi \wedge \varphi^n) \\ =& P_{=}(\neg\varphi) \ . \end{aligned} \tag{3}$$

Finally, let us note that

$$\begin{aligned} P_{=}\Big(\varphi \leftrightarrow \varphi^n\Big) &= P_{=}\Big(\varphi \wedge \varphi^n\Big) + P_{=}\Big(\neg\varphi \wedge \neg\varphi^n\Big) \\ &\overset{(2) \text{ and } (3)}{=} P_{=}(\varphi) + P_{=}(\neg\varphi) = 1 \ . \end{aligned}$$

■

Note that the proportion $\frac{|\varphi^n|}{|\Omega_n|}$ of n-states in the n-support of a sentence φ eventually equals the measure of φ. This is because $\frac{|\varphi^n|}{|\Omega_n|} = P_{=}(\varphi^n) = P_{=}(\varphi)$ for $n \geq N_\varphi$.

Lemma 25. *If φ has positive measure, then $P_{=}(\cdot|\varphi) = P_{=}(\cdot|\varphi^n)$ for all $n \geq N_\varphi$.*

Proof: Notice that by Proposition 24:

$$P_{=}(\varphi \wedge \neg\varphi^n) = 0 = P_{=}(\neg\varphi \wedge \varphi^n).$$

Exploiting the law of total probability twice and the above observation we now find for all $\psi \in S\mathcal{L}$ that

$$\begin{aligned} P_{=}(\psi|\varphi) &= \frac{P_{=}(\psi \wedge \varphi)}{P_{=}(\varphi)} \\ &= \frac{P_{=}(\psi \wedge \varphi \wedge \varphi^n) + P_{=}(\psi \wedge \varphi \wedge \neg\varphi^n)}{P_{=}(\varphi^n)} \\ &= \frac{P_{=}(\psi \wedge \varphi \wedge \varphi^n)}{P_{=}(\varphi^n)} \\ &= \frac{P_{=}(\psi \wedge \varphi \wedge \varphi^n)) + P_{=}(\psi \wedge \neg\varphi \wedge \varphi^n))}{P_{=}(\varphi^n)} \\ &= \frac{P_{=}(\psi \wedge \varphi^n)}{P_{=}(\varphi^n)} = P_{=}(\psi|\varphi^n) \ . \end{aligned}$$

■

Corollary 26. *For all $k \geq 1$, $\models \varphi^{N_\varphi+k} \leftrightarrow \varphi^{N_\varphi}$ and $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi^{N_\varphi})$.*

Proof: By Lemma 25 for all $k \geq 0$, $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi)$. This entails $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi^{N_\varphi})$ for all $k \geq 1$.

Note that $\varphi^{N_\varphi+k}$ is quantifier free. Let χ, ψ be quantifier free and satisfiable, then the probability function $P_{=}(\cdot|\psi)$ is equal to the probability function $P_{=}(\cdot|\chi)$, if and only if ψ and χ are logically equivalent; clearly, if ψ and χ are logically equivalent, then these probability functions are equal. Furthermore, if ψ and χ are not logically equivalent, then without loss of generality ψ does not entail χ, and $P_=(\psi|\psi) = 1 > P_=(\chi|\psi)$ follows.

Letting $\psi = \varphi^N$ and $\chi = \varphi^{N_\varphi+k}$ we conclude that $\vDash \varphi^{N_\varphi+k} \leftrightarrow \varphi^{N_\varphi}$. ■

Corollary 27. *If $\omega_{N_\varphi+k} \vDash \varphi^{N_\varphi}$, then $\omega_{N_\varphi+k} \vDash \varphi^{N_\varphi+k}$.*

Every $N_\varphi+k$ state $\omega_{N_\varphi+k}$ extending a state in φ^{N_φ} is such that $P_=(\omega_{N_\varphi+k} \wedge \varphi) > 0$.

Proof: We let $N := N_\varphi$. Notice that by Corollary 26, if $\omega_N \in \Omega_N$ appears in φ^N, then any extension of ω_N to $\mathcal{L}_m$ (an m-state $\omega_m \in \Omega_m$ such that $\omega_m \vDash \omega_N$ with $m = N + k > N$) will appear in φ^m. To be more precise, for all $\omega_N \in \Omega_N$ with $\omega_N \vDash \varphi_N$ and for all $m \geq N$ and $\omega_m \in \Omega_m$, if $\omega_m \vDash \omega_N$, then $\omega_m \vDash \varphi^m$. To see this suppose $\omega_N \vDash \varphi^N$, $\omega'_m \vDash \omega_N$ but $\omega'_m \nvDash \varphi^m$. Then by definition of $P_=$ we have $P_=(\omega_N \,|\, \varphi^N), P_=(\omega'_m \,|\, \omega_N) \neq 0$. Then $0 < P_=(\omega'_m \,|\, \varphi^N) = P_=(\omega'_m \,|\, \varphi^m) = 0$, where the first equality is given by Corollary 26 and second equality is given by the assumption that $\omega'_m \nvDash \varphi^m$. ■

Consider a sentence ψ with zero measure, $P_=(\psi) = 0$. Intuitively, ψ is only true in few possible worlds.[4] One way to approach this intuition is by exploiting probability axiom P3 according to which the probability of a quantified sentence is the limit of probabilities of quantifier-free sentences. This suggests that—in the limit—only few n-states "converge" to ψ. So, if $P(\psi) = c > 0$, then P has to assign a joint probability of close to c to few n-states. That is, for n large enough, there exists set of n-states S_n, with joint probability of almost c, that is arbitrarily small in comparison to the number of all n-states. The following result, for which we are obliged to Alena Vencovská, makes this precise.

Lemma 28 (Concentration of probability on few n-states)**.** *Let ψ be such that $P_=(\psi) = 0$ and $P(\psi) = c > 0$, then for any $\epsilon > 0$ there exists some $M \in \mathbb{N}$ such that for all $m \geq M$ there exists a set of m-states, S_m, such that*

$$P(\bigvee_{\omega_m \in S_m} \omega_m) \geq (1 - \epsilon) \cdot c \textit{ and } \frac{|S_m|}{|\Omega_m|} < \epsilon \ .$$

Proof: First notice that if the result holds for some $m \in \mathbb{N}$ and a set of m-states S_m, then it also holds for the set of $m+1$ states S_{m+1} defined as the extensions of S_m to $\mathcal{L}_{m+1}$. Therefore, it is enough to show that result holds for some $m \in \mathbb{N}$.

Let $\mathcal{P} = \{P, P_=\}$. We first show that there exists some $m \in \mathbb{N}$ and a quantifier-free sentence $\chi \in S\mathcal{L}_m$ such that for all $Q \in \mathcal{P}$, $Q(\psi \leftrightarrow \chi) > 1 - \epsilon \cdot c$. (We can think of χ as a finite approximation of ψ.) We proceed by induction

[4]More precisely, consider the set of term structures for $\mathcal{L}$ that have a countably infinite domain. Then this means that the proportion of those term structures that satisfy ψ is negligible. But the term structures on a countably infinite domain can be determined as the limiting extensions of terms structures on finite subsets of the domain. This means that for asymptotically large n, there are only few term structures with a domain of size n that can be extended to a term structure that satisfies ψ. Then dividing the probability mass between the term structures on the full domain in such a way as to assign a probability of $c > 0$ to ψ should inevitably distribute a probability mass close to c between few term structures on a finite subdomain of size n for large n.

on the quantifier complexity; that is we proceed by induction on n for $\psi \in \Sigma_n$ and $\psi \in \Pi_n$.

For the base case, $n = 0$, ψ is quantifier free, and we can simply pick $\chi := \psi$.

For the induction step let $\psi = \forall \vec{x} \xi(x_1, \dots, x_r) \in \Pi_g$ with $\xi \in \Sigma_{g-1}$ be in prenex normal form. The case of $\psi = \exists \vec{x} \xi(\vec{x}) \in \Sigma_g$ is analogous.

By Paris and Vencovská (2015, Lemma 3.8) for all probability functions Q

$$Q(\psi) = \lim_{n \to \infty} Q(\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r})) \ .$$

Let $n \in \mathbb{N}$ be large enough such that for all $Q \in \mathcal{P}$

$$|Q(\psi) - Q(\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r}))| < \frac{\epsilon}{2} \cdot c.$$

Now let $Q \in \mathcal{P}$. Notice that ψ logically entails $\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r})$ and thus

$$\begin{aligned}
&Q(\psi \leftrightarrow \bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r})) = Q(\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r}) \to \psi) \\
&= Q(\neg \bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r})) + Q(\psi) - Q(\neg \bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r}) \wedge \psi) \\
&= 1 - Q(\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{k_1}, \dots, t_{k_r})) + Q(\psi) - 0 \\
&> 1 - \frac{\epsilon}{2} \cdot c \ . \qquad (4)
\end{aligned}$$

By the induction hypothesis, for each $k_1, \dots k_r \in \{1, \dots, n\}$ there is a quantifier free sentence $\lambda_{\vec{k}}(a_1, \dots, a_{M(\vec{k})}) \in S\mathcal{L}_{M(\vec{k})}$ such that for all $Q \in \mathcal{P}$

$$Q(\lambda_{\vec{k}} \leftrightarrow \xi(t_{k_1}, \dots, t_{k_r})) > 1 - \frac{\epsilon}{2n^r} \cdot c \ . \qquad (5)$$

Notice that

$$\begin{aligned}
&\neg(\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{\vec{k_i}}) \leftrightarrow \bigwedge_{j_1,\dots,j_r=1}^{n} \lambda_{\vec{j}}) \\
&=(\bigvee_{k_1,\dots,k_r=1}^{n} \neg\xi(t_{\vec{k_i}}) \wedge \bigwedge_{j_1,\dots,j_r=1}^{n} \lambda_{\vec{j}}) \vee (\bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{\vec{k_i}}) \wedge \bigvee_{j_1,\dots,j_r=1}^{n} \neg\lambda_{\vec{j}}) \\
&=(\bigvee_{k_1,\dots,k_r=1}^{n} (\neg\xi(t_{\vec{k_i}}) \wedge \bigwedge_{j_1,\dots,j_r=1}^{n} \lambda_{\vec{j}})) \vee (\bigvee_{j_1,\dots,j_r=1}^{n} (\neg\lambda_{\vec{j}} \wedge \bigwedge_{k_1,\dots,k_r=1}^{n} \xi(t_{\vec{k_i}}))) \\
&\vDash(\bigvee_{k_1,\dots,k_r=1}^{n} \neg\xi(t_{\vec{k_i}}) \wedge \lambda_{k_1,\dots,k_r}) \vee (\bigvee_{j_1,\dots,j_r=1}^{n} \xi(t_{\vec{k_j}}) \wedge \neg\lambda_{j_1,\dots,j_r}) \\
&= \bigvee_{k_1,\dots,k_r=1}^{n} \neg(\lambda_{k_1,\dots,k_r} \leftrightarrow \xi(t_{\vec{k_i}}))
\end{aligned}$$

where we write $\xi(\vec{t_{k_i}})$ for $\xi(t_{k_1},\dots,t_{k_r})$.

Then

$$
\begin{aligned}
&Q(\bigwedge_{k_1,\dots,k_r=1}^{n}\xi(t_{k_1},\dots,t_{k_r})\leftrightarrow\bigwedge_{j_1,\dots,j_r=1}^{n}\lambda_{\vec{j}})\\
=&1-Q(\neg(\bigwedge_{k_1,\dots,k_r=1}^{n}\xi(t_{k_1},\dots,t_{k_r})\leftrightarrow\bigwedge_{j_1,\dots,j_r=1}^{n}\lambda_{\vec{j}}))\\
\geq&1-Q(\bigvee_{k_1,\dots,k_r=1}^{n}\neg(\lambda_{k_1,\dots,k_r}\leftrightarrow\xi(t_{k_1},\dots,t_{k_r})))\\
\overset{(5)}{>}&1-n^r\frac{\epsilon}{2n^r}\cdot c\\
=&1-\frac{\epsilon}{2}\cdot c\ .
\end{aligned}
\tag{6}
$$

Let $\Xi=\bigwedge_{k_1,\dots,k_r=1}^{n}\xi(t_{k_1},\dots,t_{k_r})$, and $\Lambda=\bigwedge_{k_1,\dots,k_r=1}^{n}\lambda_{\vec{k}}$. Then by (4) and (6) we have

$$Q(\psi\leftrightarrow\Xi)>1-\frac{\epsilon}{2}\cdot c\ ,$$

and

$$Q(\Xi\leftrightarrow\Lambda)>1-\frac{\epsilon}{2}\cdot c\ .$$

And we have

$$
\begin{aligned}
Q(\psi\leftrightarrow\Lambda)&=Q(\psi\wedge\Lambda)+Q(\neg\psi\wedge\neg\Lambda)\\
&=Q(\psi\wedge\Lambda\wedge\Xi)+Q(\psi\wedge\Lambda\wedge\neg\Xi)+Q(\neg\psi\wedge\neg\Lambda\wedge\Xi)+Q(\neg\psi\wedge\neg\Lambda\wedge\neg\Xi)\\
&\geq Q(\psi\wedge\Lambda\wedge\Xi)-Q(\psi\wedge\Lambda\wedge\neg\Xi)-Q(\neg\psi\wedge\neg\Lambda\wedge\Xi)+Q(\neg\psi\wedge\neg\Lambda\wedge\neg\Xi)\ .
\end{aligned}
$$

Noticing that

$$Q(\psi\wedge\Lambda\wedge\Xi)=Q(\psi\wedge\Xi)-Q(\psi\wedge\neg\Lambda\wedge\Xi)$$

and

$$Q(\neg\psi\wedge\neg\Lambda\wedge\neg\Xi)=Q(\neg\psi\wedge\neg\Xi)-Q(\neg\psi\wedge\Lambda\wedge\neg\Xi)$$

we get

$$Q(\psi\leftrightarrow\Lambda)\geq Q(\psi\leftrightarrow\Xi)-Q(\neg(\Lambda\leftrightarrow\Xi))>1-\epsilon\cdot c\ . \tag{7}$$

Since (7) holds for all $Q\in\mathcal{P}=\{P_=,P\}$ and $P_=(\psi)=0$, we have

$$
\begin{aligned}
1-\epsilon\cdot c<P_=(\psi\leftrightarrow\Lambda)&=P_=(\psi\wedge\Lambda)+P_=(\neg\psi\wedge\neg\Lambda)\\
&=P_=(\neg\psi\wedge\neg\Lambda)=1-P_=(\psi\vee\Lambda)\leq 1-P_=(\Lambda)
\end{aligned}
$$

and thus $P_=(\Lambda)<\epsilon\cdot c\leq\epsilon$.

Now let $m=max\{M(\vec{k})\,|\,\vec{k}\in\{1,\dots,n\}^r\}$, then $\Lambda\in S\mathcal{L}_m$ and since Λ is quantifier free, there is a set of m-states S_m, such that

$$\vDash\Lambda\leftrightarrow\bigvee_{\omega_m\in S_m}\omega_m$$

and we have $\frac{|S_m|}{|\Omega_m|} = P_=(\Lambda) < \epsilon$. Note that S_m is the set of m-states entailing Λ.

Furthermore,

$$\begin{aligned} P(\bigvee_{\omega_m \in S_m} \omega_m) = P(\Lambda) \geq P(\Lambda \wedge \psi) &= P(\Lambda \wedge \psi) + P(\neg\Lambda \wedge \neg\psi) - P(\neg\Lambda \wedge \neg\psi) \\ &= P(\psi \leftrightarrow \Lambda) - P(\neg\Lambda \wedge \neg\psi) \\ &> 1 - \epsilon \cdot c - P(\neg\Lambda \wedge \neg\psi) \\ &\geq 1 - \epsilon \cdot c - P(\neg\psi) \\ &= P(\psi) - \epsilon \cdot c = c - \epsilon \cdot c = c \cdot (1 - \epsilon) \ . \end{aligned}$$

■

There is a sense in which the states in S_m *simulate* ψ on the sublanguage $\mathcal{L}_m$. Consider an underlying domain with m elements, $t_1, t_2, \ldots, t_m$. Universal (respectively, existential) quantification over a variable x can be understood as a finite conjunction (disjunction) over all finitely many elements. Replace all the quantifications in ψ by finite conjunctions and disjunctions over these m elements. On this finite domain, the resulting quantifier free sentence is equivalent to the original sentence. It is in this sense that S_m simulates ψ on a finite domain.[5]

The next Lemma shows that any maximal entropy function must assign probability one to the support φ^{N_φ} of φ (and thus to the n-support φ^n for $n \geq N_\varphi$). Note that this lemma does not prove the existence of a maximal entropy function.

Lemma 29. *Let $\varphi \in S\mathcal{L}$ with $P_=(\varphi) \in (0, 1]$. If $P \in \mathbb{E}$ with $P(\varphi^n) < 1$ for some $n \geq N_\varphi$, then $P_=(\cdot \,|\, \varphi^n)$ has greater entropy than P.*

Proof: Let $N := N_\varphi$. If $P_=(\varphi) = 1$, then $P_=(\cdot|\varphi^N) = P_= \in \mathbb{E}$. It suffices to recall that the equivocator has greater entropy than all other probability functions.

Now consider $0 < P_=(\varphi) < 1$.

Since φ^N and φ^n are logically equivalent for $n \geq N$ (Corollary 26) and since probability functions respect logical equivalence it follows the assumption that $P(\varphi^n) < 1$ that $P(\varphi^N) < 1$ holds.

So, let P be such that $P(\varphi) = 1$ and $P(\varphi^N) < 1$, then $P(\varphi \wedge \neg\varphi^N) = c > 0$ for some $1 \geq c > 0$. Let $\psi := \varphi \wedge \neg\varphi^N$ and notice that by definition of

---

[5] One might think that the following statement can play a similar role to that played by S_m in the above proof. Let $\varphi_0^n \overset{\text{df}}{=} \bigvee\{\omega \in \Omega_n : P_=(\omega \wedge \varphi) = 0, \not\models \neg(\omega \wedge \varphi)\}$, i.e., the disjunction of n-states deductively but not inductively consistent with φ. (If there are no such states, take φ_0^n to be an arbitrary contradiction on $\mathcal{L}_n$.)

Now suppose that φ has measure zero and that $P(\varphi) = c$. Since φ has measure zero, φ^n is a contradiction on $\mathcal{L}_n$. Hence,

$$c = P(\varphi) = P(\varphi \wedge \varphi^n) + P(\varphi \wedge \varphi_0^n) = P(\varphi \wedge \varphi_0^n),$$

so $P(\varphi_0^n) \geq c$. P must concentrate probability at least c on φ_0^n.

Thus the question arises as to whether $P_=(\varphi_0^n) \longrightarrow 0$ as $n \longrightarrow \infty$. This would imply that $\frac{|\varphi_0^n|}{|\Omega_n|} = P_=(\varphi_0^n) \longrightarrow 0$ as $n \longrightarrow \infty$, in which case φ_0^n would represent an increasingly negligible number of states.

However, it turns out that while this last condition holds true for some measure-zero φ, e.g., $\forall x U_1 x$, it does not hold true for all such sentences. For example, in the case of $\exists x \forall y U_2 xy$, which also has zero measure, $P_=(\varphi_0^n) = 1$ for all n.

φ^N, $P_=(\psi) = 0$. Let $\epsilon > 0$ and take M and S_M as given by Lemma 28 and let K_M be the set of M-states in $\Omega_M \setminus S_M$ such that $P_=(\varphi \wedge \omega_M) > 0$ for $M \geq N$. Corollary 26 shows that $|K_M| = |\varphi^N|\frac{|\Omega_M|}{|\Omega_N|}$, since all M-states $\omega_M \in \Omega_M$ extending an N-state in φ^N are such that $P_=(\varphi \wedge \omega_M) > 0$. Let $b_M = P(\bigvee_{S_M} \omega_M) \geq (1-\epsilon)c > 0$ and notice that since $P(\varphi) = 1$ we have $P(\bigvee_{K_M} \omega_M) = 1 - b_M$.

Then by convexity

$$
\begin{aligned}
&H_M(P) \leq -b_M \log\left(\frac{b_M}{|S_M|}\right) - (1-b_M)\log\left(\frac{1-b_M}{|K_M|}\right) \\
&= b_M \log(|S_M|) - b_M \log(b_M) + (1-b_M)\log(|K_M|) - (1-b_M)\log(1-b_M) \ .
\end{aligned}
$$

The M-entropy of $P_=(\cdot|\varphi^N)$ is

$$
\begin{aligned}
H_M(P_=(\cdot|\varphi^N)) &= -\sum_{\omega_M \vDash \varphi^N} \frac{1}{|K_M|}\log(\frac{1}{|K_M|}) = \log(|K_M|) \\
&= \log\left(|\varphi^N| \cdot \frac{|\Omega_M|}{|\Omega_N|}\right) \ . \qquad (8)
\end{aligned}
$$

We thus note

$$
\begin{aligned}
&\frac{H_M(P) - H_M(P_=(\cdot|\varphi^N))}{\log(|K_M|)} \\
&\leq b_M \log(|S_M| - |K_M|) - \frac{b_M \log(b_M) + (1-b_M)\log(1-b_M)}{\log(|K_M|)} + (1-b_M) - 1 \ .
\end{aligned}
$$

Now consider the three summands in turn. Since $\frac{|S_M|}{|K_M|} = \frac{|S_M|\cdot|\Omega_N|}{|\Omega_M|\cdot|\varphi^N|}$ becomes arbitrarily small by Lemma 28 and $1 \geq b_M > 0$, the first term is eventually less than zero. The second term goes to zero, since K_M increases without bounds. Finally, $b_M \geq (1-\epsilon)c > 0$. This means that for all large enough M it holds that $H_M(P) - H_M(P_=(\cdot|\varphi^N)) < 0$ and hence $H_M(P) < H_M(P_=(\cdot|\varphi^N))$. This entails that $P_=(\cdot|\varphi^N)$ has greater entropy than P. Thus, $P \notin \operatorname{maxent} \mathbb{E}_\varphi$.

In particular, we note for later use that the sequence $f_n(P) := H_n(P_=(\cdot|\varphi^N)) - H_n(P)$ is bounded from below by $\frac{b_M}{2} \geq \frac{(1-\epsilon)\cdot c}{2} > 0$ for all large enough n. ■

We are now in a position to present the main result of this section:

Theorem 30 (Agreement with Bayesian Conditionalisation)**.** *For all $\varphi \in S\mathcal{L}$ with $P_=(\varphi) \in (0,1]$ and all $n \geq N_\varphi$*

$$
\operatorname{maxent} \mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} = \{P_=(\cdot|\varphi^n)\} = \{P_=(\cdot|\varphi^{N_\varphi})\} \ .
$$

Proof: We prove a stronger property, namely that $P_=(\cdot|\varphi^N)$ has greater entropy than every other probability function $P \in \mathbb{E}$.

We let $N := N_\varphi$. Given the previous lemma, it suffices to show all $P \in \mathbb{E}$ with $P(\varphi^N) = 1$ and $P \neq P_=(\cdot|\varphi^N)$ have less entropy than $P_=(\cdot|\varphi^N)$. This means that $P_=(\cdot|\varphi^N)$ has greater entropy than all other $P \in \mathbb{E} \setminus \{P_=(\cdot|\varphi^N)\}$.

Consider first the case of $P_=(\varphi) = 1$. In this case, the equivocator $P_=$ is in $\mathbb{E}$, and, since it is the probability function in $\mathbb{P}$ with maximal entropy, it is the unique member of maxent $\mathbb{E}_\varphi$. By Lemma 25, the equivocator is also $P_=(\cdot|\varphi)$. By Lemma 29, it is $P_=(\cdot|\varphi^N)$.

Now consider $0 < P_=(\varphi) < 1$. By Lemma 25, $P_=(\varphi \,|\, \varphi^N) = P_=(\varphi \,|\, \varphi^n) = P_=(\varphi \,|\, \varphi) = 1$. This establishes the two last equalities in the statement of the theorem.

Let P be a probability function with $P(\varphi) = 1$, $P \in \mathbb{E}$. If $P(\varphi^N) < 1$, then $P_=(\cdot \,|\, \varphi^N)$ has greater entropy than P (Lemma 29).

If $P(\varphi^N) = 1$ but $P \neq P_=(\cdot|\varphi^N)$, then there is some $M \geq N$ such that for all $m \geq M$, $P \neq P_=(\cdot|\varphi^N)$ on Ω_m. For all $m \geq N$, $P_=(\cdot|\varphi^N)$ equivocates over all m-states in φ^m and has strictly greater m-entropy than every other probability function $Q \in \mathbb{P}$ with $Q(\varphi^m) = 1$ and $Q(\omega_m) \neq P_=(\omega_m|\varphi^n)$ for some m-state ω_m. This entails that for all $n \geq M$ it holds that $H_n(P) < H_n(P_=(\cdot|\varphi^N))$. Hence, $P_=(\cdot|\varphi^N)$ has greater entropy then every other probability function $P \in \mathbb{E}$ with $P(\varphi^N) < 1$.

So, $P_=(\cdot|\varphi^N)$ has greater entropy then every other probability function $P \in \mathbb{E}$. ■

Example 31. For the premiss sentence $\varphi = (\exists x \forall y Uxy \wedge Ut_1t_1) \vee (\forall x \exists y \neg Uxy \wedge \neg Ut_1t_1)$,

$$\text{maxent}\, \mathbb{E}_\varphi = \{P_=(\cdot|\neg Ut_1t_1)\} \ .$$

Proof: There is only one constant mentioned in φ, t_1. So, $N_\varphi = 1$. We here consider the simple case of the language containing only the relation symbol U. The general case follows from the fact the entropy maximisation is language invariant (Paris, 1994, Chapter 6). There are two 1-states, Ut_1t_1 and $\neg Ut_1t_1$. $\varphi \wedge Ut_1t_1$ is logically equivalent to $\exists x \forall y Uxy \wedge Ut_1t_1$ and $P_=(\exists x \forall y Uxy \wedge Ut_1t_1) \leq P_=(\exists x \forall y Uxy) = 0$. $\varphi \wedge \neg Ut_1t_1$ is logically equivalent to $\forall x \exists y \neg Uxy \wedge \neg Ut_1t_1$. For this sentence it holds that $P_=(\forall x \exists y \neg Uxy \wedge \neg Ut_1t_1) = 0.5$. Hence, $\varphi^{N_\varphi} = \varphi_1 = \neg Ut_1t_1$. φ_1 is the disjunction of all 1-states $\omega_1 \in \Omega_1$ such that $P_=(\varphi \wedge \omega_1) > 0$. Thus, $\varphi_1 = \neg Ut_1t_1$. ■

The following observation shows that the maximum entropy function not only has greatest entropy in the sense defined above, but also in a cumulative sense.

Corollary 32. *If φ has positive measure, then for all $P \in \mathbb{E}_\varphi \setminus \{P_=(\cdot|\varphi)\}$,*

$$\lim_{n\to\infty} \sum_{i=1}^{n} H_i(P_=(\cdot|\varphi)) - H_i(P) = \infty \ .$$

Proof: The proof shows a slightly stronger property: for all $P \in \mathbb{E}_\varphi \setminus \{P_=(\cdot|\varphi)\}$ the sequence $f_n(P) := H_n(P_=(\cdot|\varphi)) - H_n(P)$ is such that there exists some $M \geq N_\varphi$ such that $f_n(P)$ is strictly positive and never decreasing for all $n \geq M$.

Let us first consider the case that $P(\varphi^N) < 1$. The claim of this corollary follows directly from the final observation in the proof of Lemma 29.

The second and final case is when $P(\varphi^{N_\varphi}) = 1$. Since $P_=(\cdot|\varphi) \neq P$ there has to exist some $M \geq N_\varphi$ such that for all $m \geq M$ the probability functions P and $P_=(\cdot|\varphi)$ disagree on the quantifier free sentence of $\mathcal{L}_m$. Since both functions assign non-zero probability to, at most, the m-states extending those in φ^N and $P_=(\cdot|\varphi)$ is maximally equivocal on this set of M-states, it follows that $H_M(P_=(\cdot|\varphi)) > H_M(P)$.

For all $m \geq M$ we have:

$$
\begin{aligned}
H_m(P) &= -\sum_{\omega_m \in \Omega_m} P(\omega_m)\log(P(\omega_m)) \\
&\leq -\sum_{\omega_m \in \Omega_m} P(\omega_M)\frac{|\Omega_M|}{|\Omega_m|} \cdot \log\left(P(\omega_M) \cdot \frac{|\Omega_M|}{|\Omega_m|}\right) \\
&= -\sum_{\omega_M \in \Omega_M} P(\omega_M)\log\left(P(\omega_M) \cdot \frac{|\Omega_M|}{|\Omega_m|}\right) \\
&= H_M(P) + \log\left(\frac{|\Omega_m|}{|\Omega_M|}\right) \\
H_m(P_=(\cdot|\varphi)) &\overset{(8)}{=} \log\left(|\varphi^N| \cdot \frac{|\Omega_m|}{|\Omega_N|}\right) \\
&= \log(|\varphi^N|) + \log\left(\frac{|\Omega_m| \cdot |\Omega_M|}{|\Omega_N| \cdot |\Omega_M|}\right) \\
&= H_M(P_=(\cdot|\varphi)) + \log\left(\frac{|\Omega_m|}{|\Omega_M|}\right) \quad .
\end{aligned}
$$

It thus easily follows that $H_m(P_=(\cdot|\varphi)) - H_m(P) \geq H_M(P_=(\cdot|\varphi)) - H_M(P)$ for all $m \geq M$. In turn, this implies that

$$\lim_{n\to\infty} \sum_{M=1}^{n} H_i(P_=(\cdot|\varphi)) - H_i(P) \geq \lim_{n\to\infty} (n - M) \cdot (H_M(P_=(\cdot|\varphi)) - H_M(P)) \enspace .$$

Since the last difference is strictly positive, this limit is $+\infty$. The Corollary follows trivially by adding the first $M-1$ bounded terms to the above limit. ■

Given a finite set of premisses of the form $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ we showed in Theorem 11 how a maximal entropy function can be characterised in terms of an entropy limit point. In case of a single categorical premiss, φ, if $P_=(\cdot \mid \varphi)$ is an entropy limit point then it is the unique maximum entropy function (Corollary 15). In particular, this is the case when φ is equivalent to a quantifier free sentence (Corollary 19). Theorem 30 shows that for any inductively consistent premiss φ, there exists a unique maximal entropy function, which can be determined by conditionalising the equivocator on the support of φ, the quantifier free sentence φ^{N_φ} expressible in the sublanguage $\mathcal{L}_{N_\varphi}$. For example, for $\varphi = U_1t_1 \vee \exists x \forall y U_2xy$ every 1-state is consistent with φ. However, only the 1-states entailing U_1t_1 are in the support of φ. These 1-states have the feature that almost all their extensions contribute to the probability of $P_=(\varphi)$ via probability axiom P3. What is more, Theorem 30 shows that the maximal entropy probability function equivocates between the N_φ-states, and also between their extensions. That is, the unique maximal entropy probability function divides the full probability measure equally between these N_φ-states and similarly between their extensions to any $\mathcal{L}_n$ with $n \geq N_\varphi$.

Given Theorem 30, conditionalising the equivocator function is a simple method for determining the maximal entropy probabilities in objective Bayesian inductive logic. Although this approach to inductive logic is Bayesian, conditionalisation is not taken here as a principle that is constitutive or core to the Bayesian method, but rather as an inference tool that is appropriate in certain specific circumstances. Indeed, conditionalisation has been criticised as being problematic outside an appropriate range of circumstances (Howson, 2014; Williamson, 2010). The fact that it agrees with the maximal entropy approach can be taken to justify the use of conditionalisation on learning φ, in the circumstances in which φ has positive measure and is 'simple' in the sense that it only imposes the constraint $P(\varphi) = 1$ (Williamson, 2017, Definition 5.14).

6 Jeffrey Conditionalisation

In this section, we generalise our results for conditionalisation from the case in which the premiss is a categorical sentence φ to the case in which the premiss is a sentence of the language with a specific probability attached, φ^c, with $c \in (0,1)$. Thus in this section, $\mathbb{E} = \mathbb{E}_{\varphi^c} \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi) = c\}$.

Definition 33 (Jeffrey Update of the Equivocator)**.** Where $P_=(\varphi) \in (0,1)$ we can define the Jeffrey update of the equivocator function:

$$P_{\varphi^c}(\cdot) \stackrel{\text{df}}{=} c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi) \ .$$

First, we have a straightforward generalisation of Corollary 15:

Proposition 34. *If P_{φ^c} is an entropy limit point of $\mathbb{E}_{\varphi^c}$, then*

$$\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\} \ .$$

Proof: P_{φ^c} is contained in $\mathbb{E}_{\varphi^c}$ because $P_{\varphi^c}(\varphi) = c \cdot 1 + (1-c) \cdot 0 = c$. Hence, Theorem 12 applies. ∎

We also have an analogue of Corollary 16:

Proposition 35. *If $\mathbb{H}_n$ contains P_{φ^c} for sufficiently large n, then*

$$\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\} \ .$$

Proof: If $P_{\varphi^c} \in \mathbb{H}_n$ for sufficiently large n, then P_{φ^c} is an entropy limit point of $\mathbb{E}_{\varphi^c}$. Hence, Proposition 34 applies. ∎

Thus (cf. Corollary 19), if φ is quantifier-free and P_{φ^c} is well defined ($1 > P_=(\varphi) > 0$), then $\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\}$. Interestingly, as we show shortly, this holds true even when φ contains quantifiers. First we make the following observation:

Proposition 36. $\neg\varphi^n = (\neg\varphi)^n$.

Proof: Recall from (1) that for all $n \geq N_\varphi$ it is true that

$$P_=(\varphi \wedge \omega_n) = \begin{cases} 0, & \text{if and only if } P_=(\varphi|\omega_n) = 0 \\ P_=(\omega_n), & \text{if and only if } P_=(\varphi|\omega_n) > 0 \end{cases}$$

$$P_=(\neg\varphi \wedge \omega_n) = \begin{cases} 0, & \text{if and only if } P_=(\neg\varphi|\omega_n) = 0 \\ P_=(\omega_n), & \text{if and only if } P_=(\neg\varphi|\omega_n) > 0 \ . \end{cases}$$

Since $0 < P_=(\omega_n) = P_=(\varphi \wedge \omega_n) + P_=(\neg\varphi \wedge \omega_n)$ it follows that for every fixed n-state $\omega_n \in \Omega_n$ either $P_=(\varphi \wedge \omega_n) > 0$ or $P_=(\neg\varphi \wedge \omega_n) > 0$ is true but not both. Since φ^n is the disjunction of such ω_n, in particular φ^n is quantifier free, we have

$$\neg\varphi^n = (\neg\varphi)^n$$

and $\langle \varphi^n, (\neg\varphi)^n \rangle$ is a partition. ■

We are now in a position to provide the main result of this section.

Theorem 37 (Agreement with Jeffrey Conditionalisation)**.** *For all $c \in (0,1)$ and all $\varphi \in S\mathcal{L}$ such that $P_=(\varphi) \in (0,1)$, the maximal entropy function for the premiss φ^c is obtained by Jeffrey updating the equivocator function:*

$$\text{maxent}\,\mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\} = \{c \cdot P_=(\cdot|\varphi^{N_\varphi}) + (1-c) \cdot P_=(\cdot|\neg\varphi^{N_\varphi})\} \ .$$

Theorem 30 covers the borderline cases of $c = 0$ and $c = 1$ in which the maximum entropy function is unique and given by a Bayesian conditionalisation.

Proof: The main idea in the proof comes from the intuition that it is always beneficial in terms of entropy to take the probability mass from those n-states that have few extensions to m-states that simulate φ, as m increases to infinity, and divide it (equally) between the extensions of those n-states for which almost all extensions to an m-state simulate φ as m increases to infinity.

Let $N := N_\varphi$ and note that by Theorem 30

$$c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N) = c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi) \in \mathbb{E}_{\varphi^c} \ .$$

Furthermore, $cP_=(\cdot|\varphi) + (1-c)P_=(\cdot|\neg\varphi)$ has strictly greater entropy than every other function in $Q \in \mathbb{E}$ with $Q(\varphi^N) = c$ and $Q(\neg\varphi^N) = (1-c)$ because $cP_=(\cdot|\varphi) + (1-c)P_=(\cdot|\neg\varphi)$ assigns all n-states extending φ^N the same probability and it also assigns assigns all n-states extending $(\neg\varphi)^N = \neg\varphi^N$ (Proposition 36) the same probability.

The n-entropy of P_{φ^c} is given by:

$$\begin{aligned}
&H_n(c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi)) \\
&= - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi^N}} \frac{c \cdot |\varphi^N|}{|\Omega_n|} \cdot \log\left(\frac{c \cdot |\varphi^N|}{|\Omega_n|}\right) \\
&\quad - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \neg\varphi^N}} \frac{(1-c) \cdot |\neg\varphi^N|}{|\Omega_n|} \cdot \log\left(\frac{(1-c) \cdot |\neg\varphi^N|}{|\Omega_n|}\right) \\
&= - c \cdot \log(c) - (1-c) \cdot \log(1-c) + c \cdot H_n(P_=(\cdot|\varphi^N)) \\
&\quad + (1-c) \cdot H_n(P_=(\cdot|\neg\varphi^N)) \ .
\end{aligned}$$

Now consider some other $Q \in \mathbb{E}_{\varphi^c}$ with $Q(\varphi^N) \neq c$. We show that $Q \notin \operatorname{maxent} \mathbb{E}_{\varphi}$. W.l.o.g. we assume and let $\alpha := Q(\varphi^N) < c$ where $1 - \alpha = Q(\neg\varphi^N) > (1 - c)$. Then there has to exist some state $\nu_N \models \neg\varphi^N$ (recall that this means that $P_=(\nu_N \wedge \varphi) = 0$) such that $Q(\nu_N \wedge \varphi) > 0$. The $n \geq N$-entropy of Q – under the constraint that $Q(\varphi) = c$ without insisting on $Q \in \mathbb{P}$ – is maximised, if Q could equivocate on all φ^N states and all its extensions and, furthermore, equivocate probability mass $1 - c$ over $\neg\varphi^N$ and all its extension, and, furthermore, equivocate over all n-states in a set $S \subset \Omega_n$ of extensions $\neg\varphi^N$ as in Lemma 28.

So, overall,

$$
\begin{aligned}
H_n(Q) \leq & - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi^N}} \frac{\alpha \cdot |\varphi^N|}{|\Omega_n|} \cdot \log \left(\frac{\alpha \cdot |\varphi^N|}{|\Omega_n|} \right) \\
& - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \neg\varphi^N}} \frac{(1-c) \cdot |\neg\varphi^N|}{|\Omega_n|} \cdot \log \left(\frac{(1-c) \cdot |\neg\varphi^N|}{|\Omega_n|} \right) \\
& - \sum_{S \subset \Omega_n} \frac{(c-\alpha) \cdot |S|}{|\Omega_n|} \cdot \log \left(\frac{(c-\alpha) \cdot |S|}{|\Omega_n|} \right) \quad .
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& H_n(Q) - H_n(P_{\varphi^c}) \\
\leq & - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi^N}} \frac{\alpha \cdot |\varphi^N|}{|\Omega_n|} \cdot \log \left(\frac{\alpha \cdot |\varphi^N|}{|\Omega_n|} \right) - \sum_{S \subset \Omega_n} \frac{(c-\alpha) \cdot |S|}{|\Omega_n|} \cdot \log \left(\frac{(c-\alpha) \cdot |S|}{|\Omega_n|} \right) \\
& + \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi^N}} \frac{c \cdot |\varphi^N|}{|\Omega_n|} \cdot \log \left(\frac{c \cdot |\varphi^N|}{|\Omega_n|} \right) \quad .
\end{aligned}
$$

We proved that this expression is strictly less than zero for all large enough n in Theorem 30 for $c = 1$. The general case for $1 > c > \alpha > 0$ follows from (Landes et al., 2021, Proposition 5): Letting $H_n(P) > H_n(Q)$ for $P, Q \in \mathbb{P}$ and denoting by $c \cdot P$ the result of multiplying all probabilities of n-states by $1 > c > 0$, then and only then $H_n(c \cdot P) > H_n(c \cdot Q)$. This is so, since $H_n(c \cdot P)$ is an affine function of $H_n(P)$ with a strictly positive slope. It hence follows that $H_n(Q) - H_n(P_{\varphi^c}) < 0$.

Hence, $c \cdot P_=(\cdot|\varphi) + (1 - c) \cdot P_=(\cdot|\neg\varphi)$ has greater entropy than Q. This completes the proof. ■

Corollary 38 (Generalisation to sentence with intervals attached)**.** *For all intervals $\emptyset \neq X \subset [0, 1]$ and all sentences $\varphi \in S\mathcal{L}$ such that $P_=(\varphi) \in (0, 1)$ it holds that $c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N)$ has greater entropy than every other function in $\mathbb{E}_{\varphi^X}$ where $c := \arg\min_{x \in X} \ |x - P_=(\varphi)|$. Given the premiss φ^X, the maximal entropy function is obtained by Jeffrey conditionalisation of the equivocator on φ^c where c is closest to the probability of φ under the equivocator. Hence,*

$$
\operatorname{maxent} \mathbb{E}_{\varphi^X} = \{c \cdot P_=(\cdot|\varphi^N) + (1 - c) \cdot P_=(\cdot|\neg\varphi^N)\} \quad .
$$

Proof: If $P_=(\varphi) \in X$, then $c = \arg\min_{x\in X} \quad |x - P_=(\varphi)| = P_=(\varphi) = P_=(\varphi^N)$. Hence, for all sentences $\psi \in S\mathcal{L}$

$$\begin{aligned} P_{\varphi^c}(\psi) &= P_=(\varphi^N) \cdot P_=(\psi|\varphi^N) + P_=(\neg\varphi^N) \cdot P_=(\psi \wedge \neg\varphi^N) \\ &= P_=(\psi \wedge \varphi^N) + P_=(\psi|\neg\varphi^N) \\ &= P_=(\psi) \ . \end{aligned}$$

Since $P_= \in \mathbb{E}_{\varphi^X}$ the result follows.

If $P_=(\varphi) \notin X$, then for all $P \in \mathbb{E}_{\varphi^X}$ it holds that $x := P(\varphi) \neq P_=(\varphi)$. By the proof of Theorem 37 we see that $x \cdot P_=(\cdot|\varphi^N) + (1-x) \cdot P_=(\cdot|\neg\varphi^N)$ has greatest entropy among all functions with $x = P(\varphi)$. Hence,

$$\text{maxent}\,\mathbb{E}_{\varphi^X} \subseteq \{x \cdot P_=(\cdot|\varphi^N) + (1-x) \cdot P_=(\cdot|\neg\varphi^N) \ : \ x \in X\} \ .$$

Letting $P_x := x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)$, we now compute the n-entropies for all these probability functions for $n \geq N$ to be equal to

$$\begin{aligned} H_N(P_x) =& H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ =& - \sum_{\omega_N \models \varphi^N} x \cdot \frac{\varphi^N}{|\Omega_N|} \log\left(x \cdot \frac{|\varphi^N|}{|\Omega_N|}\right) \\ & - \sum_{\omega_N \models \neg\varphi^N} (1-x) \cdot \frac{|\Omega_N| - |\varphi^N|}{|\Omega_N|} \log\left((1-x) \cdot \frac{|\Omega_N| - |\varphi^N|}{|\Omega_N|}\right) \\ =& - x \cdot \log\left(\frac{x}{\frac{|\Omega_N|}{|\varphi^N|}}\right) - (1-x) \cdot \log\left(\frac{1-x}{\frac{|\Omega_N|}{|\Omega^N| - |\varphi^N|}}\right) \end{aligned}$$

and

$$\begin{aligned} H_n(P_x) =& H_n(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ =& - x \cdot \log\left(\frac{x}{\frac{|\Omega_N|}{|\varphi^N|} \cdot \frac{|\Omega_n|}{|\Omega_N|}}\right) - (1-x) \cdot \log\left(\frac{1-x}{\frac{|\Omega_N|}{|\Omega^N| - |\varphi^N|} \cdot \frac{|\Omega_n|}{|\Omega_N|}}\right) \\ =& H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ & - x \cdot \log\left(\frac{|\Omega_N|}{|\Omega_n|}\right) - (1-x) \cdot \log\left(\frac{|\Omega_N|}{|\Omega_n|}\right) \\ =& H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) + \log(|\Omega_n|) - \log(|\Omega_N|) \\ =& H_N(P_x) + \log(|\Omega_n|) - \log(|\Omega_N|) \ . \end{aligned}$$

It is hence holds for all $x, y \in [0,1]$ and all $n > N$ that

$$H_n(P_x) > H_n(P_y), \text{ if and only if } H_N(P_x) > H_N(P_y) \ . \tag{9}$$

Let us next note that $P_{P_=(\varphi^N)} = P_=$. Furthermore, every P_x is a convex combination of $P_=(\cdot|\varphi)$ and of $P_=(\cdot|\neg\varphi)$. Along this line from $P_=(\cdot|\varphi)$ to $P_=(\cdot|\neg\varphi)$ N-entropy is maximised by $P_{P_=(\varphi^N)} = P_=$ since it is the equivocator (on Ω_N). Since the P_x (on Ω_N) all are part of a line segment and H_N is strictly concave, it follows that N-entropy is uniquely maximised by the equivocator and strictly decreases the further one moves in one direction from the equivocator. Hence, P_c has strictly the greatest N-entropy among all other P_x for $x \in X \setminus \{c\}$.

Applying the above equivalence (9) we find that P_c (since $c \in X$ is the closest to $P_=(\varphi)$) also has the greatest n-entropy among all P_x for $x \in X$ for large enough n. P_c has hence greater entropy than every other probability function in $P \in \mathbb{E}_{\varphi^X} \setminus \{P_c\}$. ■

7 Preservation of Inductive Tautologies

Having developed the entropy limit point method for determining maximal entropy functions, and having demonstrated concordance with Bayesian conditionalisation and Jeffrey conditionalisation, we will now discuss the maximal entropy approach from a general perspective. In this section, we outline some logical features of objective Bayesian inductive logic, while in §8 we will explore the extent to which inferences are invariant under permutations of the constants, and in §9 we investigate some cases involving categorical premisses with zero measure.

First we show that, in objective Bayesian inductive logic, inductive tautologies (i.e., probability 1 inferences in the absence of any premisses) are preserved after learning the probability of any proposition that is inductively consistent:

Theorem 39 (Preservation of Inductive Tautologies, PIT)**.** *If* $\overset{\circ}{|\!\!\approx}\theta$ *and* $\not\overset{\circ}{|\!\!\approx}\neg\varphi$*, then* $\varphi^c \overset{\circ}{|\!\!\approx}\theta$ *for any* $c \in (0,1]$*.*

Proof: To simplify notation, we use $P^\dagger$ denote the unique probability function with maximal entropy if there exists such a function.

First, note that applying the assumption $P_=(\theta) = 1$ to $P_=(\theta\wedge\varphi)+P_=(\neg\theta\wedge\varphi) = P_=(\varphi)$ entails $P_=(\theta \wedge \varphi) = P_=(\varphi)$ for all sentences $\varphi \in S\mathcal{L}$.

If $c = 1$, then by Theorem 30,

$$P^\dagger(\theta) = P_=(\theta|\varphi) = \frac{P_=(\theta\wedge\varphi)}{P_=(\varphi)} = \frac{P_=(\varphi)}{P_=(\varphi)} = 1 \ .$$

So, $\varphi^1 \overset{\circ}{|\!\!\approx}\theta$.

If, on the other hand, $c \in (0,1)$, then by Theorem 37,

$$\begin{aligned} P^\dagger(\theta) &= c \cdot P_=(\theta|\varphi) + (1-c)\cdot P_=(\theta|\neg\varphi) = c + (1-c) \\ &= 1 \ . \end{aligned}$$

So, $\varphi^c \overset{\circ}{|\!\!\approx}\theta$. ■

PIT implies that inductive contradictions are also preserved after learning the probability of any proposition that is not an inductive contradiction: if $\overset{\circ}{|\!\!\approx}\neg\theta$ and $\not\overset{\circ}{|\!\!\approx}\neg\varphi$, then $\varphi^c \overset{\circ}{|\!\!\approx}\neg\theta$ for any $c \in (0,1]$.

PIT is loosely related to the Obstinacy principle of Paris (1994, p. 99), which provides a condition under which inferences from $\varphi_1^{X_1},...,\varphi_k^{X_k}$ are preserved upon learning $\pi_1^{W_1},...,\pi_l^{W_l}$. In the present setting, Obstinacy can be formulated as follows. Consider $\mathbb{E} = \{P : P \text{ satisfies } \varphi_1^{X_1},...,\varphi_k^{X_k}\}$ and $\mathbb{F} = \{P : P \text{ satisfies } \pi_1^{W_1},...,\pi_l^{W_l}\}$. Then:

Theorem 40 (**Obstinacy**). *If* $\text{maxent}\,\mathbb{E} \subseteq \mathbb{F}$*, then* $\text{maxent}\,\mathbb{E} \subseteq \text{maxent}(\mathbb{E} \cap \mathbb{F})$*.*

Proof: If $P \in \text{maxent}\,\mathbb{E}$ then no function in $\mathbb{E}$ dominates P in n-entropy for sufficiently large n. In particular, no function in $\mathbb{E} \cap \mathbb{F}$ dominates P in n-entropy for sufficiently large n. Thus, $P \in \text{maxent}(\mathbb{E} \cap \mathbb{F})$ and $\text{maxent}\,\mathbb{E} \subseteq \text{maxent}(\mathbb{E} \cap \mathbb{F})$. ■

PIT can also be thought of as a variant of the Rational Monotonicity rule of inference in non-monotonic logic (Lehmann and Magidor, 1992, §3.4):

Rational Monotonicity. If $\psi \approx\!\!\!| \; \theta$ and $\psi \not\approx\!\!\!| \; \neg\varphi$, then $\psi \wedge \varphi \approx\!\!\!| \; \theta$.

PIT specialises Rational Monotonicity to the case in which ψ is an inductive tautology and then generalises it to the case in which φ is uncertain.

PIT can also be interpreted as an absolute continuity condition (Billingsley, 1979, p. 422): if $\neg\theta$ has zero measure, i.e., $P_{=}(\neg\theta) = 0$, then any $P^{\dagger} \in \text{maxent}\,\mathbb{E}_{\varphi^c}$ also gives zero probability to $\neg\theta$, where φ has positive measure and $c > 0$. Note that the equivocator function $P_{=}$ corresponds to Lebesgue measure when probability functions on $\mathcal{L}$ are mapped to probability measures on the unit interval (Williamson, 2017, §2.6.3). Thus, 'zero measure' in the present sense (Definition 4) corresponds to zero Lebesgue measure.

8 Invariance under Permutations

Williamson (2010, Proposition 5.10) shows that the maximal entropy approach is invariant under those finite and infinite permutations of the atomic sentences that list atomic sentences involving only $t_1, ..., t_n$ before those involving t_{n+1} for each n. In this section, we explore invariance under permutations of the constants themselves.

Definition 41. Let f be a reordering of constants, i.e, f is bijective. For $\varphi \in S\mathcal{L}$ we write $f(\varphi)$ for the result of reordering the constants in φ according to f. We use $f(P)$ to denote the probability function obtained from P by permuting the constants of $\varphi \in S\mathcal{L}$ according to f: $f(P)(\varphi(\vec{t})) := P(\varphi(f(\vec{t})))$ for all $\varphi \in S\mathcal{L}$.

Lemma 42. *If* $P \in \mathbb{P}$ *and* f *is a permutation, then* $f(P) \in \mathbb{P}$*.*

Proof: It is clear that $f(P)$ satisfies P1 and P2.

Concerning P3, we need to show the next equality, the latter equalities follow from the definition of $f(P)$.

$$
\begin{aligned}
f(P)(\exists x\theta(x,\vec{t})) &= \sup_m f(P)(\bigvee_{i=1}^{m} \theta(t_i,\vec{t})) \\
&= \sup_m P(f(\bigvee_{i=1}^{m} \theta(t_i,\vec{t}))) \\
&= \sup_m P(\bigvee_{i=1}^{m} f(\theta(t_i,\vec{t}))) \\
&= \sup_m P\left(\bigvee_{i=1}^{m} \left(\theta\left(\frac{f(t_i), f(\vec{t})}{t_i, \vec{t}}\right)\right)\right) \quad .
\end{aligned}
$$

As usual, put $N := \max\{i : t_i \in \theta(\vec{t})\}$ and also let $N_f := \max\{j : t_j \in f(\theta(\vec{t}))\}$.

Let us now fix m and consider $M_m \geq \max\{f(1), \dots, f(m), N_f\}$, then $\bigvee_{i=1}^{m} \theta\left(\frac{f(t_i), f(\vec{t})}{t_i, \vec{t}}\right) \vDash \bigvee_{i=1}^{M_m} \theta(t_i, f(\vec{t}))$ and thus

$$P\left(\bigvee_{i=1}^{m} \theta\left(\frac{f(t_i), f(\vec{t})}{t_i, \vec{t}}\right)\right) \leq P(\bigvee_{i=1}^{M_m} \theta(t_i, f(\vec{t}))) \ .$$

Similarly, let $J_m \geq \max\{f^{-1}(1), \dots, f^{-1}(m), N_f\}$, then $\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t})) \vDash \bigvee_{i=1}^{J_m} f(\theta(t_i, \vec{t}))$ and so

$$P\left(\bigvee_{i=1}^{J_m} \theta\left(\frac{f(t_i), f(\vec{t})}{t_i, \vec{t}}\right)\right) \geq P(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))) \ .$$

We next note that $(P(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))))_{m \in \mathbb{N}}$ is an increasing non-negative sequence which converges by P3 to $P(\exists x \theta(x, f(\vec{t})))$. This entails that $\sup_m f(P)(\bigvee_{i=1}^{m} \theta(t_i, \vec{t}))$ also converges to $P(\exists x \theta(x, f(\vec{t})))$.

So,

$$\begin{aligned}
\sup_m f(P)(\bigvee_{i=1}^{m} \theta(t_i, \vec{t})) &= \sup_m P(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))) \\
&= P(\exists x \theta(x, f(\vec{t}))) \\
&= f(P)(\exists x \theta(x, \vec{t})) \ ,
\end{aligned}$$

where the last equality is just definition of $f(P)$. Hence, $f(P)$ satisfies P3. ■

The concept of 'greater entropy' is well defined in the sense that it is preserved under any permutation that preserves the probability functions that it permutes:

Proposition 43 (Independence of ordering of constant symbols)**.** *For any reordering of constants f and probability functions P, Q such that $f(P) = P$ and $f(Q) = Q$, P has greater entropy than Q, if and only if $f(P)$ has greater entropy than $f(Q)$.*

Proof: If $f(P) = P$ and $f(Q) = Q$ then $H_n(P) = H_n(f(P))$ and $H_n(Q) = H_n(f(Q))$. So, $H_n(P) > H_n(Q)$, if and only if $H_n(f(P)) > H_n(f(Q))$. Hence, P has greater n-entropy than Q for sufficiently large n, if and only if $f(P)$ has greater n-entropy than $f(Q)$ for sufficiently large n. ■

On the other hand, if a permutation f changes the two probability functions of interest, then the permuted functions can compare differently with respect to which has greater entropy:

Proposition 44 (Dependence on ordering of constant symbols)**.** *There exists an infinite reordering of constants f and probability functions P, Q such that P has greater entropy than Q but $f(Q)$ has greater entropy than $f(P)$.*

Proof: To simplify matters we consider a language only containing a single relation symbol, U, which is unary. It is apparent from the proof that the proof strategy applies to all languages in our sense.

Let f be the following bijection on $\mathbb{N}$. $f(2n+1) := 2n-1$ for all $n \geq 1$, $f(1) = 2$ and $f(2n) = 2n+2$. Intuitively, the even numbers and 1 are postponed to the future and the odd numbers, with the exception of 1 are brought forward.

It is important in the following that for all n it holds that f is not a bijection on $\{1, \ldots, n\}$. For all even n and $n = 1$ it holds that $f(n) > n$. For all other odd n it holds that $f^{-1}(n) = n + 2 > n$. This fact will be used without further mention.

Next define a probability function $P \in \mathbb{P}$ by having all constant symbols be independent of each other. This entails that n-entropies can be written as a sum of n 1-entropies. This follows from, for example, (Landes and Williamson, 2016, Equation 1).

For all $n \geq 1$ we now let

$$P(Ut_1) := \frac{1}{2} \qquad P(Ut_{2n}|\omega_{2n-1}) := \frac{1}{2} \qquad P(Ut_{2n+1}) := 1 \ ,$$

whenever $P(\omega_{2n-1}) > 0$.

We can then compute the n-entropies as follows for all $n \geq 1$

$$H_1(P) = \log(2) \qquad H_{2n}(P) = (n+1)\log(2)$$
$$H_{2n+1}(P) = H_{2n}(P) = (n+1)\log(2) \ ,$$

since the even n are maximally equivocal as is 1 and all other odd n are deterministic.

Ignoring the constant factor ($\log(2)$), the n-entropies of P can then be represented by the sequence $\langle 1, 2, 2, 3, 3, \ldots \rangle$. Figuratively speaking, the individual levels of n-entropy increase $(H_{n+1}(P) - H_n(P))$ are represented by $\langle 1, 1, 0, 1, 0, 1, 0, 1, 0, \ldots \rangle$. 0 here represents a deterministic behaviour and 1 represents a fully equivocal behaviour.

We now compute the n-entropies as follows for all $n \geq 1$

$$H_1(f(P)) = 0 \quad H_{2n}(f(P)) = n\log(2) \quad H_{2n+1}(f(P)) = H_{2n}(f(P)) = n\log(2) \ .$$

Clearly, for all $n \geq 1$ it holds that $H_n(P) > H_n(f(P))$.

Figuratively speaking, the individual levels have the following entropies for $f(P)$ ignoring the constant factor ($\log(2)$): $\langle 0, 1, 0, 1, 0, 1, 0, 1, 0, \ldots \rangle$ and n-entropies $\langle 0, 1, 1, 2, 2, 3, 3, 4, 4, \ldots \rangle$. Clearly, this last sequence is pointwise smaller than the corresponding sequence for P.

Now define a probability function Q which also makes all constant symbols independent of each other. Implicitly define Q by

$$H_1(Q) = 0.6 \cdot \log(2) \qquad H_2(Q) = 1.2 \cdot \log(2) \quad H_3(Q) = 1.8 \cdot \log(2)$$
$$H_{n+1}(Q) = H_n(Q) + 0.5 \cdot \log(2)$$

for all $n \geq 4$. That is, we need to find a value $Q(Ut_i)$ such that

$$-Q(Ut_i)\log(Q(Ut_i)) - (1 - Q(Ut_i))\log(1 - Q(Ut_i)) = \alpha \log(2) \ ,$$

where $\alpha \in \{0.5, 0.6\}$.

We note that Q is well defined under the assumption that $Q(Ut_i) \leq 0.5$ since i) 1-entropy is strictly concave and strictly increasing for $Q(Ut_i) \in [0, 0.5]$, ii) $H_1(P) \in [0, \log(2)]$ for all $P \in \mathbb{P}$, iii) H_1 is continuous, iv) H_1 is a bijective map from $[0, 0.5]$ onto $[0, \log(2)]$ and finally v) the intermediate value theorem holds.

Apparently, for $i \in \{1, 2, 3\}$ it holds that $H_i(P) > H_i(Q)$. That $H_i(P) > H_i(Q)$ holds for all greater i, too, follows from the definition of Q.

Figuratively speaking, the individual levels have the following entropies for Q ignoring the constant factor $(\log(2))$: $\langle 0.6, 0.6, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, \ldots\rangle$ and n-entropies $\langle 0.6, 1.2, 1.8, 2.3, 2.8, 3.3, \ldots\rangle$. Clearly, this last sequence is pointwise smaller than the corresponding sequence for P.

We compute the n-entropies for $f(Q)$ as follows for all $1 \leq n \leq 4$

$$H_1(f(Q)) = 0.6 \cdot \log(2) \qquad H_2(f(Q)) = 1.2 \cdot \log(2)$$
$$H_3(f(Q)) = 1.7 \cdot \log(2) \qquad H_4(f(Q)) = 2.3 \cdot \log(2) \ .$$

For all larger $n \geq 5$ we observe

$$H_n(f(Q)) = H_4(f(Q)) + \frac{(n-4)}{2} \cdot \log(2) > 0 \ .$$

Figuratively speaking, the individual levels have the following entropies for Q ignoring the constant factor $(\log(2))$: $\langle 0.6, 0.6, 0.5, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, \ldots\rangle$ and n-entropies $\langle 0.6, 1.2, 1.7, 2.3, 2.8, 3.3, \ldots\rangle$. Clearly, this last sequence is pointwise greater than the corresponding sequence for $f(P)$. ■

The proof shows in fact that for any language there exist probability functions $P, Q \in \mathbb{P}$ such that P has greater entropy than $f(P)$ and P has greater entropy than Q, yet $f(Q)$ has greater entropy than $f(P)$.

Interestingly, despite the possibility exposed by Proposition 44, our results show that in many natural cases, the function that has *maximal* entropy is invariant under reordering the constants:[6]

Theorem 45 (Invariance under Permutations of Constant Symbols)**.** *If* $1 > P_{=}(\varphi) > 0$ *and* $0 < c \leq 1$*, then for* $\{P^\dagger\} = \operatorname{maxent} \mathbb{E}_{\varphi^c}$ *and* $\{P^\dagger_f\} = \operatorname{maxent} \mathbb{E}_{f(\varphi)^c}$ *it holds that for all* $\psi \in S\mathcal{L}$ *that*

$$P^\dagger(\psi) = P^\dagger_f(f(\psi)) \ .$$

Proof: Let us first recall that by Lemma 42 we have $f(P) \in \mathbb{P}$. Furthermore, from the definition of $f(P)$ we immediately obtain that $P \in \mathbb{E}_{\varphi^c}$, if and only if $f(P) \in \mathbb{E}_{f(\varphi)^c}$.

[6] Landes et al. (2021, Footnote 2) show that Paris' approach to maximising entropy, which appeals to limits of entropy maximisers on finite languages, is invariant under finite and infinite permutations of constant symbols where it is well defined. They demonstrate that Paris' approach agrees with the maximal entropy approach in many cases, and conjecture that this agreement extends to all cases in which Paris' limiting function is well defined. In all such cases, invariance of this limit function implies that the maximal entropy approach is invariant under permutations of constants.

After observing that $\models f(\varphi^m) \leftrightarrow f(\varphi)^m$ and that $\models f(\neg\varphi^m) \leftrightarrow f(\neg\varphi)^m$ for all large enough m, we apply Theorem 37 and find

$$\begin{aligned}
P^\dagger =& c \cdot P_=(\cdot|\varphi^m) + (1-c) \cdot P_=(\cdot|\neg\varphi^m) \\
P^\dagger_f =& c \cdot P_=(\cdot|f(\varphi^m)) + (1-c) \cdot P_=(\cdot|f(\neg\varphi^m)) \ .
\end{aligned}$$

It now suffices to note that the equivocator function is as symmetrical as can be: for all $\chi, \rho \in QFS\mathcal{L}$ it holds that

$$P_=(\chi|\rho) = P_=(f(\chi)|f(\rho)) \ .$$

Hence $P^\dagger(\chi) = P^\dagger_f(f(\chi))$ for all quantifier free sentences $\chi \in QFS\mathcal{L}$. Gaifman's Theorem Gaifman (1964) then delivers the result that $P^\dagger(\cdot) = P^\dagger_f(f(\cdot))$.
■

As expected this result generalises easily to a single premiss with an attached uncertainty interval.

Corollary 46. *If* $1 > P_=(\varphi) > 0$ *and interval* $\emptyset \neq X \subset [0,1]$, *then for* $\{P^\dagger\} = \text{maxent}\,\mathbb{E}_{\varphi^X}$ *and* $\{P^\dagger_f\} = \text{maxent}\,\mathbb{E}_{f(\varphi)^X}$ *it holds that for all* $\psi \in S\mathcal{L}$ *that*

$$P^\dagger(\psi) = P^\dagger_f(f(\psi)) \ .$$

Proof: For both premisses a unique maximum entropy function exists which is equal to a Jeffrey update of the equivocator. These Jeffrey (or simply Bayesian) updates are with respect to φ^{N_φ}, respectively, the logically equivalent $f(\varphi^{N_\varphi})$ and $(f(\varphi))^{N_\varphi}$. Furthermore, both Jeffrey updates are with respect to the same $x \in X$ (Corollary 38).

Finally, let us apply the proof of Theorem 45 to note that for all $\psi \in S\mathcal{L}$ it holds that

$$P^\dagger(\psi) = P^\dagger_f(f(\psi)) \ .$$

■

9 Zero Measure Premisses

As Example 13 illustrates, there are cases of zero-measure premisses that are entirely unproblematic and that can be handled using the entropy limit point techniques introduced in §3.[7] However, some zero-measure premisses are more problematic, in that they generate sets $\mathbb{E}$ of probability functions in which there is no function with maximal entropy. We will focus on these pathological cases in this section. We first provide some examples of such cases and then we discuss how best to proceed when they arise. We argue that these cases suggest a refinement to the definition of maximal entropy and that they motivate drawing inferences from any function with sufficiently great entropy.

[7] More generally, if φ is a universally quantified claim about a conjunction of literals then it has zero measure but can be handled straightforwardly (Landes et al., 2021).

To simplify the exposition in this section we assume in this section that the underlying language $\mathcal{L}$ contains only the single relation symbol employed in the respective propositions. The general case follows from the fact entropy maximisation is language invariant (Paris, 1994, Chapter 6), because maximal entropy functions equivocate over all sentences mentioning only relation symbols that are not mentioned by any premiss.

Proposition 47. *For $\varphi = \exists x \forall y Uxy$ and any $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy than P. Hence,* $\operatorname{maxent} \mathbb{E}_\varphi = \emptyset$.

Proof: Suppose for contradiction that $\operatorname{maxent} \mathbb{E}_\varphi \neq \emptyset$ and let $P \in \operatorname{maxent} \mathbb{E}_\varphi$. We now show that this entails a contradiction. This is achieved by first defining a probability function $P' \in \mathbb{E}_\varphi \setminus \{P\}$ such that $H_n(P') \geq H_n(P)$ for all large enough n. It is not necessarily the case that P' has greater entropy than P. However, all probability functions that are a convex combination of P and P' are in $\mathbb{E}_\varphi$ ($\mathbb{E}_\varphi$ is convex) and have strictly greater n-entropy than P for all large enough n ($H_n(\cdot)$ is concave). Hence, all the convex combinations are in $\mathbb{E}_\varphi$ and have greater entropy than P. Contradiction.

Note that $P_=(\varphi) = 0 < 1 = P(\varphi)$. Hence, $P \neq P_=$. Let us now define a probability function $P' \in \mathbb{E}$ by shifting all witnessing of $\exists x \forall y Uxy$ by one and then adding a constant t_1 such that Ut_1t^* is independent from all other literals for all $t^* \neq t_1$. Intuitively, the literals $\pm Ut_it_k$ are replaced by $\pm Ut_{i+1}t_k$.

Formally, let $\omega_n \in \Omega_n = \bigwedge_{i,k=1}^n U^{\epsilon_{i,k}} t_i t_k$ be an arbitrary n-state. Then define P' by

$$\begin{aligned} P'(\omega_n) :&= P(\bigwedge_{i=2}^n \bigwedge_{k=1}^n U^{\epsilon_{i-1,k}} t_{i-1} t_k) \cdot P_=(\bigwedge_{k=1}^n U^{\epsilon_{1,k}} t_1 t_k) \\ &= \frac{P(\bigwedge_{i=1}^{n-1} \bigwedge_{k=1}^n U^{\epsilon_{i,k}} t_i t_k)}{2^n} \quad . \end{aligned}$$

Firstly, we note $P'(\forall y U t_1 y) = \lim_{n\to\infty} P'(\bigwedge_{k=1}^n U t_1 t_k) = \lim_{n\to\infty} 2^{-n} = 0$. So, according to P' the constant t_1 is not a witness of the existential premiss sentence φ.

We next show that $P \neq P'$. Firstly, note that since $\lim_{n\to\infty} P(\bigvee_{i=1}^n \forall y U t_i y) = P(\exists x \forall y Uxy) = 1$ the following minimum is a finite number and thus obtains $\min\{i \in \mathbb{N} : P(\forall y U t_i y) > 0\}$. Armed with this observation, we note that secondly and finally that

$$\begin{aligned} \min\{i \in \mathbb{N} \,:\, P'(\forall y U t_i y) > 0\} &= \min\{i \in \mathbb{N} \,:\, \lim_{n\to\infty} P'(\bigwedge_{k=1}^n U t_i t_k) > 0\} \\ &= \min\{i \in \mathbb{N} \,:\, \lim_{n\to\infty} P(\bigwedge_{k=1}^n U t_{i+1} t_k) > 0\} \\ &= 1 + \min\{i \in \mathbb{N} \,:\, \lim_{n\to\infty} P(\bigwedge_{k=1}^n U t_i t_k) > 0\} \\ &= 1 + \min\{i \in \mathbb{N} \,:\, P(\forall y U t_i y) > 0\} \quad . \end{aligned}$$

So, $P \neq P'$.

We also observe that for all $i \in \mathbb{N}$, $P'(\forall y U t_i y) = P(\forall y U t_{i+1} y)$ and furthermore $P'(\bigvee_{i\in I} \forall y U t_i y) = P(\bigvee_{i\in I} \forall y U t_{i+1} y)$ for all finite index sets I. So,

$$\begin{aligned}
P'(\exists x \forall y U x y) &= \lim_{n\to\infty} P'(\bigvee_{i=1}^{n} \forall y U t_i y) \\
&\geq \lim_{n\to\infty} P'(\bigvee_{i=2}^{n} \forall y U t_i y) \\
&= \lim_{n\to\infty} P(\bigvee_{i=1}^{n-1} \forall y U t_i y) \\
&= \lim_{n\to\infty} P(\bigvee_{i=1}^{n} \forall y U t_i y) \\
&= 1 \ .
\end{aligned}$$

This means that $P'(\exists x \forall y U x y) = 1$ and thus, as advertised, $P' \in \mathbb{E}_\varphi$.

We now calculate n-entropies of P and P' and find for $n \geq 1$ that:

$$\begin{aligned}
H_n(P) = -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\leq r\leq n\\ 1\leq s\leq n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\leq u\leq n}} & P(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k) \\
& \cdot \log(P(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k))
\end{aligned}$$

$$\begin{aligned}
H_n(P') &= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\leq r,s\leq n}} P'(\bigwedge_{i,k=1}^{n} U^{\epsilon_{i,k}} t_i t_k) \cdot \log(P'(\bigwedge_{i,k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)) \\
&= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\leq r\leq n\\ 1\leq s\leq n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\leq u\leq n}} P'(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k) \\
&\qquad \cdot \log(P'(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)) \\
&= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\leq r\leq n\\ 1\leq s\leq n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\leq u\leq n}} \frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)}{2^n} \\
&\qquad \cdot \log(\frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)}{2^n}) \ .
\end{aligned}$$

Holding the first summation fixed, we note that since n-entropy is maximised by maximally equivocating that $H_n(P) \leq H_n(P')$. For example, if P is flat on $\bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_n t_k$, $P(\bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_n t_k) = 2^{-n}$ for all $\epsilon_{n,k}$ with $1 \le k \le n$, and all these conjunctions are independent of $\bigwedge_{i=1}^{n-1} \bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_i t_k$ for all ϵ, then $H_n(P) = H_n(P')$.

Now define $Q := \frac{P+P'}{2}$. Since $\mathbb{E}_\varphi$ is convex and $P, P' \in \mathbb{E}_\varphi$, we observe that $Q \in \mathbb{E}_\varphi$.

Since n-entropy is a strictly concave function we conclude that $H_n(Q) > H_n(P)$ whenever P and P' disagree on $\mathcal{L}_n$. Since $P \neq P'$ there has to exists some finite M and quantifier free sentence $\psi \in QFS\mathcal{L}_M$ such that $P(\psi) \neq P'(\psi)$ (Gaifman's Theorem). Since $\mathcal{L}_m \subset \mathcal{L}_{m+1}$ for all m we have that P disagrees with P' on $\mathcal{L}_m$ for all $m \ge M$. We have hence found a $Q \in \mathbb{E}$ such that $H_n(Q) > H_n(P)$ for all large enough n. Hence, $P \notin \operatorname{maxent} \mathbb{E}_\varphi$. Contradiction. ■

We generalise this result to higher quantifier complexity in Appendix 3. These results are summarised in the following theorem.

Theorem 48 (Zero Measure Premisses). *For all $n \ge 1$ and*

- $\varphi = \exists v_{2n} \forall v_{2n-1} \ldots \exists v_2 \forall v_1 U v_1 v_2 \ldots v_{2n} \in \Sigma_{2n}$ *or*
- $\varphi = \forall v_{2n+1} \ldots \exists v_2 \forall v_1 U v_1 v_2 \ldots v_{2n+1} \in \Pi_{2n+1}$,

it holds that for all $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy. Hence, $\operatorname{maxent} \mathbb{E}_\varphi = \emptyset$.

Having introduced some pathological cases in which there is no maximal entropy function, we now turn to the question as to what to do in such cases.

For simplicity of exposition, we focus on the case in which we have a single premiss, $\varphi = \exists x \forall y Uxy$, considered in Proposition 47. We call a proposition of the form $\forall y U t_i y$ a *witness proposition.* A probability function P that satisfies φ distributes probability 1 to the witness propositions, $\lim_{k\to\infty} P(\bigvee_{i=1}^{k} \forall y U t_i y) = P(\exists x \forall y U x) = 1$. We call a constant t_i a *witness* if P gives positive probability to the corresponding witness proposition $\forall y U t_i y$. Now, the equivocator function, which is the probability function with maximal entropy, gives φ measure zero, $P_=(\exists x \forall y U x) = 0$, and thus it has no witnesses. Given P, one can construct a function Q that has greater entropy than P by making Q 'closer to' the equivocator in one or both of two ways:

1. Delaying the witnesses. If there are infinitely many witnesses, then one can create Q by increasing the index of each witness in an appropriate way in order to make Q more like the equivocator than P for each fixed n. For example, if $t_{i_1}, t_{i_2}, \ldots$ are the witnesses for P, one can construct Q with witnesses $t_{i_2}, t_{i_3}, \ldots$, ensuring that $Q(\forall y U t_{i_1} y) = 0$ and $Q(\forall y U t_{i_j} y) = P(\forall y U t_{i_{j-1}} y)$ for each $j > 1$.

2. Flattening the distribution over witness propositions. Entropy can be increased by increasing the number of witnesses, if there are finitely many, and distributing probability more equally to the witnesses, decreasing the rate at which the probability of $\bigvee_{i=1}^{k} \forall y U t_i y$ converges to 1.

The approach taken in the proof of Proposition 47 involved a mixture of these strategies: delaying witnesses to give P', and then flattening the distribution by taking a convex combination of P and P', to yield Q.

One might argue that although the first of these two strategies increases n-entropy for sufficiently large n, it does not on its own lead to a function that is more equivocal in an intuitive sense. Hence, this seems to be a case in which the formal concept of maximal entropy fails to adequately explicate the concept of being maximally equivocal. (In contrast, the second strategy is unproblematic: flattening the distribution over witness propositions does seem to be a genuine way of generating a more equivocal probability function.)

The explication of maximal entropy can however be refined to avoid this problem: we can deem P to have greater entropy than Q just when, for every reordering f of the constants that do not appear in the premisses, $f(P)$ dominates $f(Q)$ in n-entropy for sufficiently large n. Note that this refinement relativises the greater-entropy relation to the premisses.

This refinement eradicates the first of the two strategies: delaying witnesses no longer increases entropy, because there are reorderings with respect to which the witnesses are not delayed. The refinement leaves intact the second kind of strategy.

If we accept this refinement, the question then becomes: what policy should be adopted when there is no maximal entropy function because of increases in entropy of the second kind?

Williamson (2010, pp. 29–30) suggests a pragmatic policy: to take inferences to be determined by probability functions with *sufficiently* great entropy. Here, the cut-off between functions that have sufficiently great entropy and those that do not may depend on features of the problem or on the users of the logic, and may not be precise. Choosing a probability function with sufficiently great entropy amounts to a choice of P such that $P(\bigvee_{i=1}^{k} \forall y U t_i y)$ converges to 1 sufficiently slowly.

Further desiderata may be imposed. For example, one might suggest equivocating between the constants by treating them equally. The thought here is that each constant should be a witness, because the premiss gives no grounds for discriminating between constants that are witnesses and those that are not. This line of reasoning motivates giving each witness proposition the same probability $s > 0$ and making witness propositions probabilistically independent.[8] In which case, $P(\bigvee_{i=1}^{k} \forall y U t_i y) = 1 - (1-s)^k$, which converges to 1 as required. Now, decreasing s (and distributing the corresponding probability equally amongst n-states) will lead to a probability function with greater entropy—this is an application of the second of the two strategies outlined above. The pragmatic policy then amounts to drawing inferences from probability functions that correspond to values of s that are sufficiently small. One approach here is to take s to be sufficiently small just when taking s any smaller would not make a significant difference with respect to practical purposes.

In sum, we see that although these pathological examples require refinements to the overall approach, there is scope to devise policies that allow one to extend objective Bayesian inductive logic even to these difficult measure-zero cases.

[8]Such a distribution fits well with the maximal entropy approach, since it encapsulates symmetry and independence properties that have been used to motivate entropy maximisation (Paris and Vencovská, 1997; Paris, 1998).

10 Conclusion

Objective Bayesian inductive logic defines inductive entailment from a set of (possibly probabilistic) premisses in terms of maximal entropy probability functions that satisfy the given premisses. To be more precise, a set of premisses inductively entails a conclusion if every probability function with maximal entropy that satisfies the premisses also satisfies the conclusion. This is a very natural approach to inductive logic that has been studied extensively in the literature in the context of reasoning with propositional languages. An immediate task that arises with this approach is then to find these maximal entropy probability functions in order to perform inference. This is a straightforward, although possibly computationally expensive, problem when working with propositional languages. For more expressive languages, however, it is not clear how one should proceed to determine these maximal entropy probability functions. In this paper, we have studied this problem for premisses and conclusion that are given in terms of constraints on the probabilities of sentences of a first order language.

To do so we first introduced the notion of an *entropy limit point* and discussed its use for determining maximal entropy probability functions. Next we distinguished what we call the measure-zero sentences from those that have positive measure. Measure-zero sentences are sentences that are assigned probability zero by the equivocator function $P_=$. Intuitively, measure-zero sentences are those that have very few models. To be more precise, these are sentences for which the proportion of term structures with a countably infinite domain that satisfy them is negligible. We showed that for categorical premisses with positive measure, the maximal entropy approach agrees with Bayesian conditionalisation. This then generalizes to Jeffrey conditionalisation when dealing with a non-categorical premiss that is given in terms of a constraint on the probability of some sentence. With these results in place we then showed that although the comparative entropy of probability function does in general depend on the ordering of constants in the language, the probability functions with maximal entropy remain invariant under such permutations in the cases where it agrees with Bayesian of Jeffrey conditionalisation.

These results not only clarify which probabilities the maximal entropy probability functions assign for inductive inference but also give a constructive method for calculating the maximal entropy probabilities. On the one hand, this shows that the maximal entropy approach agrees with standard conceptions of baseline rationality. On the other, it witnesses the stability and generality of Bayesian conditionalisation as a process of probabilistic learning.

Finally, we turned our attention to inference from zero-measure premisses and identified a certain class of zero-measure sentences for which there is no maximal entropy probability function. This leaves the question of inductive inference from these pathological zero-measure premisses open. The issue is then to understand which inferences from zero measure premisses are rational and how to systematically characterize such inferences in terms of a unified inference process, and we developed a strategy for doing this.

Another interesting question is what more can be said about inductive inference from multiple non-categorical premisses. So far, our results on objective Bayesian inductive logic have concerned languages containing only relation symbols. It is natural to extend these considerations to languages also containing a symbol for equality and function symbols, which have already been studied in Pure Inductive Logic (Landes, 2009; Landes et al., 2009; Paris and Vencovská, 2015; Howarth and Paris, 2019; Paris and Vencovská, 2019). Finally, our hope here is that these results can also suggest new avenues for investigating the open cases of the *entropy limit conjecture* that concerns the equivalence of the two main approaches to inductive inference introduced in §1.

Acknowledgements

Blinded for review.

Appendix 1. Proofs of Proposition 11 and Theorem 12

First let us recount some basic information-theoretic facts.

The *n-divergence* of two probability functions P and Q is defined as the Kullback-Leibler divergence of P from Q on $\mathcal{L}_n$:

$$d_n(P,Q) \stackrel{\text{df}}{=} \sum_{\omega\in\Omega_n} P(\omega)\log\frac{P(\omega)}{Q(\omega)}.$$

A Pythagorean theorem holds for the n-divergence d_n (Cover and Thomas, 1991, Theorem 11.6.1):

$$d_n(P,Q) \geq d_n(P,R_n) + d_n(R_n,Q),$$

for any convex $\mathbb{F} \subseteq \mathbb{P}$, if $P \in \mathbb{F}$ and $Q \notin \mathbb{F}$, where $R_n \in \arg\inf_{S\in\mathbb{F}} d_n(S,Q)$.

Consequently, for any $P \in \mathbb{E}$ and $Q_n \in \mathbb{H}_n$ (Landes et al., 2021, corollary 32):

$$H_n(Q_n) - H_n(P) \geq d_n(P,Q_n).$$

Pinsker's inequality connects the L_1 distance to n-divergence (see, e.g., Cover and Thomas, 1991, Lemma 11.6.1):

$$d_n(P,Q) \geq \frac{1}{2}\|P-Q\|_n^2.$$

Proposition 11. If P is an entropy limit point of $\mathbb{E}$ then there are $Q_n \in \mathbb{H}_n$ such that $\|Q_n - P\|_n \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof: Putting our last two information-theoretic facts together we have that

$$\begin{aligned} H_n(Q_n) - H_n(P) &\geq d_n(P,Q_n) \\ &\geq \frac{1}{2}\|P-Q_n\|_n^2, \end{aligned}$$

for $Q_n \in \mathbb{H}_n$ and $P \in \mathbb{E}$.

Now, if P is an entropy limit point of $\mathbb{E}$ then there are $Q_n \in \mathbb{H}_n$ such that $|H_n(Q_n) - H_n(P)| \longrightarrow 0$ as $n \longrightarrow \infty$. Hence $\|P-Q_n\|_n^2$ also converge to zero, as required. ■

Theorem 12. If $\mathbb{E}$ contains an entropy limit point P then

$$\operatorname{maxent}\mathbb{E} = \{P\}.$$

Proof: First we shall show that $P \in \operatorname{maxent} \mathbb{E}$; later we shall see that there is no other member of $\operatorname{maxent} \mathbb{E}$.

First, then, assume for contradiction that $P \notin \operatorname{maxent} \mathbb{E}$. Then there is some $Q \in \mathbb{E}$ such that Q has greater entropy than P. That is, for sufficiently large n, $H_n(Q_n) \geq H_n(Q) > H_n(P)$, where the $Q_n \in \mathbb{H}_n$ converge in entropy (and, by Proposition 11, in L_1) to P. N.b., $Q \neq P$. Hence, for sufficiently large n,

$$\begin{array}{rcl} H_n(Q_n) - H_n(P) & > & H_n(Q_n) - H_n(Q) \\ & \geq & d_n(Q, Q_n) \\ & \geq & \dfrac{1}{2} \left\| Q - Q_n \right\|_n^2 . \end{array}$$

Since the Q_n converge in entropy to P, they converge in L_1 to Q. By the uniqueness of L_1 limit points, $Q = P$: a contradiction. Hence $P \in \operatorname{maxent} \mathbb{E}$, as required.

Next we shall see that P is the unique member of $\operatorname{maxent} \mathbb{E}$. Suppose for contradiction that there is some $P^\dagger \in \operatorname{maxent} \mathbb{E}$ such that $P^\dagger \neq P$. Then P cannot eventually dominate $P^\dagger$ in n-entropy—i.e., there is some infinite set $J \subseteq \mathbb{N}$ such that for $n \in J$,

$$H_n(P^\dagger) \geq H_n(P).$$

Let $R \stackrel{\text{df}}{=} \lambda P^\dagger + (1-\lambda)P$ for some $\lambda \in (0,1)$. Now by the log-sum inequality (Cover and Thomas, 1991, Theorem 2.7.1), for all $n \in J$ large enough that $P^\dagger(\omega_n) \neq P(\omega_n)$ for some $\omega_n \in \Omega_n$,

$$\begin{array}{rcl} H_n(R) & > & \lambda H_n(P^\dagger) + (1-\lambda) H_n(P) \\ & \geq & \lambda H_n(P) + (1-\lambda) H_n(P) \\ & = & H_n(P). \end{array}$$

Hence,

$$\begin{array}{rcl} H_n(Q_n) - H_n(P) & > & H_n(Q_n) - H_n(R) \\ & \geq & d_n(R, Q_n), \end{array}$$

for large enough $n \in J$.

Now by Pinsker's inequality and the definition of R,

$$\begin{array}{rcl} d_n(R, Q_n) & \geq & \dfrac{1}{2} \left\| R - Q_n \right\|_n^2 \\ & = & \dfrac{1}{2} \left\| P - Q_n + \lambda(P^\dagger - P) \right\|_n^2 \\ & = & \dfrac{1}{2} \left(\displaystyle\sum_{\omega_n \in \Omega_n} \left| P(\omega_n) - Q_n(\omega_n) + \lambda(P^\dagger(\omega_n) - P(\omega_n)) \right| \right)^2 . \end{array}$$

Let $f_n(\varphi) \stackrel{\text{df}}{=} P(\varphi) - Q_n(\varphi) + \lambda(P^\dagger(\varphi) - P(\varphi))$ and $\rho_n \stackrel{\text{df}}{=} \bigvee_{f_n(\omega_n)>0} \omega_n$.

Then,

$$\begin{aligned}
\sum_{\omega_n\in\Omega_n} |f_n(\omega_n)| &= \sum_{\omega_n:f_n(\omega_n)>0} f_n(\omega_n) - \sum_{\omega_n:f_n(\omega_n)\leq 0} f_n(\omega_n)\\
&= \sum_{\omega_n:f_n(\omega_n)>0} f_n(\omega_n) - \sum_{\omega_n:f_n(\omega_n)\not>0} f_n(\omega_n)\\
&= f_n(\rho_n) - f_n(\neg\rho_n)\\
&= 2f_n(\rho_n)
\end{aligned}$$

after substituting $P(\neg\rho_n) = 1 - P(\rho_n)$ etc.

Let us consider the behaviour of

$$f_n(\rho_n) = P(\rho_n) - Q_n(\rho_n) + \lambda(P^\dagger(\rho_n) - P(\rho_n))$$

as $n \longrightarrow \infty$. Now, $P(\rho_n) - Q_n(\rho_n) \longrightarrow 0$ as $n \longrightarrow \infty$, because Q_n converges in L_1 to P. However, $\lambda(P^\dagger(\rho_n) - P(\rho_n)) \not\longrightarrow 0$ as $n \longrightarrow \infty$, as we shall now see. $P^\dagger \neq P$ by assumption, so they must differ on some quantifier-free sentence ψ, a sentence of $\mathcal{L}_m$, say. Suppose without loss of generality that $P^\dagger(\psi) > P(\psi)$ (otherwise take $\neg\psi$ instead) and let $\delta = P^\dagger(\psi) - P(\psi) > 0$. Now for $n \geq m$,

$$f_n(\rho_n) = \sum_{\omega_n:f_n(\omega_n)>0} f_n(\omega_n) \geq \sum_{\omega_n\models\psi} f_n(\omega_n) = f_n(\psi) \ .$$

Since Q_n converges in L_1 to P we can consider $n > m$ large enough that (Cover and Thomas, 1991, Equation 11.137):

$$\|Q_n - P\|_n = 2 \max_{\varphi\in S\mathcal{L}_n} (Q_n(\varphi) - P(\varphi)) < \lambda\delta \ .$$

In particular, since ψ is quantifier-free, $Q_n(\psi) - P(\psi) \leq \max_{\varphi\in S\mathcal{L}_n}(Q_n(\varphi) - P(\varphi)) < \lambda\delta/2$. For any such n,

$$\begin{aligned}
f_n(\rho_n) &\geq f_n(\psi)\\
&= P(\psi) - Q_n(\psi) + \lambda(P^\dagger(\psi) - P(\psi))\\
&> -\frac{\lambda\delta}{2} + \lambda\delta\\
&= \frac{\lambda\delta}{2} \ .
\end{aligned}$$

Putting the above parts together, we have that for sufficiently large $n \in J$,

$$H_n(Q_n) - H_n(P) > d_n(R, Q_n) \geq \frac{(2f_n(\rho_n))^2}{2} > \frac{\lambda^2\delta^2}{2} > 0 \ .$$

However, that these $H_n(Q_n) - H_n(P)$ are bounded away from zero contradicts the assumption that the Q_n converge in entropy to P. Hence, P is the unique member of maxent $\mathbb{E}$, as required. ■

Appendix 2. Alternative proof of Corollary 16

This appendix provides a more direct proof of Corollary 16, which identifies an important scenario in which the equivocator function conditioned on a categorical constraint is the maximal entropy function.

Corollary 16. If $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large n then

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

Proof: There are two cases: either $P_=(\varphi) = 1$ or $P_=(\varphi) < 1$.

If $P_=(\varphi) = 1$ then $P_= \in \mathbb{E}_\varphi$ and $P_=(\cdot|\varphi) = P_=(\cdot)$. $P_=$ is the unique member of $\text{maxent}\,\mathbb{E}_\varphi$ because the equivocator function has greater entropy than any other probability function, so $\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}$, as required.

If $P_=(\varphi) < 1$ then we can proceed as follows.

Since $P_=(\varphi) > 0$, $P_=(\cdot|\varphi)$ is well defined. $P_=(\varphi|\varphi) = 1$ so $P_=(\cdot|\varphi) \in \mathbb{E}$. Thus $\mathbb{E}_\varphi \neq \emptyset$.

Suppose for contradiction that $\text{maxent}\,\mathbb{E}_\varphi \neq \{P_=(\cdot|\varphi)\}$. Then in $\mathbb{E}_\varphi$ there must be some $P^\dagger \neq P_=(\cdot|\varphi)$ that is not eventually dominated in entropy by $P_=(\cdot|\varphi)$. That is, there is some infinite $J \subseteq \mathbb{N}$ such that $H_n(P^\dagger) \geq H_n(P_=(\cdot|\varphi))$ for all $n \in J$. (To see this consider that there are three cases: (i) if $\text{maxent}\,\mathbb{E}_\varphi = \emptyset$ then every member of $\mathbb{E}_\varphi$ is eventually dominated by some other in entropy, so $P_=(\cdot|\varphi)$ is dominated by some $P^\dagger$ and $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$; (ii) if $P_=(\cdot|\varphi) \notin \text{maxent}\,\mathbb{E}_\varphi = \{P^\dagger, \ldots\}$ then $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$; (iii) if $\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi), P^\dagger, \ldots\}$ then $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$.)

Define a probability function $Q \stackrel{\text{df}}{=} \lambda P^\dagger + (1-\lambda)P_=(\cdot|\varphi)$ for some $\lambda \in (0,1)$. By the log-sum inequality (Cover and Thomas, 1991, Theorem 2.7.1), for all $n \in J$ large enough that $P^\dagger(\omega) \neq P_=(\omega|\varphi)$ for some $\omega \in \Omega_n$,

$$\begin{aligned} H_n(Q) &> \lambda H_n(P^\dagger) + (1-\lambda) H_n(P_=(\cdot|\varphi)) \\ &\geq \lambda H_n(P_=(\cdot|\varphi)) + (1-\lambda) H_n(P_=(\cdot|\varphi)) \\ &= H_n(P_=(\cdot|\varphi)). \end{aligned}$$

However, that $H_n(Q) > H_n(P_=(\cdot|\varphi))$ for sufficiently large $n \in J$ contradicts the assumption that $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large n. Hence $\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}$, as required. ■

Appendix 3. Zero measure Premisses of Higher Quantifier Complexity

Proposition 49 (Σ_{2m}). *For* $\varphi = \exists x_{2m} \forall x_{2m-1} \ldots \forall x_1 U x_{2m} x_{2m-1} \ldots x_1 \in \Sigma_{2m}$ *it holds that for all* $P \in \mathbb{E}_\varphi$ *there exists a probability function* $Q \in \mathbb{E}_\varphi$ *which has greater entropy. Hence,* $\text{maxent}\,\mathbb{E}_\varphi = \emptyset$.

Proof: For ease of notation we will write $Ut_i\vec{t}$ for $Ut_i t_{k_{2m-1}} \ldots t_{k_1}$ and $\bigwedge_{t=1}^n Ut_i\vec{t}$ for $\bigwedge_{k_{2m-1}=1}^n \cdots \bigwedge_{k_1=1}^n Ut_i t_{k_{2m-1}} \ldots t_{k_1}$

Suppose for contradiction that $\text{maxent}\,\mathbb{E} \neq \emptyset$ and let $P \in \text{maxent}\,\mathbb{E}$. Note that $P_=(\varphi) = 0 < 1 = P(\varphi)$. Hence, $P \neq P_=$.

Let us now define a probability function $P' \in \mathbb{E}$ by shifting all witnessing of $\exists x_{2m} \forall x_{2m-1} \exists x_{2m-2} \ldots . \forall x_1 U\vec{x}$ by one and then adding a constant t_1 such that $Ut_1\vec{t}$ is independent from all other literals for all $\vec{t}$. Intuitively, the literals $\pm Ut_i\vec{t}$ are replaced by $\pm Ut_{i+1}\vec{t}$.

Formally, let $\omega_n \in \Omega_n = \bigwedge_{i,t=1}^n U^{\epsilon_{i,\vec{t}}} t_i\vec{t}$ be an arbitrary n-state. Then define P' by

$$\begin{aligned} P'(\omega_n) :&= P(\bigwedge_{i=2}^{n}\bigwedge_{t=1}^{n} U^{\epsilon_{i-1,\vec{t}}} t_{i-1}\vec{t}) \cdot P_{=}(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1\vec{t}) \\ &= \frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i\vec{t})}{2^{n^{2m-1}}} \ . \end{aligned}$$

Firstly, we note that

$$P'(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_1\vec{x}) = \lim_{n\to\infty} P'(\bigwedge_{j=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_1t_j\vec{x}) =$$

$$\lim_{n\to\infty} P_{=}(\bigwedge_{j=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_1t_j\vec{x}) = 0 \tag{10}$$

(reference). So, according to P' the constant t_1 is not a witness of the existential premiss sentence φ.

We next show that $P \neq P'$. Firstly, note that

$$\lim_{n\to\infty} P(\bigvee_{i=1}^{n} \forall y\exists x_{2m-2}....\forall x_1 Ut_1y\vec{x}) = P(\exists z\forall y\exists x_{2m-2}....\forall x_1 Uzy\vec{x})) = 1$$

and thus $\min\{i \in \mathbb{N} \ : \ P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}) > 0\}$ is a finite number and thus obtains . With this and (10), we have

$$\begin{aligned} &\min\{i \in \mathbb{N} \ : \ P'(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_ix_{2m-1}\vec{x}) > 0\} \\ &= \min\{i \in \mathbb{N} \ : \ \lim_{n\to\infty} P'(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_it_k\vec{x}) > 0\} \\ &= \min\{i \in \mathbb{N} \ : \ \lim_{n\to\infty} P(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_{i-1}t_k\vec{x}) > 0\} \\ &= 1 + \min\{i \in \mathbb{N} \ : \ \lim_{n\to\infty} P(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_it_k\vec{x}) > 0\} \\ &= 1 + \min\{i \in \mathbb{N} \ : \ P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_ix_{2m-1}\vec{x}) > 0\} \ . \end{aligned}$$

So, $P \neq P'$.

We also observe that for all $i \geq 2$ it holds that

$$P'(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}) = P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_{i-1}\vec{x})$$

and furthermore

$$P'(\bigvee_{i\in I} \forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}) = P(\bigvee_{i\in I} \forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_{i-1}\vec{x})$$

for all finite index sets I. So,

$$
\begin{aligned}
P'(\exists y \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U y \vec{x}) &= \lim_{n\to\infty} P'(\bigvee_{i=1}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}) \\
&\geq \lim_{n\to\infty} P'(\bigvee_{i=2}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}) \\
&= \lim_{n\to\infty} P(\bigvee_{i=1}^{n-1} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}) \\
&= \lim_{n\to\infty} P(\bigvee_{i=1}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}) \\
&= 1 \ .
\end{aligned}
$$

This means that $P'(\exists x \forall y U x y) = 1$ and thus, as advertised, $P' \in \mathbb{E}$.

We now calculate n-entropies of P and P' and find for $n \geq 1$ that:

$$
\begin{aligned}
H_n(P) = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} P(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}) \\
&\cdot \log(P(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})) \\
H_n(P') = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 1 \leq i \leq n}} P'(\bigwedge_{i,t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}) \cdot \log(P'(\bigwedge_{i,t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})) \\
= &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} P'(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}) \\
&\cdot \log(P'(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})) \\
= &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} \frac{P(\bigwedge_{i=1}^{n-1} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})}{2^{n^{2m-1}}} \\
&\cdot \log(\frac{P(\bigwedge_{i=1}^{n-1} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})}{2^{n^{2m-1}}}) \ .
\end{aligned}
$$

Holding the first summation fixed, we note that since n-entropy is maximised by maximally equivocating $H_n(P) \leq H_n(P')$. Now define $Q := \frac{P+P'}{2}$. Since $\mathbb{E}$ is convex and $P, P' \in \mathbb{E}$, we observe that $Q \in \mathbb{E}$.

Since n-entropy is a strictly concave function we conclude that $H_n(Q) > H_n(P)$ whenever P and P' disagree on $\mathcal{L}_n$. Since $P \neq P'$ there has to exists some finite M and quantifier free sentence $\psi \in QFS\mathcal{L}_M$ such that $P(\psi) \neq P'(\psi)$ (Gaifman's Theorem). Since $\mathcal{L}_m \subset \mathcal{L}_{m+1}$ for all m we have that P disagrees with P' on $\mathcal{L}_m$ for all $m \geq M$. We have hence found a $Q \in \mathbb{E}$ such that $H_n(Q) > H_n(P)$ for all large enough n. Hence,

$P \notin \operatorname{maxent} \mathbb{E}$. Contradiction. ■

Proposition 50 (Π_3)**.** *For $\varphi = \forall x \exists y \forall z Sxyz \in \Pi_3$ it holds that for all $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy. Hence,* $\operatorname{maxent} \mathbb{E}_\varphi = \emptyset$.

Proof: Let us first note that

$$\begin{aligned}\mathbb{E}_\varphi &= \{P \in \mathbb{P} \;:\; P(\varphi) = 1\} \\ &= \{P \in \mathbb{P} \;:\; P(\exists y \forall z S t_1 y z) = 1,\; P(\exists y \forall z S t_2 y z) = 1, \ldots, \} \;. \end{aligned} \tag{11}$$

Assume for contradiction that $P \in \operatorname{maxent} \mathbb{E}_\varphi$. Since $P_=(\varphi) = 0$, P cannot be the equivocator. However, since $P \in \mathbb{E}_\varphi$, it must also holds that for all t_i ($i \in \mathbb{N}$) there has to exist some minimal $t_{k_i^*}$ ($k_i^* \geq 1$) such that $P(\forall z S t_i t_{k_i^*} z) > 0$.

We now define a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy than P, which contradicts that $P \in \operatorname{maxent} \mathbb{E}_\varphi$. First, we postpone for all i the witnessing (see Proposition 47) to $k_i^* + 1$. This is again achieved by first defining a probability function $P' \in \mathbb{E}_\varphi \setminus \{P\}$ such that $H_n(P') \geq H_n(P)$ for all large enough n:

$$P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l) := \frac{P(\bigwedge_{k=1}^{n-1} \bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l)}{2^n} \;.$$

As we saw in Proposition 47, it holds that $P'(\exists y \forall z S t_i y z) = 1$ for all $i \in \mathbb{N}$. Furthermore, for all $i \in \mathbb{N}$ there exist an $n_i \in \mathbb{N}$ and $\epsilon_{k,l} \in \{0,1\}^{n_i \times n_i}$ such that $P'(\bigwedge_{k=1}^{n_i} \bigwedge_{l=1}^{n_i} S^{\epsilon_{k,l}} t_i t_k t_l) \neq P(\bigwedge_{k=1}^{n_i} \bigwedge_{l=1}^{n_i} S^{\epsilon_{k,l}} t_i t_k t_l)$.

Given the way we wrote $\mathbb{E}_\varphi$ (see (11)), we see that every extension of P' to a probability function – which so far not be defined on the entire language – will be in $\mathbb{E}_\varphi$ since membership in $\mathbb{E}_\varphi$ solely depends on sub-states where the first constant is fixed to some t_i.

We now define P' on an arbitrary n-state ω_n of the language, and hence via Gaifman's Theorem on the entire language by

$$P'(\omega_n) := \prod_{i=1}^{n} P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l) \;.$$

Because of the additivity of the entropy function (Csiszár, 2008, P. 63), we also find for all $n \in \mathbb{N}$ that

$$H_n(P') = -\sum_{i=1}^{n} \sum_{\substack{\epsilon_{i,r,s} \in \{0,1\} \\ 1 \leq r \leq n \\ 1 \leq s \leq n}} P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_1 t_k t_l) \cdot \log(P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_1 t_k t_l)) \;.$$

Since the entropy function is maximised for independent variables we also find:

$$H_n(P) \geq -\sum_{i=1}^{n} \sum_{\substack{\epsilon_{i,r,s} \in \{0,1\} \\ 1 \leq r \leq n \\ 1 \leq s \leq n}} P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l) \cdot \log(P'(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l)) \;.$$

Now recall that for all large enough fixed $i \in \mathbb{N}$ we saw in Proposition 47 that the following inequality holds

$$
\begin{aligned}
{}_iH_n(P') :&= - \sum_{\substack{\epsilon_{r,s} \in \{0,1\} \\ 1 \le r \le n \\ 1 \le s \le n}} P'(\bigwedge_{k=1}^n \bigwedge_{l=1}^n S^{\epsilon_{k,l}} t_i t_k t_l) \cdot \log(P'(\bigwedge_{k=1}^n \bigwedge_{l=1}^n S^{\epsilon_{k,l}} t_i t_k t_l)) \\
&= - \sum_{\substack{\epsilon_{r,s} \in \{0,1\} \\ 1 \le r \le n \\ 1 \le s \le n}} P'(\bigwedge_{k=1}^n \bigwedge_{l=1}^n U^{\epsilon_{k,l}} t_k t_l) \cdot \log(P'(\bigwedge_{k=1}^n \bigwedge_{l=1}^n U^{\epsilon_{k,l}} t_k t_l)) \\
&\ge - \sum_{\substack{\epsilon_{r,s} \in \{0,1\} \\ 1 \le r \le n \\ 1 \le s \le n}} P(\bigwedge_{k=1}^n \bigwedge_{l=1}^n U^{\epsilon_{k,l}} t_k t_l) \cdot \log(P(\bigwedge_{k=1}^n \bigwedge_{l=1}^n U^{\epsilon_{k,l}} t_k t_l)) \\
&= - \sum_{\substack{\epsilon_{r,s} \in \{0,1\} \\ 1 \le r \le n \\ 1 \le s \le n}} P(\bigwedge_{k=1}^n \bigwedge_{l=1}^n S^{\epsilon_{k,l}} t_i t_k t_l) \cdot \log(P(\bigwedge_{k=1}^n \bigwedge_{l=1}^n S^{\epsilon_{k,l}} t_i t_k t_l)) := {}_iH_n(P) \ .
\end{aligned}
$$

So, we have for all large enough $n \in \mathbb{N}$ that

$$
H_n(P') = \sum_{i=1}^n {}_iH_n(P') \ge \sum_{i=1}^n {}_iH_n(P) \ge H_n(P) \ .
$$

We again put $Q := \frac{P+P'}{2}$ and note that since $P \neq P'$ that $Q \neq P$. Since $P' \in \mathbb{E}_\varphi$ we easily find by applying the convexity of $\mathbb{E}_\varphi$ that $Q \in \mathbb{E}_\varphi$. Furthermore, $H_n(Q) > H_n(P)$ for all large enough $n \in \mathbb{N}$ since Q is a convex combination of P and P' and $H_n(P') \ge H_n(P)$ for all $n \in \mathbb{N}$. ■

Proposition 51 (Π_{2m+3}). *For $\varphi = \forall v_1 \exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R v_1 w_1 \ldots v_m w_m xyz \in \Pi_{2m+1}$ it holds that for all $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy. Hence,* maxent $\mathbb{E}_\varphi = \emptyset$.

Proof: The proof proceeds by induction on the quantifier complexity m.

The *base case* $m = 0$ is Proposition 50.

The *induction step* for $m \ge 1$ assumes the result for $m - 1 \ge 0$. The proof follows the blue print laid out in the base case.

Let us first note that

$$
\begin{aligned}
\mathbb{E}_\varphi &= \{P \in \mathbb{P} \ : \ P(\varphi) = 1\} \\
&= \{P \in \mathbb{P} \ : \ P(\exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_1 w_1 \ldots v_m w_m xyz) = 1, \\
&\qquad P(\exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_2 w_1 \ldots v_m w_m xyz) = 1, \\
&\qquad \ldots, \} \ . \qquad (12)
\end{aligned}
$$

Assume for contradiction that $P \in$ maxent $\mathbb{E}_\varphi$. Since $P_=(\varphi) = 0$, P cannot be the equivocator. However, since $P \in \mathbb{E}_\varphi$, it must also holds that for all t_i ($i \in \mathbb{N}$) there has to exist some minimal $t_{k_i^*}$ ($k_i^* \ge 1$) such that $P(\forall v_2 \exists w_2 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_i t_{k_i^*} v_2 w_2 \ldots v_m w_m xyz) > 0$. We now postpone this witnessing as usual.

We begin by assigning probabilities to substates fixing t_i

$$P'(\bigwedge_{b_1=1}^{n}\bigwedge_{a_2=1}^{n}\dots\bigwedge_{a_{m+1}=1}^{n}\bigwedge_{b_{m+1}=1}^{n}\bigwedge_{a_{m+2}=1}^{n} R^{\epsilon_{b_1,a_2,\dots,a_{m+2}}} t_i t_{b_1} t_{a_2}\dots t_{a_{m+2}}) := \frac{P(\bigwedge_{b_1=1}^{n-1}\bigwedge_{a_2=1}^{n}\dots\bigwedge_{a_{m+1}=1}^{n}\bigwedge_{b_{m+1}=1}^{n}\bigwedge_{a_{m+2}=1}^{n} R^{\epsilon_{b_1,a_2,\dots,a_{m+2}}} t_i t_{b_1} t_{a_2}\dots t_{a_{m+2}})}{2^n} .$$

Again, it holds that $P'(\exists w_1\dots\forall v_m\exists w_m\forall x\exists y\forall z R t_i w_1\dots v_m w_m xyz)=1$ for all $i\in\mathbb{N}$. Furthermore, for all $i\in\mathbb{N}$ there exist an $n_i\in\mathbb{N}$ and $\vec{\epsilon}\in\{0,1\}^{n_i^{2m+2}}$ such that

$$\begin{aligned}
&P'(\bigwedge_{b_1=1}^{n_i}\bigwedge_{a_2=1}^{n_i}\dots\bigwedge_{a_{m+1}=1}^{n_i}\bigwedge_{b_{m+1}=1}^{n_i}\bigwedge_{a_{m+2}=1}^{n_i} R^{\epsilon_{b_1,a_2,\dots,a_{m+2}}} t_i t_{b_1} t_{a_2}\dots t_{a_{m+2}})\\
&\quad :\neq P'(\bigwedge_{b_1=1}^{n_i}\bigwedge_{a_2=1}^{n_i}\dots\bigwedge_{a_{m+1}=1}^{n_i}\bigwedge_{b_{m+1}=1}^{n_i}\bigwedge_{a_{m+2}=1}^{n_i} R^{\epsilon_{b_1,a_2,\dots,a_{m+2}}} t_i t_{b_1} t_{a_2}\dots t_{a_{m+2}}) .
\end{aligned}$$

In particular, $P'\neq P$.

We now define P' on an arbitrary n-state ω_n of the language, and hence via Gaifman's Theorem on the entire language by fixing $\vec{\epsilon}_i\in\{0,1\}^{n^{2m+2}}$ for $1\leq i\leq n$ and letting

$$P'(\omega_n) := \prod_{i=1}^{n} P'(\bigwedge_{\vec{\epsilon}_i\in\{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i \vec{t}) .$$

Because of the additivity of the entropy function (Csiszár, 2008, P. 63), we also find for all $n\in\mathbb{N}$ that

$$\begin{aligned}
H_n(P') &= -\sum_{i=1}^{n}\sum_{\vec{\epsilon}_i\in\{0,1\}^{n^{2m+2}}} P'(\bigwedge_{\vec{\epsilon}_i\in\{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i\vec{t})\cdot\log(P'(\bigwedge_{\vec{\epsilon}_i\in\{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i\vec{t}))\\
&:= \sum_{i=1}^{n} {}_iH_{n,2m+2}(P') .
\end{aligned}$$

We now use the proof of Proposition 49 to obtain that for all i and all large enough n (depending on i) that

$${}_iH_{n,2m+2}(P') \geq {}_iH_{n,2m+2}(P) .$$

${}_iH_{n,2m}(P)$ is the n-entropy of a probability function P on a language containing one $2m+2$-ary relation symbol U, $\varphi = \exists w_1\forall v_2\exists w_2\dots\exists w_{m+1}\forall v_{m+2} U w_1 v_2 w_2\dots w_{m+1} v_{m+2}\in\Pi_{2m+2}$ and $P\in\mathbb{E}_\varphi$.

Since n-entropy is maximised by probability functions with a maximal probabilistic independences, we again have

$$H_n(P) \geq \sum_{i=1}^{n} {}_iH_{n,2m+2}(P) ,$$

which overall gives the not-necessarily strict inequality:

$$H_n(P') = \sum_{i=1}^{n} {}_iH_{n,2m+2}(P') \geq \sum_{i=1}^{n} {}_iH_{n,2m+2}(P) \geq H_n(P) .$$

Taking Q to be any convex combination of P and P', we see that $H_n(Q) > H_n(P)$ for all large enough n. This entails that Q has greater entropy than P. $\blacksquare$

Bibliography

Balestrino, A., Caiti, A., and Crisostomi, E. (2006). Efficient numerical approximation of maximum entropy estimates. *International Journal of Control*, 79(9):1145–1155.

Barnett, O. and Paris, J. B. (2008). Maximum Entropy Inference with Quantified Knowledge. *Logic Journal of IGPL*, 16(1):85–98.

Billingsley, P. (1979). *Probability and measure*. John Wiley and Sons, New York, third (1995) edition.

Carnap, R. (1952). *The continuum of inductive methods*. University of Chicago Press, Chicago IL.

Caticha, A. and Giffin, A. (2006). Updating Probabilities. In *Proceedings of MaxEnt*, volume 872, pages 31–42.

Chen, B., Hu, J., and Zhu, Y. (2010). Computing maximum entropy densities: A hybrid approach. *Signal Processing: An International Journal*, 4(2):114–122.

Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley and Sons, New York, second (2006) edition.

Csiszár, I. (2008). Axiomatic Characterizations of Information Measures. *Entropy*, 10(3):261–273.

Gaifman, H. (1964). Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2(1):1–18.

Goldman, S. A. (1987). Efficient methods for calculating maximum entropy distributions. Master's thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Goldman, S. A. and Rivest, R. (1988). A non-iterative maximum entropy algorithm. In Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 2*, pages 133–148. Elsevier, North-Holland.

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic logics and probabilistic networks*. Synthese Library. Springer, Dordrecht.

Howarth, E. and Paris, J. B. (2019). Pure Inductive Logic with Functions. *Journal of Symbolic Logic*, pages 1–22.

Howson, C. (2014). Finite additivity, another lottery paradox and conditionalisation. *Synthese*, 191(5):989–1012.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Jaynes, E. T. (2003). *Probability theory: the logic of science.* Cambridge University Press, Cambridge.

Landes, J. (2009). *The Principle of Spectrum Exchangeability within Inductive Logic.* PhD thesis, Manchester Institute for Mathematical Sciences.

Landes, J. (2021a). A Triple Uniqueness of the Maximum Entropy Approach. In Vejnarová, J. and Wilson, N., editors, *Proceedings of ECSQARU*, volume 12897 of *LNAI*, pages 644–656, Cham. Springer.

Landes, J. (2021b). The Entropy-limit (Conjecture) for Σ_2-Premisses. *Studia Logica*, 109:423–442.

Landes, J., Paris, J. B., and Vencovská, A. (2009). Representation Theorems for probability functions satisfying Spectrum Exchangeability in Inductive Logic. *International Journal of Approximate Reasoning*, 51(1):35–55.

Landes, J., Rafiee Rad, S., and Williamson, J. (2021). Towards the Entropy-Limit Conjecture. *Annals of Pure and Applied Logic*, 172(2):102870.

Landes, J. and Williamson, J. (2015). Justifying objective Bayesianism on predicate languages. *Entropy*, 17(4):2459–2543.

Landes, J. and Williamson, J. (2016). Objective Bayesian nets from consistent datasets. In Giffin, A. and Knuth, K. H., editors, *Proceedings of MaxEnt*, volume 1757, pages 020007–1 – 020007–8. AIP.

Lehmann, D. and Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60.

Ormoneit, D. and White, H. (1999). An efficient algorithm to compute maximum entropy densities. *Econometric Reviews*, 18(2):127–140.

Paris, J. and Vencovská, A. (2015). *Pure Inductive Logic.* Cambridge University Press, Cambridge.

Paris, J. B. (1994). *The uncertain reasoner's companion.* Cambridge University Press, Cambridge.

Paris, J. B. (1998). Common Sense and Maximum Entropy. *Synthese*, 117:75–93.

Paris, J. B. and Rafiee Rad, S. (2010). A Note on the Least Informative Model of a Theory. In Ferreira, F., Löwe, B., Mayordomo, E., and Mendes Gomes, L., editors, *Proceedings of CiE*, pages 342–351, Berlin. Springer.

Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4(3):183–223.

Paris, J. B. and Vencovská, A. (1997). In Defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17(1):77–103.

Paris, J. B. and Vencovská, A. (2019). Six Problems in Pure Inductive Logic. *Journal of Philosophical Logic.*

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, San Mateo CA.

Rafiee Rad, S. (2009). *Inference Processes for Probabilistic First Order Languages*. PhD thesis, Manchester Institute for Mathematical Sciences.

Rafiee Rad, S. (2017). Equivocation axiom on first order languages. *Studia Logica*.

Rafiee Rad, S. (2018). Maximum Entropy Models for Σ_1 Sentences. *Journal of Applied Logics - IfCoLoG Journal of Logics and their Applications*, 5(1):287–300.

Rafiee Rad, S. (2021). On probabilistic characterisation of models of first order theories. *Annals of Pure and Applied Logic*.

Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, 53(4):467–491.

Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423.

Williams, P. M. (1980). Bayesian Conditionalisation and the Principle of Minimum Information. *British Journal for the Philosophy of Science*, 31(2):131–144.

Williamson, J. (2008). Objective Bayesian probabilistic logic. *Journal of Algorithms*, 63(4):167–183.

Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford University Press, Oxford.

Williamson, J. (2017). *Lectures on inductive logic*. Oxford University Press, Oxford.