

# Towards the Entropy-Limit Conjecture

Draft of June 11, 2019.

## Abstract

The maximum entropy principle is widely used to determine non-committal probabilities on a finite domain, subject to a set of constraints, but its application to continuous domains is notoriously problematic. This paper concerns an intermediate case, where the domain is a first-order predicate language. Two strategies have been put forward for applying the maximum entropy principle on such a domain: (i) applying it to finite sublanguages and taking the pointwise limit of the resulting probabilities as the size  $n$  of the sublanguage increases; (ii) selecting a probability function on the language as a whole whose entropy on finite sublanguages of size  $n$  is not dominated by that of any other probability function for sufficiently large  $n$ . The entropy-limit conjecture says that, where these two approaches yield determinate probabilities, the two methods yield the same probabilities. If this conjecture is found to be true, it would provide a boost to the project of seeking a single canonical inductive logic—a project which faltered when Carnap’s attempts in this direction succeeded only in determining a continuum of inductive methods. The truth of the conjecture would also boost the project of providing a canonical characterisation of normal or default models of first-order theories.

Hitherto, the entropy-limit conjecture has been verified for languages which contain only unary predicate symbols and also for the case in which the constraints can be captured by a categorical statement of  $\Sigma_1$  quantifier complexity. This paper shows that the entropy-limit conjecture also holds for categorical statements of  $\Pi_1$  complexity, for various non-categorical constraints, and in certain other general situations.

Keywords: Maximum Entropy, Objective Bayesianism, Probabilistic Constraints on Predicate Languages, Inductive Logic, Normal Models, Default Models.

## 1 Introduction

Inductive logic seeks to determine how much certainty to attach to a conclusion proposition  $\psi$ , given premiss propositions  $\varphi_1, \dots, \varphi_k$  to which attach measures of certainty  $X_1, \dots, X_k$  respectively. That is, the main task is to find  $Y$  such that

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \psi^Y,$$

where  $\approx$  signifies the inductive entailment relation. Often,  $X_1, \dots, X_k, Y$  are probabilities or sets of probabilities. There are many possible semantics for inductive logic [21]. One key approach stems from the work of Carnap, who

provided a continuum of inductive entailment relations [9, 10, 11, 35]. An alternative approach, which is the focus of this paper, is to apply the maximum entropy principle of Jaynes [24, 25]. According to this approach one should consider, from all the probability functions that satisfy the premisses, those with maximum entropy, and let  $Y$  be the set of probability values that these functions give to the conclusion  $\psi$ .

If the underlying logical language is a finite propositional language, then this latter proposal is rather straightforward to implement and has many nice properties [36].<sup>1</sup> However, if the language is a first-order predicate language  $\mathcal{L}$  with infinitely many constant symbols, certain intriguing questions arise. In particular, there are two main ways to implement the proposal in the predicate-language case, and it is not entirely clear as to whether the resulting inductive logics agree.

One approach, due to Barnett and Paris [7], proceeds as follows: (i) reinterpret the premisses as constraints on the probabilities of sentences of a *finite* predicate language  $\mathcal{L}_n$  that has  $n$  constant symbols; (ii) determine the function  $P^n$  that maximises entropy on this finite language, subject to constraints imposed by the reinterpreted premisses; (iii) draw inductive inferences using the function  $P^\infty$  defined by  $P^\infty(\theta) \stackrel{\text{df}}{=} \lim_{n \rightarrow \infty} P^n(\theta)$  for any sentence  $\theta$  of  $\mathcal{L}$ . (The technical details will be explained below.)

A second approach, explored by Williamson [45, 46], proceeds as follows: (i) consider probability functions defined on the language  $\mathcal{L}$  as a whole; (ii) deem one probability function  $P$  to have greater entropy than another function  $Q$  if  $H_n(P)$ , where  $H_n$  is the entropy function on the finite sublanguage  $\mathcal{L}_n$ , dominates  $H_n(Q)$  for sufficiently large  $n$ ; (iii) draw inductive inferences using those functions  $P^\dagger$ , from all the probability functions on  $\mathcal{L}$  that satisfy the premisses, that have maximal entropy (i.e., no other function satisfying the premisses has greater entropy). Again, see below for details.

The first approach, which we shall call the *entropy-limit* approach, has the advantage that it is more constructive, so it can be easier to calculate the probabilities required for inductive inference. The second approach, which we shall call the *maximal-entropy* approach, has the advantage that it yields determinate results in certain cases where the entropy-limit approach does not. This is because the entropy-limit approach faces what is known as the finite model problem: contingent premisses can become inconsistent when reinterpreted as applying to a finite domain.

These approaches to inductive logic would be strengthened if it could be shown that they give the same same results where they are both applicable. Then one could use the maximal-entropy approach to provide a general semantics for inductive logic, but use the entropy-limit approach where a more constructive approach is helpful.

Following some results of Rafiee Rad [39], discussed in the next section, Williamson [46, p. 191] articulated the following conjecture:

**Entropy-limit Conjecture.** Where  $P^\infty$  exists and satisfies the constraints imposed by the premisses, it is the unique function with maximal entropy

---

<sup>1</sup>Maximum entropy probability functions are of interest in numerous applications, including the statistical analysis of multiple data-sets (meta analysis) [2, 3], information geometry [1], multi-expert reasoning [4], logical frameworks in artificial intelligence [31, 32] and quantum mechanics [12], among others [8, 15, 26, 27, 28, 29, 30, 47].

from all those that satisfy the premisses, i.e.,  $P^\dagger = P^\infty$  .

If the entropy-limit conjecture is true, this would lend support to the claim that maximising entropy leads to a canonical inductive logic—a goal that has hitherto proved very elusive [46]. In this paper, we provide new evidence for the entropy-limit conjecture. We show that the entropy-limit conjecture is true for a single premiss that takes the form of a categorical  $\Pi_1$  sentence  $\forall \vec{x}\theta(\vec{x})$  where  $\theta(\vec{x})$  is quantifier-free (§3); for various scenarios in which there are multiple non-categorical premisses,  $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$  where  $X_i$  is a probability or set of probabilities that attaches to the sentence  $\varphi_i$  (§4); and for certain general cases in which convergence of the  $n$ -entropy maximiser  $P^n$  to the entropy limit  $P^\infty$  is sufficiently fast (§5). While the general status of the entropy-limit conjecture remains open, these new results verify important consequences of the conjecture. Thus, when taken together with previous results (outlined in §2), these new results provide inductive support for the general entropy-limit conjecture.

**Normal models.** While the quest for a viable inductive logic provides key motivation for this research, the results of this paper are also relevant to a rather different problem: the characterisation of the most normal model of a first-order theory.

Consider a first-order language  $\mathcal{L}$ . Let  $T$  be a finite consistent set of first-order axioms in  $\mathcal{L}$ . There have been different approaches to defining the *default* or *most normal* model for  $T$ , depending on how one would interpret the default model. One can, for example, consider the prime models (the smallest canonical models) as default models (see Chang and Keisler [14, p. 96] and Hodges [22, p. 336] for instance). Another approach would be to interpret normality in terms of closure properties and require default models to be, for example, existentially closed. Other approaches have consider the default model as the ‘average’ model and try to characterise this in terms of the distribution of models (see for example Bacchus et al. [5, 6], Grove et al. [18, 19, 20]). Another way that this question can be interpreted was posed in [37], and studied further in [38, 40, 41], as:

Given a finite (consistent) set  $T$  of first-order axioms, from a language  $\mathcal{L}$  and a structure  $\mathcal{M}$  with domain  $\{t_1, t_2, \dots\}$  over  $\mathcal{L}$ , which we only know to be a model of  $T$ , what probability should we assign to a sentence  $\theta(t_1, \dots, t_n)$  being true in  $\mathcal{M}$ ?

Then any set of first-order axioms can be seen as imposing a probability function over the sentences of the language, in which the probability assigned to a sentence  $\theta$  is interpreted as the probability that it will hold in a random model of  $T$ . The question is how a set of first-order axioms can determine a probability function in the most natural way [42].

The constraint that  $\mathcal{M}$  is a model for  $T$  requires the probability assignment to give probability 1 to all sentences in  $T$  and consequently to all sentences logically implied by them. There is, however, a large set of probability functions that will satisfy this constraint but will differ on the probability that they will assign to other sentences in  $\mathcal{L}$ . One can further trim this set by imposing extra conditions on the way that these probabilities are to be assigned. And, by doing so specify what it means for  $\mathcal{M}$  to be the default model.

One example is to make this assignment of probabilities in such a way that captures the notion of averageness or typicality. This in the literature is referred to as the limiting centre of mass assignment (see for example [33, 36, 37]). Another approach, followed in [38, 40, 41], and with which we will be concerned here, characterises a default model as being maximally uninformative the model with respect to the sentences of the language not implied by  $T$ . These maximally uninformative probability assignments are taken to be maximum entropy probability functions.

If the entropy-limit conjecture is true, this would lend support to the claim that maximising entropy leads to a canonical model characterisation for first-order theories. Thus, the results of this paper are relevant to the characterisation of normal models.

## 2 The formal framework

In this section we set out the rudiments of the formal framework and some notational conventions, and we survey previous work relevant to the entropy-limit conjecture.

**The predicate languages.** Throughout this paper we consider a first-order predicate language  $\mathcal{L}$ , with countably many constant symbols  $t_1, t_2, \dots$  and finitely many relation symbols,  $U_1, \dots, U_n$ . The atomic sentences, i.e., sentences of the form  $U_i t_{i_1} \dots t_{i_k}$  where  $k$  is the arity of the relation  $U_i$ , will be denoted by  $a_1, a_2, \dots$ , ordered in such a way that atomic sentences involving only constants among  $t_1, \dots, t_n$  occur before those atomic sentences that also involve  $t_{n+1}$ . We denote the sentences of  $\mathcal{L}$  by  $S\mathcal{L}$ .

We will also consider the finite sublanguages  $\mathcal{L}_n$  of  $\mathcal{L}$ , where  $\mathcal{L}_n$  has only the first  $n$  constant symbols  $t_1, \dots, t_n$  but the same relation symbols as  $\mathcal{L}$ .  $\mathcal{L}_n$  has finitely many atomic sentences  $a_1, \dots, a_{r_n}$ . We call the state descriptions of  $\mathcal{L}_n$  (i.e., the sentences of the form  $\pm a_1 \wedge \dots \wedge \pm a_{r_n}$ ),  $n$ -states. We let  $\Omega_n$  be the set of  $n$ -states for each  $n$ . Note that  $|\Omega_n| = 2^{r_n}$ , and every  $n$ -state  $\omega_n \in \Omega_n$  has  $|\Omega_{n+1}|/|\Omega_n| = 2^{r_{n+1}-r_n}$  many  $n+1$ -states,  $\omega_{n+1}$  which extend it (i.e.,  $\omega_{n+1} \models \omega_n$ ). We denote the sentences of  $\mathcal{L}_n$  by  $S\mathcal{L}_n$ .

We use  $N_\varphi$  (or, when  $\varphi$  is clear from the context, simply  $N$ ) to refer to the largest number  $n$  such that  $t_n$  appears in  $\varphi \in S\mathcal{L}$ .

For a sentence  $\varphi \in S\mathcal{L}$  and fixed  $n \geq N_\varphi$ , we can reinterpret  $\varphi$  as a sentence of  $\mathcal{L}_n$ , by interpreting  $\exists x \theta(x)$  as  $\theta(t_1) \vee \dots \vee \theta(t_n)$  and  $\forall x \theta(x)$  as  $\theta(t_1) \wedge \dots \wedge \theta(t_n)$ . We use the notation  $(\varphi)_n$ , or if there is no ambiguity, simply  $\varphi_n$ , to denote this reinterpretation of  $\varphi$  in  $\mathcal{L}_n$ . For any sentence  $\varphi$  we denote by  $[\varphi]_n$  the set of  $n$ -states that satisfy  $\varphi$ . We denote the number of  $n$ -states in  $[\varphi]_n$  by  $|\varphi|_n$ .

**Example 1.** If  $\varphi$  is  $\forall x \theta(x)$ , then  $[\varphi_n]_n = \{\omega \in \Omega_n : \omega \models \bigwedge_{i=1}^n \theta(t_i)\}$ . For quantification over  $l$  variables we have  $(\forall \vec{x} \theta(\vec{x}))_n = \bigwedge_{j=1}^l \bigwedge_{i_j=1}^n \theta(t_{i_1}, \dots, t_{i_k})$  and  $(\exists \vec{x} \theta(\vec{x}))_n = \bigvee_{j=1}^l \bigvee_{i_j=1}^n \theta(t_{i_1}, \dots, t_{i_k})$ .

**Example 2.** The 1-state  $\omega_1 := Ut_1 t_1$  is a member of  $[(\forall x \exists y Uxy)_1]_1$  but the 2-state  $\omega_2 := Ut_1 t_1 \wedge Ut_1 t_2 \wedge \neg Ut_2 t_1 \wedge \neg Ut_2 t_2$  that extends  $\omega_1$  is not in  $[(\forall x \exists y Uxy)_2]_2$ , even though  $\omega_2 \models \omega_1$ . The 1-state  $\nu_1 := \neg Ut_1 t_1$  is not in  $[(\forall x \exists y Uxy)_1]_1$ ; however, the extending 2-state  $\nu_2 := \neg Ut_1 t_1 \wedge Ut_1 t_2 \wedge Ut_2 t_1 \wedge Ut_2 t_2$  is in  $[(\forall x \exists y Uxy)_2]_2$ .

**Probability.** A probability function  $P$  on  $\mathcal{L}$  is a function  $P : S\mathcal{L} \rightarrow \mathbb{R}_{\geq 0}$  such that:

**P1:** If  $\tau$  is a tautology, i.e.,  $\models \tau$ , then  $P(\tau) = 1$ .

**P2:** If  $\theta$  and  $\varphi$  are mutually exclusive, i.e.,  $\models \neg(\theta \wedge \varphi)$ , then  $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$ .

**P3:**  $P(\exists x\theta(x)) = \sup_m P(\bigvee_{i=1}^m \theta(t_i))$ .

A probability function on  $\mathcal{L}_n$  is defined similarly. We shall use the notation  $\mathbb{P}$  and  $\mathbb{P}_n$  to denote the set of all probability functions on  $\mathcal{L}$  and  $\mathcal{L}_n$  respectively. Conditional probability is defined here in terms of unconditional probabilities:  $P(\theta|\varphi) := P(\theta \wedge \varphi)/P(\varphi)$  if  $P(\varphi) > 0$ .

A probability function is determined by the values it gives to the  $n$ -states, for each  $n$  [46, Theorem 2.34]. Consequently, a probability function is determined by the values it gives to the quantifier-free sentences, a result known as *Gaifman's Theorem* [17].

**Example 3 (Equivocator function).** The equivocator function  $P_{=}$  is defined by:

$$P_{=}(\omega_n) \stackrel{\text{df}}{=} \frac{1}{|\Omega_n|} = \frac{1}{2^n} \text{ for each } n\text{-state } \omega_n \in \Omega_n \text{ and each } n = 1, 2, \dots$$

The equivocator function is a probability function on  $\mathcal{L}$ . It is called the equivocator function because it equivocates between  $n$ -states: it is the only probability function that gives each  $n$ -state the same probability, for each  $n$ . The restriction  $P_{=}|_{\mathcal{L}_n}$  of  $P_{=}$  to  $\mathcal{L}_n$  is a probability function on  $\mathcal{L}_n$ , for any  $n$ . To simplify notation, we will use  $P_{=}$  to refer to these restrictions, as well as to the function on  $\mathcal{L}$  itself. In addition, we will say that a sentence  $\theta$  has measure  $x$  if  $x$  is the probability that the equivocator function attaches to  $\theta$ ,  $P_{=}(\theta) = x$ .

**Entropy.** The  $n$ -entropy of a probability function  $P$  (which is defined on either  $\mathcal{L}$  or  $\mathcal{L}_n$ ) is defined as:

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

We follow the usual conventions in taking  $0 \log 0 = 0$  and the logarithm to be the natural logarithm.

We now turn to entailment relationships in inductive logic of the form

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \psi^Y,$$

where  $\varphi_1, \dots, \varphi_k, \psi \in S\mathcal{L}$  and each  $X_i$  is a member or subset of the unit interval or is omitted (in which case the premiss is called *categorical*). We introduce the two key approaches to making sense of such a relationship, the entropy-limit approach and the maximal-entropy approach.

**The entropy-limit approach.** Suppose  $X_1, \dots, X_n$  are probabilities or closed intervals of probabilities. Let  $N = \max\{N_{\varphi_1}, \dots, N_{\varphi_k}\}$ , so that  $t_N$  is the constant symbol, of all those occurring in  $\varphi_1, \dots, \varphi_k$ , with the largest index. For fixed  $n \geq N$ , reinterpret  $\varphi_1, \dots, \varphi_k$  as statements of  $\mathcal{L}_n$ . Let  $\mathbb{E}_n$  be the set of probability functions on  $\mathcal{L}_n$  that satisfy  $(\varphi_1)_n^{X_1}, \dots, (\varphi_k)_n^{X_k}$ . If  $\mathbb{E}_n \neq \emptyset$  consider the entropy maximiser:

$$P^n \stackrel{\text{df}}{=} \arg \max_{P \in \mathbb{E}_n} H_n(P).$$

Since  $X_1, \dots, X_n$  are probabilities or closed intervals of probabilities,  $\mathbb{E}_n$  is closed and convex and  $P^n$  is uniquely determined. Several considerations point to  $P^n$  as the most appropriate probability function for drawing inferences from premisses on  $\mathcal{L}_n$ , [36]. When characterising normal models of a set of first order axioms, i.e., when  $X_1, \dots, X_n = 1$ ,  $P^n$  can be regarded as the normal probabilistic characterisation of a random model of  $\{\varphi_1, \dots, \varphi_n\}$  with respect to  $\mathcal{L}_n$ , where the normality is understood in terms of being minimally constrained [42].

However, the premisses are intended as statements on  $\mathcal{L}$ , not  $\mathcal{L}_n$ , and the question arises as to what would be the most appropriate probability function for drawing inferences from these premisses when they are interpreted as statements about an infinite domain, or what the default characterisation of random model of the premisses would be with respect to the full language  $\mathcal{L}$ . If it exists, one can consider the function  $P^\infty$  defined on  $\mathcal{L}$  as a pointwise limit of maximum entropy functions [7]:

$$P^\infty(\theta) \stackrel{\text{df}}{=} \lim_{n \rightarrow \infty} P^n(\theta).$$

The entropy-limit approach takes  $P^\infty$  for inference, attaching probability  $Y = P^\infty(\psi)$  to sentence  $\psi$ .

There is one complication about the definition of  $P^\infty$  which we need to address. While Barnett and Paris [7] define  $P^\infty$  in terms of a pointwise limit where the limit is taken independently for each sentence of  $\mathcal{L}$ , Paris and Rafiee Rad [38], Rafiee Rad [39] and Rafiee Rad [41] define  $P^\infty$  in a slightly different way: take the pointwise limit on quantifier-free sentences and extend this to the (unique) probability function on  $\mathcal{L}$  as a whole which agrees with the values obtained on the quantifier-free sentences, assuming that the pointwise limit exists and satisfies the axioms of probability on quantifier-free sentences of  $\mathcal{L}$  [17]. The Rad-Paris definition circumvents a problem that can arise with the Barnett-Paris definition, namely that the pointwise limit on  $\mathcal{L}$  as whole may exist but may fail to be a probability function (see Appendix A.1 for a discussion of this point). Since the entropy-limit conjecture with respect to the Rad-Paris definition implies the entropy-limit conjecture with respect to the Barnett-Paris definition, we consider the Rad-Paris definition of  $P^\infty$  in this paper, with the aim of proving stronger results.

Note that  $P^n$  and  $\mathbb{E}_n$  are defined on  $\mathcal{L}_n$  not  $\mathcal{L}$ . To simplify notation, when  $P$  is defined on  $\mathcal{L}$ , we will say  $P \in \mathbb{E}_n$  to mean that the restriction of  $P$  to  $\mathcal{L}_n$  is in  $\mathbb{E}_n$ ,  $P|_{\mathcal{L}_n} \in \mathbb{E}_n$ .

**The maximal-entropy approach.** This alternative approach avoids appealing to the finite sublanguages  $\mathcal{L}_n$ . Instead, consider  $\mathbb{E}$ , the set of probability functions on  $\mathcal{L}$  that satisfy the premisses  $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ . For probability functions  $P$  and  $Q$  defined on  $\mathcal{L}$ ,  $P$  is deemed to have greater entropy than function  $Q$  if it has greater  $n$ -entropy for sufficiently large  $n$ , i.e., if there is some natural

number  $N$  such that for all  $n \geq N$ ,  $H_n(P) > H_n(Q)$ . Then we can consider the set of probability functions in  $\mathbb{E}$  with maximal entropy:

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

If  $\text{maxent } \mathbb{E} \neq \emptyset$ , one can draw inferences using the maximal entropy functions  $P^\dagger$ . Thus, the maximal-entropy approach attaches the set of probabilities  $Y = \{P^\dagger(\psi) : P^\dagger \in \text{maxent } \mathbb{E}\}$  to  $\psi$ . Alternatively, if the premisses are categorical one can take  $P^\dagger$  as the default probabilistic descriptions of a random model of the premisses.

See Landes and Williamson [34] and Williamson [46, Chapter 9] for one kind of justification of this approach.

**What is known so far.** The entropy-limit conjecture says that if  $P^\infty$  exists and is in  $\mathbb{E}$ , then  $\text{maxent } \mathbb{E} = \{P^\infty\}$ . The majority of work in the literature concerning the conjecture deals with the special case of categorical premisses and concerns the probabilistic characterisation of models of a set of first order axioms. Barnett and Paris [7] study monadic first order languages and show that the limit entropy approach is well defined, i.e.,  $P^\infty$  exists, for a generalised set of linear constraints (i.e. categorical and non-categorical premisses) on such languages. Rafiee Rad [39] considers the special case of a set of first order axioms on monadic languages and shows both the exact form of  $P^\infty$  and that the entropy-limit conjecture holds for these languages—see Rafiee Rad [39, Theorem 29], and Rafiee Rad [42] for a more general case. Similarly, Rafiee Rad [41] shows the exact form of  $P^\infty$  and that the conjecture is true in the categorical  $\Sigma_1$  case, i.e., the case in which the premiss propositions  $\varphi_1, \dots, \varphi_k$  are all  $\Sigma_1$  statements [41, Corollary 1].

Rafiee Rad [40] shows that there exist cases in which  $\text{maxent } \mathbb{E} = \emptyset$ : for any probability function satisfying the premiss  $\exists x \forall y Rxy$  there is another probability function with greater entropy that also satisfies that premiss. This involves considering sets of axioms with quantifier complexity of  $\Sigma_2$ . (It is not known whether the entropy limit exists in this scenario.) Similarly, [39, §3.2] provides cases, with premisses of  $\Pi_2$  quantifier complexity, in which  $P^\infty$  is not well defined. (It is not yet known whether or not the maximal entropy approach can yield an answer for those cases.) See also Rafiee Rad [42] for a more general case. On the other hand, Rafiee Rad [39, §4.1] shows that there are cases in which  $P^\infty$  does not exist but  $P^\dagger$  does.

This leaves open the case concerning sets of axioms with quantifier complexity  $\Pi_1$  as well as non-categorical premisses for polyadic languages. Paris and Rafiee Rad [38] investigate the existence of  $P^\infty$  for sets of  $\Pi_1$  sentences and show that for a special case, which they call the *slow*  $\Pi_1$  sentences, the entropy-limit approach is well defined.

**Plan of the paper.** In §3 we show the entropy-limit conjecture holds in cases involving categorical premisses (i.e., premisses that take the form of sentences of the predicate language  $\mathcal{L}$  without probabilities attached) of  $\Pi_1$  quantifier complexity. In §4 we extend these cases to ones in which the premiss sentences do have probabilities attached. In §5 we provide a general result which shows that the entropy-limit conjecture holds in certain general cases where the  $P^n$  converge fast enough to  $P^\infty$ . We sum up in §6.

Premises	$P^\dagger$	$P^\infty$	$P^\dagger = P^\infty?$
$\Delta_0$	exists and unique	exists and unique	✓
$\Sigma_1$	exists and unique	exists and unique	✓
$\Pi_1$	?	?	?
$\Sigma_2$	not always well-defined	?	?
$\Pi_2$	?	not always well-defined	?

Table 1: Summary of what is known so far with respect to entropy maximisers for categorical premisses.

**Summary of key notation.** Key notation is summarised in Table 2. Note that we use  $\chi$  to denote the set of constraints (premisses) currently under investigation. In cases where the constraints vary, we subscript  $P^\infty$  or  $P^\dagger$  with the constraint currently operating.

$r_n$	number of atomic sentences of $\mathcal{L}_n$
$\Omega_n$	set of $n$ -states. N.b., $ \Omega_n  = 2^{r_n}$
$N_\varphi$ or $N$	maximum $n$ such that $t_n$ appears in $\varphi$
$(\varphi)_n$ or $\varphi_n$	reinterpretation of $\varphi$ in $\mathcal{L}_n$
$[\varphi]_n$	set of $n$ -states satisfying $\varphi$
$ \varphi _n$	number of $n$ -states satisfying $\varphi$
$P_-$	equivocator function
$P^\infty$	entropy-limit function
$P^\dagger$	maximal-entropy function
$\chi$	set of constraints under investigation

Table 2: Key notation

### 3 The categorical $\Pi_1$ constraint

In this section we show that the entropy-limit conjecture holds in the case in which there is a single categorical constraint  $\varphi$  which takes the form of a satisfiable  $\Pi_1$  sentence  $\forall \vec{x}\theta(\vec{x})$ . As we shall now explain, this situation splits naturally into two cases: that in which  $\varphi$  has non-zero measure,  $P_-(\forall \vec{x}\theta(\vec{x})) > 0$ , explored in §3.1, and that in which  $\varphi$  has zero measure,  $P_-(\forall \vec{x}\theta(\vec{x})) = 0$ , explored in §3.2.

For all satisfiable  $\exists \vec{x}\theta(\vec{x}) \in \Sigma_1$  it is known that for  $N = N_{\exists \vec{x}\theta(\vec{x})}$  ([41, Theorem 4] and [41, Corollary 1]):

$$P_{\exists \vec{x}\theta(\vec{x})}^\infty(\cdot) = P_-(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x}\theta(\vec{x}) \text{ is consistent}}} \omega_N) = P_-(\cdot \mid \exists \vec{x}\theta(\vec{x})) = P_{\exists \vec{x}\theta(\vec{x})}^\dagger(\cdot) .$$

Intuitively, since  $\Sigma_1$  and  $\Pi_1$  are natural duals, one suspects that something similar is true for all satisfiable  $\forall \vec{x}\theta(\vec{x}) \in \Pi_1$  and  $N = N_{\forall \vec{x}\theta(\vec{x})}$ :

$$P_{\forall \vec{x}\theta(\vec{x})}^\infty(\cdot) = P_-(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \models \forall \vec{x}\theta(\vec{x})}} \omega_N) = P_-(\cdot \mid \forall \vec{x}\theta(\vec{x})) = P_{\forall \vec{x}\theta(\vec{x})}^\dagger(\cdot) .$$

Unfortunately, this intuition only gets us so far, because there are no  $\omega_N$  to include in the disjunction iff  $\forall \vec{x} \theta(\vec{x})$  has measure zero (Proposition 8). In this case, the intuition breaks down because conditioning on a zero-probability sentence is not defined. On the other hand, the intuition is correct as long as at least one such  $\omega_N$  exists (Corollary 9). The set of such  $\Pi_1$ -sentences is characterised in Proposition 8.

The case of satisfiable  $\forall \vec{x} \theta(\vec{x}) \in \Pi_1$  with measure zero is much harder, since  $P_{=}(\cdot | \forall \vec{x} \theta(\vec{x}))$  is simply not defined. Nevertheless, in Theorem 14 we prove that Entropy Limit Conjecture does also hold for all measure zero  $\forall \vec{x} \theta(\vec{x}) \in \Pi_1$ . The technical difficulty is precisely that of defining a “conditional probability” on sentences which have measure zero.

The following proposition plays a key part in many of our later proofs.

**Proposition 4.** *For all  $\emptyset \neq S$  and all  $x > 0$  it holds that*

$$\begin{aligned} & \arg \sup_{f: S \rightarrow [0, x] : \sum_{y \in S} f(y) = x} - \sum_{y \in S} f(y) \log(f(y)) \\ &= \frac{1}{x} \arg \sup_{g: S \rightarrow [0, 1] : \sum_{y \in S} g(y) = 1} - \sum_{y \in S} g(y) \log(g(y)) . \end{aligned}$$

This proposition implies that entropy maximisation over a subset  $S$  of  $n$ -states under a linear constraint can be achieved by maximising the  $n$ -entropy of probability functions assigning  $S$  joint probability one and pointwise re-scaling the maximal  $n$ -entropy function.

**Proof:** It suffices to note that the  $n$ -entropy of  $g/x$  over  $S$  is an affine-linear transformation of the  $n$ -entropy of the probability function  $g$  over  $S$ :

$$\begin{aligned} - \sum_{y \in S} \frac{g(y)}{x} \log\left(\frac{g(y)}{x}\right) &= - \frac{1}{x} \sum_{y \in S} g(y) (\log(g(y)) - \log(x)) \\ &= \frac{\log(x)}{x} + \frac{H_n(g)}{x} . \end{aligned}$$

■

We use this proposition to show that distributing probability mass more uniformly increases  $n$ -entropy in the following sense. Consider two probability functions  $P, Q \in \mathbb{P}$  which disagree on non-empty subset  $S$  of  $n$ -states such that  $a := \sum_{\omega_n \in S} P(\omega_n) = \sum_{\omega_n \in S} Q(\omega_n)$  where  $P$  is uniform on  $S$  ( $P(\omega_n) = a/|S|$ ) and  $Q$  is not. Then  $P$  has greater  $n$ -entropy than  $Q$ ,  $H_n(P) > H_n(Q)$ .

The following lemma will be important in the remainder of this section. Recall that  $P_{=}$  is used to refer to the equivocator function both on  $\mathcal{L}$  and its finite sublanguages  $\mathcal{L}_n$ :

**Lemma 5** (Local Entropy Maximisation). *For all  $\gamma \in S\mathcal{L}$  with  $\mathbb{E} = \{P \in \mathbb{P} : P(\gamma) = 1\}$ , all  $n \in \mathbb{N}$  and all  $\psi \in S\mathcal{L}_n$ , if  $\gamma_n$  is satisfiable, then*

$$P^n(\cdot) = P_{=}(\cdot | \gamma_n) .$$

**Proof:** By definition of  $P^n$ ,  $P^n$  can only assign non-zero probability to those  $n$ -states which are in  $[\gamma_n]_n$ . Since entropy is maximal if the probabilities are uniform (Proposition 4), it follows that  $P^n$  has to assign equal probability to all

$n$ -states in  $[\gamma_n]_n$ —if  $[\gamma_n]_n \neq \emptyset$ . By assumption we have that  $[\gamma_n]_n \neq \emptyset$ , and thus all  $n$ -states in  $[\gamma_n]_n$  are assigned the same probability by  $P^n$  and these probabilities sum to one. Hence,  $P^n(\psi) = P_=(\psi|\gamma_n)$  for all  $n \in \mathbb{N}$  and all  $\psi \in S\mathcal{L}_n$ . ■

First we consider the case in which  $\forall \vec{x}\theta(\vec{x})$  has positive measure.

### 3.1 Non-zero measure, $P_=(\forall \vec{x}\theta(\vec{x})) > 0$

**Remark 6.** Using Lemma 5 it is easy to see that for all sentences  $\gamma \in S\mathcal{L}$  with positive measure, if for all  $\chi \in QFS\mathcal{L}$   $\lim_{n \rightarrow \infty} P_=(\chi|\gamma_n)$  exists, then

$$P_\gamma^\infty(\chi) = \lim_{n \rightarrow \infty} P_=(\chi|\gamma_n) . \quad (1)$$

In other words,  $P^\infty$  is obtained by considering the limit of equivocators conditionalised on the premiss reinterpreted on  $\mathcal{L}_n$ . The above lemma tells us what these probability functions, the  $P^n$ , look like.

**Theorem 7.** For all  $\gamma \in \Pi_1 \cup \Sigma_1$ , if one of the following two conditions hold

1.  $\lim_{n \rightarrow \infty} P_=(\gamma_n) \in (0, 1)$  and for all  $\chi \in QFS\mathcal{L}$ ,  $\lim_{n \rightarrow \infty} P_=(\chi|\gamma_n)$  exists,
2.  $\lim_{n \rightarrow \infty} P_=(\gamma_n) = 0$ ,

then

$$P_{\neg\gamma}^\infty = \frac{P_=(\neg\gamma) - P_=(\gamma) \cdot P_\gamma^\infty}{P_=(\neg\gamma)} . \quad (2)$$

**Proof:** For all  $\chi \in QFS\mathcal{L}$  and all large enough  $n$  such that  $P_=(\neg(\gamma_n)) > 0$ , we have, by the law of total probability, that<sup>2</sup>

$$P_=(\chi) = P_=(\gamma_n) P_=(\chi|\gamma_n) + P_=(\neg\gamma_n) P_=(\chi|\neg\gamma_n) . \quad (3)$$

Then notice that by assumption  $\lim_{n \rightarrow \infty} P_=(\gamma_n) \in [0, 1]$  exists and  $\lim_{n \rightarrow \infty} P_=(\neg\gamma_n) = \lim_{n \rightarrow \infty} 1 - P_=(\gamma_n) = 1 - \lim_{n \rightarrow \infty} P_=(\gamma_n) \in (0, 1]$ . Then since the limits exist, it must be the case that  $\lim_{n \rightarrow \infty} P_=(\gamma_n) = P_=(\gamma)$ , by Gaifman's Theorem.<sup>3</sup> Then  $\lim_{n \rightarrow \infty} P_=(\neg\gamma_n) = 1 - P_=(\gamma) = P_=(\neg\gamma)$ .

<sup>2</sup>If  $P_=(\gamma_n) = 0$ , then put for sake of simplicity  $P_=(\gamma_n) \cdot P_=(\chi|\gamma_n) := 0$ . In actual fact,  $P_=(\chi|\gamma_n)$  is undefined.

<sup>3</sup>Note that for quantifier free  $\gamma(\vec{x})$ , we have  $Q\vec{x}\gamma(\vec{x}) \in \Pi_1 \cup \Sigma_1$  for  $Q \in \{\forall, \exists\}$ . The equality  $P(Q\vec{x}\gamma(\vec{x})) = \lim_{n \rightarrow \infty} P(\gamma_n) \in [0, 1]$  for all probability functions  $P$  is an application of Gaifman's Theorem [35, Lemma 3.8]. In particular, these limits always exist.

It is, however, not always the case that  $P(\delta) = \lim_{n \rightarrow \infty} P(\delta_n)$ . Consider the contingent sentence  $\delta \in \Pi_2$ , which only has infinite models:

$$\delta = \forall xzw \exists y (U_1xy \wedge [(U_1xw \wedge U_1wz) \rightarrow U_1xz] \wedge \neg U_1xx) .$$

It is easy to check that open orders satisfy  $\delta$ . On the other hand,  $\delta$  does not have a finite model: for every element in the support  $a$  of a model of  $\delta$  there has to exist some other element  $b$  which is greater than  $a$ ,  $U_1ab$ . Note that  $a \neq b$  ( $U_1$  is irreflexive). If  $U_1$  were cyclic, then transitivity entails  $U_1dd$  for some  $d$  which contradicts  $\delta$ . Hence, for every element in the finite model there must be some other element which is greater. Contradiction. So,  $\delta$  only has infinite models. It holds for all probability functions  $P \in \mathbb{P}$  that  $P(\delta_n) = 0$ , since there are no finite models of  $\delta$ . Since,  $\delta$  is satisfiable, there exist probability functions  $P' \in \mathbb{P}$  such that  $P'(\delta) > 0$  ([35, p. 189]). And thus  $P'(\delta_n) = 0 < P'(\delta)$  for such a probability function  $P'$ .

Assume that  $P_=(\gamma) > 0$  and thus for large enough  $n$ :  $P_=(\gamma_n) > 0$  and  $P(\cdot|\gamma_n)$  is well-defined. By Remark 6,  $\lim_{n \rightarrow \infty} P_=(\chi|\gamma_n) = P_\gamma^\infty(\chi)$ . Taking the limit of 3 we obtain

$$\begin{aligned} P_=(\chi) &= P_=(\gamma) \cdot P_\gamma^\infty(\chi) + P_=(\neg\gamma) \cdot \lim_{n \rightarrow \infty} P_=(\chi|\neg\gamma_n) \\ &= P_=(\gamma) \cdot P_\gamma^\infty(\chi) + P_=(\neg\gamma) \cdot P_{\neg\gamma}^\infty(\chi) \end{aligned}$$

where the last equality follows from Remark 6, if the limit exists. But notice that

$$\lim_{n \rightarrow \infty} P_=(\chi|\neg(\gamma_n)) = \frac{P_=(\chi) - P_=(\gamma)P_\gamma^\infty(\chi)}{P_=(\neg\gamma)}.$$

Thus, the limit exists as the right hand side is well defined. Then (2) holds for all quantifier free sentences and thus by Gaifman's Theorem it holds for all sentences in  $S\mathcal{L}$ .

Next assume that  $P_=(\gamma) = 0$ . Taking the limit of 3 we obtain

$$\begin{aligned} P_=(\chi) &= P_=(\gamma) \cdot P_\gamma^\infty(\chi) + P_=(\neg\gamma) \cdot \lim_{n \rightarrow \infty} P_=(\chi|\neg\gamma_n) \\ &= P_=(\neg\gamma) \cdot \lim_{n \rightarrow \infty} P_=(\chi|\neg\gamma_n) = \lim_{n \rightarrow \infty} P_=(\chi|\neg\gamma_n). \end{aligned}$$

So, the limit on the right hand side exists and is equal to  $P_=(\neg\gamma)$ . Now conclude as above that  $P_=(\chi) = P_{\neg\gamma}^\infty(\chi)$ . ■

While Theorem 7 is informative about the entropy limit, it leaves open certain questions. When is it the case that  $\forall \vec{x} \theta(\vec{x}) \in \Pi_1$  has positive measure? What exactly does the entropy limit look like? And does the entropy-limit conjecture hold in that case? We shall address these questions in turn.

**Proposition 8.** *For all  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  the following are equivalent*

1.  $P_=(\varphi) > 0$
2. Let  $\forall \vec{x} \bigwedge_{i \in I} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})$  be the conjunctive normal form of  $\varphi$ . Then for all  $i \in I$  there exists a  $j_i^*$  such that  $\lambda_{ij_i^*}(\vec{x}, \vec{t})$  does not mention a variable and  $\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t})$  is consistent.
3.  $\models \varphi \iff (\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \vee \varphi)$ .
4. There exists an  $\omega_{N_\varphi} \in \Omega_{N_\varphi}$  such that  $\omega_{N_\varphi} \models \varphi$ .

For example,  $P_=(\forall x(Ux \rightarrow Ut_1)) = P_=(Ut_1 \vee \forall x \neg Ux) = 0.5 > 0$  but  $P_=(\forall x((U_1x \vee U_2t_1) \wedge (U_3x \vee \neg U_2t_1))) = 0$ . In the latter example, there is no  $n$ -state entailing  $\forall x((U_1x \vee U_2t_1) \wedge (U_3x \vee \neg U_2t_1))$  because the conjunction of the two literals not mentioning a variable,  $U_2t_1 \wedge \neg U_2t_1$ , is inconsistent.

**Proof:** To simplify notation we let  $N := N_\varphi$ .

2  $\Rightarrow$  1: If  $\varphi \in \Pi_1$  is of this form, then

$$P_=(\varphi) = P_=(\forall \vec{x} \bigwedge_{i \in I} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})) \geq P_=(\forall \vec{x} \bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{x}, \vec{t}))$$

$$= P_{=} \left( \bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \right) = \frac{1}{2^{|I|}} > 0.$$

1  $\Rightarrow$  2. We show that the negation of 2 entails the negation of 1,  $P_{=}(\varphi) = 0$ . If  $\varphi \in \Pi_1$  is not of this form then denote by  $\emptyset \subseteq I^* \subseteq I$  those indices for which every  $\lambda_{ij_i}$  contains a variable. We consider two cases: first suppose that  $I^*$  is not empty and let  $i \in I^*$ .

Note that  $P_{=}(\forall \vec{x} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})) \stackrel{P3}{=} \lim_{n \rightarrow \infty} P_{=}((\forall \vec{x} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t}))_n)$ . Let us suppose for the moment that  $\varphi$  contains only a single variable, say  $x$ . Let  $n > N$  and let  $\omega_n$  be an  $n$ -state which satisfies  $(\forall x \bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_n$ . We now count the number  $n + 1$ -states which extend  $\omega_n$  and satisfy  $(\forall x \bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_{n+1}$ . Notice that  $\omega_n$  has  $2^{r_{n+1}-r_n}$  extensions to  $\mathcal{L}_{n+1}$ . Those that satisfy  $(\forall x \bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_{n+1}$  are precisely those  $n + 1$ -states which satisfy  $(\forall x \bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_n \wedge \bigvee_{j \in J_i} \lambda_{ij}(t_{n+1}, \vec{t})$ . There are

$$\frac{2^{|J_i|} - 1}{2^{J_i}} \cdot 2^{r_{n+1}-r_n}$$

such  $n + 1$ -states. So,

$$P_{=}(\forall x (\bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_{n+1}) = \frac{2^{|J_i|} - 1}{2^{|J_i|}} P_{=}(\forall x (\bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_n).$$

Thus,

$$\lim_{n \rightarrow \infty} P_{=}(\forall x (\bigvee_{j \in J_i} \lambda_{ij}(x, \vec{t}))_n) \leq \lim_{n \rightarrow \infty} \left( \frac{2^{|J_i|} - 1}{2^{|J_i|}} \right)^{n-N} = 0.$$

If  $\varphi$  contains two or more variables, then the  $()_n$ -operation leads to more conjunctions than in the single-variable case. Hence, when counting the  $n + 1$ -states which satisfy  $(\forall \vec{x} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{a}))_{n+1}$  there is an even greater number which we subtract from  $2^{|J_i|}$ . The limit is hence equal to zero, too. And so

$$\begin{aligned} P_{=}(\varphi) &= P_{=}(\forall \vec{x} \bigwedge_{i \in I} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})) \leq P_{=}(\forall \vec{x} \bigwedge_{i \in I^*} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})) \\ &\leq \sup_{i \in I^*} P_{=}(\forall \vec{x} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t})) = \sup_{i \in I^*} \lim_{n \rightarrow \infty} P_{=}(\forall \vec{x} (\bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t}))_n) = 0. \end{aligned}$$

For the second case, suppose that  $I^*$  is empty. Then for every  $i \in I$  there exists at least one  $\lambda_{ij_i}$  which does not contain a variable. Furthermore, since we are assuming that  $\varphi$  is not of the form given in (2), for every such choice of  $j_i$ ,  $\bigwedge_{i \in I} \lambda_{ij_i}$  is inconsistent.

So, if the  $n$ -state  $\omega_n$  is such that  $\omega_n \models \varphi_n = (\forall \vec{x} \bigwedge_{i \in I} \bigvee_{j \in J_i} \lambda_{ij}(\vec{x}, \vec{t}))_n$  ( $n \geq N$ ),  $\omega_n$  cannot satisfy  $\varphi_n$  by only satisfying a variable-free literal from each conjunct (since, as mentioned above, they are jointly inconsistent). Then there has to exist an  $i_0 \in I$  such that  $\omega_n$  is inconsistent with all variable-free literals in  $\bigvee_{j \in J_{i_0}} \lambda_{i_0 j_{i_0}}(\vec{t})$ . Thus,  $\omega_n$  and all its extensions  $\omega_m$  which satisfy  $\varphi$  must satisfy literals in  $\bigvee_{j \in J_{i_0}} \lambda_{i_0 j_{i_0}}$ , which contain a variable. For the purposes of counting extensions, we might as well ignore the variable-free literals of  $\bigvee_{j \in J_{i_0}} \lambda_{i_0 j_{i_0}}$ . We may now proceed as if  $I^*$  is not empty.

3  $\Rightarrow$  1. Since  $\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t})$  is consistent, this easily follows:

$$P_{=}(\varphi) = P_{=}(\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \vee \varphi) \geq P_{=}(\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t})) > 0 .$$

2  $\Rightarrow$  3.  $\models \varphi \rightarrow (\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \vee \varphi)$  is trivially true.

For the direction from right to left it suffices to notice that by 2,  $\models \bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \rightarrow \varphi$ .

3  $\Rightarrow$  4. Notice that  $\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \in QFSL_N$  since each conjunct only involves constants, and  $N$  is the largest constant appearing in  $\theta(\vec{x})$ . And that it is consistent by assumption since we have shown that 3 implies 1 and 1 implies 2. Hence, the conjunction is entailed by some  $N$ -state  $\omega_N \in \Omega_N$ . Then  $\omega_N$  entails the logically weaker  $\bigwedge_{i \in I} \lambda_{ij_i^*}(\vec{t}) \vee \varphi$ .

4  $\Rightarrow$  1.  $P_{=}(\varphi) \geq P_{=}(\omega_N) > 0$ , where the strict inequality follows from the definition of  $P_{=}$ .  $\blacksquare$

**Corollary 9.** For all  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  such that  $P_{=}(\varphi) > 0$  it holds that

$$P_{\varphi}^{\infty}(\cdot) = P_{=}(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \models \varphi}} \omega_N) .$$

where  $N = N_{\varphi}$ , i.e., the maximum  $n$  such that  $t_n$  appears in  $\varphi$ .

**Proof:** Since  $\varphi \in \Pi_1$ , if  $P_{=}(\varphi) = 1$  then  $\varphi$  is a tautology, as is  $\varphi_N$ , so  $\{\omega_N \in \Omega_N : \omega_N \models \varphi\} = \Omega_N$ . Hence,  $P_{=} \in \mathbb{E}_n$  for all  $n \geq N$  and so for all  $n > N$ ,  $P_{\varphi}^n = P_{=}$  and thus  $P_{\varphi}^{\infty} = P_{=} = P_{=}(\cdot \mid \Omega_N)$ .

If  $P_{=}(\varphi) < 1$  then, first, by [41, Theorem 4]:

$$P_{\exists \vec{x} \neg \theta(\vec{x})}^{\infty}(\cdot) = P_{=}(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \neg \theta(\vec{x}) \text{ is consistent}}} \omega_N) .$$

and notice that  $\lim_{n \rightarrow \infty} P_{=}(\cdot \mid (\exists \vec{x} \neg \theta(\vec{x}))_n) = 1 - P_{=}(\varphi)$  [35, Lemma 3.8]. The limit is hence well-defined and in  $[0, 1)$ . Thus we can use Theorem 7 to obtain

$$P_{\varphi}^{\infty}(\cdot) = \frac{P_{=}(\cdot) - P_{=}(\exists \vec{x} \neg \theta(\vec{x})) \cdot P_{=}(\cdot \mid \exists \vec{x} \neg \theta(\vec{x}))}{P_{=}(\varphi)} . \quad (4)$$

Next, applying [41, Lemma 5], we have

$$P_{=}(\exists \vec{x} \neg \theta(\vec{x})) = P_{=}(\bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \neg \theta(\vec{x}) \text{ is consistent}}} \omega_N) .$$

So,

$$P_{=}(\varphi) = 1 - P_{=}(\bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \neg \theta(\vec{x}) \text{ is consistent}}} \omega_N) .$$

Inserting this back into (4) gives

$$P_{\varphi}^{\infty}(\cdot) [1 - P_{=}(\bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \neg \theta(\vec{x}) \text{ is consistent}}} \omega_N)]$$

$$\begin{aligned}
&= P_{=}(\neg \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \rightarrow \theta(\vec{x}) \text{ is consistent}}} \omega_N) \cdot P_{\varphi}^{\infty}(\cdot) \\
&= P_{=}(\cdot) - P_{=}(\bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \rightarrow \theta(\vec{x}) \text{ is consistent}}} \omega_N) \cdot P_{=}(\cdot | \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \rightarrow \theta(\vec{x}) \text{ is consistent}}} \omega_N) .
\end{aligned}$$

Applying the Theorem of Total Probability to  $P_{=}(\cdot)$  on the right hand side and simplifying the equation, we obtain

$$\begin{aligned}
P_{\varphi}^{\infty}(\cdot) &= P_{=}(\cdot | \neg(\bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \exists \vec{x} \rightarrow \theta(\vec{x}) \text{ is consistent}}} \omega_N)) \\
&= P_{=}(\cdot | \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \models \forall \vec{x} \theta(\vec{x})}} \omega_N) .
\end{aligned}$$

■

**Lemma 10.** *Suppose  $\varphi \in \Pi_1$  is such that  $P_{=}(\varphi) > 0$  and  $N = N_{\varphi}$ . Then for all  $N$ -states  $\omega_N$  such that  $\omega_N \in [\varphi_N]_N$  and  $\omega_N \not\models \varphi$  it holds that*

$$|\{\omega_n \in \Omega_n : \omega_n \models \omega_N \ \& \ \omega_n \in [\varphi_n]_n\}| \cdot \frac{|\Omega_N|}{|\Omega_n|} .$$

*converges exponentially (or faster) to zero in  $n > N$ .*

This lemma says that  $N$ -states compatible with the constraint on  $\mathcal{L}_N$  which do not entail  $\varphi$  have only very few extensions which satisfy the constraints on more expressive languages.

**Proof:** As above, write  $\varphi$  in conjunctive normal form, say  $\bigwedge_{i \in I} \bigvee_{j_i \in J_i} \lambda_{ij_i}(\vec{x}, \vec{t})$ , with the standard convention that  $\lambda_{ij_i}(\vec{x}, \vec{t})$  means that  $\lambda_{ij_i}(\vec{x}, \vec{t})$  mentions at most the variables in  $\vec{x}$  and at most the constant symbols in  $\vec{t}$  but not necessarily all of them.

First notice that for  $\omega_N \in [\varphi_N]_N$  such that  $\omega_N \not\models \varphi$ , there has to exist at least one disjunction, say  $\bigvee_{j_i \in J_i} \lambda_{ij_i}(\vec{x}, \vec{t})$ , such that  $\omega_N$  fails to entail all literals in  $\lambda_{ij_i}(\vec{x}, \vec{t})$  that do not mention a variable. To see this, notice that if  $\omega_N$  entails one such literal in every disjunction then  $\omega_N$  entails the whole disjunction and thus  $\varphi$ . That contradicts our assumption. Notice that if the literal does mention a variable then  $\omega_N$  cannot entail it, since  $\omega_N$  only mentions constants  $t_1, \dots, t_N$  but no variable. We now let  $\Delta(\vec{x}, \vec{t}) := \bigvee_{j_i \in J_i} \lambda_{ij_i}(\vec{x}, \vec{t})$  be one such disjunction.

Next notice that this also holds for any  $\omega_n$  extending  $\omega_N$ . That is every  $\omega_n$  which extends  $\omega_N$  fails to satisfy all literals in  $\Delta(\vec{x}, \vec{t})$  which do not mention a variable. To see this notice that all literals in  $\Delta(\vec{x}, \vec{t})$  only mention constants  $t_1, \dots, t_N$ , and since  $\omega_n$  agrees with  $\omega_N$  on  $\mathcal{L}_n$ , if it satisfies any such literal, that literal is satisfied by  $\omega_N$ , which cannot be the case, as just discussed above.

Now consider an  $n$ -state  $\omega_n \in [\varphi_n]_n$  which is logically equivalent to  $(\bigwedge_{i \in I} \bigvee_{j_i \in J_i} \lambda_{ij_i}(\vec{x}, \vec{t}))_n$ . Since  $\omega_n$  does not satisfy a single literal  $(\lambda_{ij_i}(\vec{x}, \vec{t}))_n$  of  $\Delta(\vec{x}, \vec{t})$  which does not mention a variable,  $\omega_n$  must satisfy the interpretation of at least one of its literals, say  $\lambda_{ij_i}(\vec{x}, \vec{t})$ , mentioning a variable in  $\mathcal{L}_n$ . In  $\lambda_{ij_i}(\vec{x}, \vec{t})$  a variable has been replaced by a constant by the  $(\ )_n$ -operation.

Let us consider  $\omega_{n+1}$  that extends  $\omega_n$  and  $\omega_{n+1} \models (\forall \vec{x} \Delta(\vec{x}, \vec{t}))_{n+1}$ . Since  $\omega_n$  mentions all constants in  $\Delta$  ( $n \geq N$ ), its extension,  $\omega_{n+1}$  cannot satisfy a literal in  $(\Delta(\vec{x}, \vec{t}))_{n+1}$  with no variable replaced by a constant. This is so because by the discussion above  $\omega_n$  does not satisfy any such literal in  $\Delta(\vec{x}, \vec{t})$ . But since  $\omega_{n+1} \models (\forall \vec{x} \Delta(\vec{x}, \vec{t}))_{n+1}$  it has to satisfy, in  $\mathcal{L}_{n+1}$ , the interpretation of one literal in  $\Delta(\vec{x}, \vec{t})$  with variables. That is, it has to satisfy some literal, with variables, of  $\Delta(\vec{x}, \vec{t})$  in which the variable is replaced by a constant.

Since  $\varphi \in \Pi_1$ , for all variables there is at least one  $\forall$ -quantifier in front of the  $\Delta(\vec{x}, \vec{t})$  which binds them. Since  $\omega_{n+1} \models (\forall \vec{x} \Delta(\vec{x}, \vec{t}))_{n+1}$ , we have  $\omega_{n+1} \models \Delta(t_{n+1}, \dots, t_{n+1}, \vec{t})$  (we have instantiated all the universally quantified variables with  $t_{n+1}$ ). And so

$$\begin{aligned} & \{\omega_{n+1} \in \Omega_{n+1} \mid \omega_{n+1} \models \omega_n \wedge \varphi_{n+1}\} \\ & \subseteq \{\omega_{n+1} \in \Omega_{n+1} \mid \omega_{n+1} \models \omega_n \ \& \ \omega_{n+1} \models \Delta(t_{n+1}, \dots, t_{n+1}, \vec{t})\} \\ & = \{\omega_{n+1} \in \Omega_{n+1} \mid \omega_{n+1} \models \omega_n \ \& \ \omega_{n+1} \not\models \neg \Delta(t_{n+1}, \dots, t_{n+1}, \vec{t})\} \\ & = \{\omega_{n+1} \in \Omega_{n+1} \mid \omega_{n+1} \models \omega_n \ \& \ \omega_{n+1} \not\models \bigwedge_{j_i} \neg \lambda_{ij_i}(t_{n+1}, \dots, t_{n+1}, \vec{t})\} . \end{aligned}$$

Let  $d$  be the maximal number of literals in any disjunction in the CNF of  $\varphi$ . In particular,  $\bigwedge_{j_i} \neg \lambda_{ij_i}(t_{n+1}, \dots, t_{n+1}, \vec{t})$  has no more than  $d$  literals — which is independent of  $n$ . And thus

$$|\{\omega_{n+1} \in \Omega_{n+1} \mid \omega_{n+1} \models \omega_n \wedge \varphi_{n+1}\}| \leq \frac{|\Omega_{n+1}|}{|\Omega_n|} \frac{2^d - 1}{2^d} .$$

And thus for all  $n > N$

$$\begin{aligned} |\{\omega_n \in \Omega_n : \omega_n \models \omega_N \ \& \ \omega_n \in [\varphi_n]_n\}| \cdot \frac{|\Omega_N|}{|\Omega_n|} & \leq \frac{|\Omega_N|}{|\Omega_n|} \frac{|\Omega_n|}{|\Omega_N|} \left(\frac{2^d - 1}{2^d}\right)^{n-N} \\ & \leq \left(\frac{2^d - 1}{2^d}\right)^{n-N} . \end{aligned}$$

■

**Theorem 11.** For all  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  with positive measure,

$$\text{maxent } \mathbb{E} = \{P_\varphi^\dagger\} = \{P_\varphi^\infty\} = \{P_=(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \models \varphi}} \omega_N)\} .$$

where  $N = N_\varphi$ , i.e., the maximum  $n$  such that  $t_n$  appears in  $\varphi$ .

**Proof:** If  $\varphi$  is a tautology, then  $P_\varphi^\dagger = P_\varphi^\infty = P_=(\cdot \mid \Omega_n)$ .

Otherwise let  $\forall \vec{x} \bigwedge_i \bigvee_{j_i} \lambda_{ij_i}(\vec{x}, \vec{t})$  be the conjunctive normal form of  $\varphi$ . Then from Proposition 8 every  $\omega_N \in \Omega_N$  such that  $\omega_N \models \varphi$  entails at least one  $\lambda_{ij_i}^*(\vec{t})$  (i.e., literal with no variable) for each  $i$ . Hence, every extension  $\omega_n$  of  $\omega_N$  also entails  $\varphi$  and thus  $\varphi_n$ . To see this notice that such literals only mention constants and will thus be quantifier free sentences in  $\mathcal{L}_N$ , and every extension of  $\omega_N$  agrees with  $\omega_N$  on  $\mathcal{L}_N$ . Also note that if  $\omega'_N \notin [\varphi_N]_N$  then for all  $n > N$  and all its extensions  $\omega'_n$  it holds that  $\omega'_n \notin [\varphi_n]_n$ . And so we find for all  $n \geq N$  that  $P_\varphi^n(\omega'_N) = 0 = P_\varphi^n(\omega'_n)$ .

Now consider an  $N$ -state  $\omega_N \in [\varphi_N]_N$  which does not entail  $\varphi$ . By Lemma 10, the ratio of its extensions that satisfy  $\varphi_n$  decreases at least exponentially quickly in  $n$ . Since  $P^n$  equivocates on those  $n$ -states which are models of  $\varphi_n$  (Lemma 5), it follows that

$$P_\varphi^\infty(\omega_n) = \begin{cases} \frac{1}{|\{\omega_n \in \Omega_n | \omega_n \models \varphi\}|} = \frac{1}{|\varphi_N|_N} \cdot \frac{|\Omega_N|}{|\Omega_n|} & \text{if } \omega_n \models \varphi \\ 0 & \text{otherwise .} \end{cases}$$

To complete the proof we need to show that  $P_\varphi^\infty = P_\varphi^\dagger$ . To show this we show that  $P_\varphi^\infty$  defined above has greater entropy than every probability function  $P' \in \mathbb{E}$  in the sense required by the maximal entropy approach.

Use  $S$  to denote the set of  $N$ -states which entail  $\varphi$ . First notice that  $P_\varphi^\infty$  defined above has greater entropy than any probability function  $P'$  which assigns probability one jointly to the  $N$ -states in  $S$ . To see this notice that for each  $n > N$ ,  $P'$  assigns non-zero probability only to extensions of state descriptions in  $S$  but so does  $P_\varphi^\infty$  and  $P_\varphi^\infty$  does so in a completely equivocal way, dividing the probability equally between them. So for each  $n > N$ ,  $P_\varphi^\infty$  has strictly greater  $n$ -entropy than  $P'$  for all  $n$  such that  $P$  and  $P'$  disagree on  $\mathcal{L}_n$ .

Next we calculate the  $n$ -entropy of  $P_\varphi^\infty$ :

$$\begin{aligned} H_n(P_\varphi^\infty) &= - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \in S}} P_\varphi^\infty(\omega_n) \log(P_\varphi^\infty(\omega_n)) \\ &= - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \in S}} \frac{|\Omega_N|}{|S| \cdot |\Omega_n|} \log\left(\frac{|\Omega_N|}{|S| \cdot |\Omega_n|}\right) \\ &= - \log\left(\frac{|\Omega_N|}{|S| \cdot |\Omega_n|}\right) = \log(|\Omega_n|) - \log\left(\frac{|\Omega_N|}{|S|}\right) . \end{aligned}$$

where  $\omega_n|_N$  is the restriction of  $\omega_n$  to the first  $N$  constants.

If  $P'$  is a probability function in  $\mathbb{E}$  which assigns joint probability  $1 - k$  to the  $N$ -states in  $S$ , then it must assign joint probability  $k$  to the  $N$ -states not in  $S$ . To maximise  $n$ -entropy,  $P'$  equivocates on the  $n$ -states extending those  $N$ -states not in  $S$  – as much as possible. We find

$$H_n(P') = - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \in S}} P'(\omega_n) \log(P'(\omega_n)) - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \notin S \\ \omega_n \models \varphi_n}} P'(\omega_n) \log(P'(\omega_n)) \quad (5)$$

To calculate this we first notice a couple of things: first that each  $\omega_N \in \Omega_N$  has the same number of extensions to  $\mathcal{L}_n$  for  $n > N$ , which is  $\frac{|\Omega_n|}{|\Omega_N|}$ . So, the number of  $\omega_n$  in the first sum is  $|S| \cdot \frac{|\Omega_n|}{|\Omega_N|}$  and the probability mass  $1 - k$  is divided equally between them to maximise  $n$ -entropy, assigning each such  $\omega_n$  measure  $\frac{(1-k) \cdot |\Omega_N|}{|S| \cdot |\Omega_n|}$ . Second, the number of  $\omega_n$  in the second sum is  $M = |\{\omega_n \in \Omega_n | (\omega_n)|_N \notin S, \omega_n \models \varphi_n\}|$  and these are jointly assigned probability mass  $k$ , so the entropy on this set is maximal, if this probability is divided equally between

them. Then for (5) we have

$$\begin{aligned}
H_n(P') &\leq - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \in \mathcal{S}}} \frac{(1-k) \cdot |\Omega_N|}{|S| \cdot |\Omega_n|} \log\left(\frac{(1-k) \cdot |\Omega_N|}{|S| \cdot |\Omega_n|}\right) - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n|_N \notin \mathcal{S} \\ \omega_n \models \varphi_n}} \frac{k}{M} \cdot \log\left(\frac{k}{M}\right) \\
&= -(1-k) \cdot \log\left(\frac{(1-k) \cdot |\Omega_N|}{|S| \cdot |\Omega_n|}\right) - k \cdot \log\left(\frac{k}{M}\right) \\
&= (1-k) \log(|\Omega_n|) - (1-k) \cdot \log\left(\frac{(1-k) \cdot |\Omega_N|}{|S|}\right) + k \log(M) - k \log(k) .
\end{aligned}$$

By Lemma 10, for large  $n$ ,  $M \ll |\Omega_n|$ . Thus, for large  $n$

$$\begin{aligned}
&(1-k) \log(|\Omega_n|) - (1-k) \cdot \log\left(\frac{(1-k) \cdot |\Omega_N|}{|S|}\right) + k \log(M) - k \log(k) \\
&\ll \log(|\Omega_n|) - \log\left(\frac{|\Omega_N|}{|S|}\right) .
\end{aligned}$$

So,  $P_\varphi^\infty$  has greater entropy than  $P'$ . And so,  $P_\varphi^\infty$  has greater entropy than all other  $P' \in \mathbb{E}$ . Hence, we have  $P_\varphi^\dagger = P_\varphi^\infty$ .  $\blacksquare$

### 3.2 Zero measure, $P_=(\forall \vec{x} \theta(\vec{x})) = 0$

If  $\varphi = \forall \vec{x} \theta(\vec{x})$  has measure zero, then another strategy is required, as explained at the start of the section. We can, however, solve one case easily (Proposition 12). We then show why this solution strategy does not work in the general measure-zero case.

**Proposition 12.** *For all conjunctions of literals  $\theta(\vec{x}, \vec{t})$  it holds for consistent  $\varphi = \forall \vec{x} \theta(\vec{x}, \vec{t}) \in \Pi_1$  with  $P_=(\varphi) = 0$  that*

$$\text{maxent } \mathbb{E} = \{P_\varphi^\dagger\} = \{P_\varphi^\infty\} .$$

**Proof:** Let  $n \geq N$  and  $\varphi = \forall \vec{x} \bigwedge_{i=1}^I \lambda_i(\vec{x}, \vec{t})$  where all  $\lambda_i$  are literals. Note that for all  $\omega_n \in \Omega_n$  it holds that  $\omega_n \in [\varphi_n]_n$ , if and only if  $[\varphi_n]_n$  is a sub-formula of  $\omega_n$ . Hence, all  $\omega_n \in [\varphi_n]_n$  have equally many  $k+n$ -states extending them which are in  $[\varphi_{n+k}]_{n+k}$ . Since the entropy maximisers on finite languages  $\mathcal{L}_m$  assign all those states which do not satisfy  $\varphi_m$  probability zero for large enough  $m$ , all probability mass is assigned to those states in  $[\varphi_n]_n$ .

First, note that  $P_=(\cdot|\varphi_n)$  has maximum  $n$ -entropy among all probability functions with  $P(\varphi_n) = 1$ . Thus,  $P_\varphi^n = P_=(\cdot|\varphi_n)$ .

Second, observe that  $P_=(\cdot|\varphi_{n+1})$  agrees with  $P_=(\cdot|\varphi_n)$  on  $\Omega_n$  since  $\theta(\vec{x})$  is a contingent conjunction of literals. To see this notice that  $P_=(\cdot|\varphi_{n+1})$  divides the probability mass equally between the  $n+1$ -states that satisfy  $\varphi_{n+1}$  but these are all extensions of  $n$ -states that satisfy  $\varphi_n$  and all of these have an equal number of extensions to  $n+1$ -states that satisfy  $\varphi_{n+1}$ .

We hence find for all  $m \geq N$  and all  $m$ -states  $\omega_m$  that  $P_\varphi^\infty(\omega_m) = \lim_{n \rightarrow \infty} P_\varphi^n(\omega_m) = \lim_{n \rightarrow \infty} P_=(\omega_m|\varphi_n) = P_=(\omega_m|\varphi_m)$ . This shows that  $P_\varphi^\infty$  is well defined on all  $m$ -states for all  $m > N$  and it satisfies P1 and P2 since

it is a limit of probability functions. Thus, by Gaifman's Theorem it can be uniquely extended to  $\mathcal{L}$ .

Also,  $\lim_{n \rightarrow \infty} P_{=}(·|\varphi_n)$  eventually dominates all other probability functions in  $\mathbb{E}$  in entropy, by the first observation. So,  $P^\dagger(\cdot) = \lim_{n \rightarrow \infty} P_{=}(·|\varphi_n)$ . ■

As soon as there are disjunctions in  $\forall \vec{x}\theta(\vec{x}) \in \Pi_1$ , matters are more involved, because different disjuncts in  $\theta(\vec{x})$  may have different consequences:

**Example 13.** For  $\varphi = \forall xy((U_2t_1 \wedge U_1x) \vee \neg U_2y) \in \Pi_1$  and the language containing only the two unary relation symbols  $U_1, U_2$ , we have  $P_{=}(\forall xy((U_2t_1 \wedge U_1x) \vee \neg U_2y)) = 0$ .

There are two sorts of  $n$ -states which entail  $\varphi_n$ . Those which entail  $U_2t_1 \wedge \bigwedge_{i=1}^n U_1t_i$  and those which entail  $\bigwedge_{i=1}^n \neg U_2t_i$ . No state can entail both sentences. At every level  $n \geq 2$ , exactly half of all extensions satisfying  $\varphi_n$  are extensions satisfying  $\varphi_{n+1}$ . Furthermore, at every level there are twice as many  $n$ -states which entail  $\bigwedge_{i=1}^{n+1} \neg U_2t_i$  than those which entail  $U_2t_1 \wedge \bigwedge_{i=1}^{n+1} U_1t_i$ .

Hence,  $P_\varphi^n((\forall x \neg U_2x)_n) = 2P_\varphi^n((\forall x(U_2t_1 \wedge U_1x))_n)$ . The disjuncts in  $\varphi$  ( $U_2t_1 \wedge U_1x$  and  $\neg U_2y$ ) are thus treated differently in the entropy maximising process.

**Theorem 14.** If  $\chi = \{\forall \vec{x}\theta(\vec{x})\}$  where  $\forall \vec{x}\theta(\vec{x}) \in \Pi_1$ , and  $P^\infty$  exists, then

$$\text{maxent } \mathbb{E} = \{P^\infty\} = \{P^\dagger\} .$$

**Proof:** Let  $\varphi$  be  $\forall \vec{x}\theta(\vec{x})$ ,  $N = N_\varphi$  and  $t > N$ . Let's first observe that, then for all  $t \geq N$  and all  $t$ -states  $\omega_t \in \Omega_t$  using Lemma 5

$$\begin{aligned} P^\infty(\omega_t) &= \lim_{n \rightarrow \infty} P^n(\omega_t) = \lim_{n \rightarrow \infty} \sum_{\substack{\omega_n \in [\varphi_n]_n \\ \omega_n \models \omega_t}} P^n(\omega_n) \\ &= \lim_{n \rightarrow \infty} \frac{|\{\omega_n \in [(\forall \vec{x}\theta(\vec{x}))_n]_n \mid \omega_n \models \omega_t\}|}{|\varphi_n|_n} . \end{aligned}$$

Since  $P^\infty$  exists by assumption, this limit is well-defined, i.e., it takes a definite value for all  $\omega_t \in \Omega_t$ .

This, then, defines  $P^\infty$  on all  $t$ -states, for all  $t \geq N$ . But this means that  $P^\infty$  is uniquely determined. By the fact that  $P^\infty$  is a unique probability function, it must be the case that **P2** holds, in particular, for all  $t$ -states  $\omega_t \in \Omega_t$  it holds that  $P^\infty(\omega_t) = \sum_{\omega_{t+1} \models \omega_t} P^\infty(\omega_{t+1})$ . Since  $1 = \lim_{n \rightarrow \infty} P^n(\varphi_n) = P^\infty(\varphi)$  it follows that  $P^\infty \in \mathbb{E}$ .

Next, we show that  $P^\infty = P^\dagger$ . To show this we will show that  $P^\infty$  has greater entropy than every other probability function in  $\mathbb{E}$ . So let  $Q \in \mathbb{E} \setminus \{P^\infty\}$ . If  $t \geq N$ , then for every  $t$ -state  $\omega_t$  that is inconsistent with  $\varphi_t$ , it holds that  $\omega_t$  is inconsistent with  $\varphi$ . Hence,  $Q(\omega_t) = 0$ . We consider two cases: first we look at those  $Q$  that assign non-zero probability to some  $t$ -state,  $t > N$  with vanishingly few extensions that satisfy  $\varphi_n$  for large  $n$ , and next those  $Q$  that assign zero probability to all such  $t$ -states.

**Case 1.** Suppose that there is some  $t$ -state  $\omega_t$  that satisfies  $\varphi_t$  and  $Q(\omega_t) > 0$  such that there exists some other  $t$ -state  $\mu_t$  that satisfies  $\varphi_t$  with many more extensions compatible with  $\varphi$  than  $\omega_t$ . Then the limit of  $\frac{|\{\omega_n \in [\varphi_n]_n \mid \omega_n \models \omega_t\}|}{|\{\nu_n \in [\varphi_n]_n \mid \nu_n \models \mu_t\}|}$  is zero. Suppose that  $t \geq N$  is minimal with this property. Now define a

probability function  $P \in \mathbb{E}$  which agrees with  $Q$  everywhere except for  $\omega_t, \nu_t$  and (at least some of) their extensions. Let  $P(\omega_t) := 0$  and  $P(\nu_t) := Q(\nu_t) + Q(\omega_t) > Q(\nu_t)$ . Note that this forces  $P(\omega_n) = 0$  for all extensions  $\omega_n$  of  $\omega_t$ . For the extensions  $\nu_n$  of  $\nu_t$  we define a real number  $\alpha > 0$  by the unique solution of  $\alpha \cdot P^\infty(\nu_t) = Q(\omega_t) + Q(\nu_t)$ . Then simply put  $P(\nu_n) := \alpha \cdot P^\infty(\nu_n)$  for all extensions  $\nu_n$  of  $\nu_t$ . We need to show is that this is a probability function. For this is it is enough to observe that for all  $k \geq 0$  and all  $\nu_{n+k}$  that satisfy  $\varphi_{n+k}$  extending  $\nu_n$ ,

$$\begin{aligned} P(\nu_{n+k}) &= \alpha P^\infty(\nu_{n+k}) = \alpha \sum_{\substack{\nu_{n+k+1} \models \nu_{n+k} \\ \nu_{n+k+1} \in [\varphi_{n+k+1}]_{n+k+1}}} P^\infty(\nu_{n+k+1}) \\ &= \alpha \sum_{\nu_{n+k+1} \models \nu_{n+k}} P^\infty(\nu_{n+k+1}) \\ &= \sum_{\nu_{n+k+1} \models \nu_{n+k}} P(\nu_{n+k+1}) . \end{aligned}$$

Next, note that for all large enough  $n \geq t$  it holds that (Proposition 4)

$$- \sum_{\substack{\lambda_n \in \Omega_n \\ \lambda_n \models (\omega_t \vee \nu_t)}} Q(\lambda_n) \cdot \log(Q(\lambda_n)) < - \sum_{\substack{\lambda_n \in \Omega_n \\ \lambda_n \models (\omega_t \vee \nu_t)}} P(\lambda_n) \cdot \log(P(\lambda_n)) .$$

Since  $Q$  and  $P$  only disagree on  $\omega_t, \nu_t$  and (at least some of) their extensions, this means that  $H_n(Q) < H_n(P)$  for all large enough  $n$ . This entails that  $Q \notin \text{maxent } \mathbb{E}$ .

**Case 2.** Consider a  $Q \in \mathbb{E}$  which assigns zero probability to all  $\omega_t$  which have vanishingly few extensions satisfying  $\varphi_n$  for large  $n$ . Suppose furthermore that  $Q$  does not always assign probabilities according to asymptotic ratios of the number of extensions, i.e., there exists a minimal  $t \geq N$  and two  $t$ -states  $\omega_t, \nu_t \in [\varphi_t]_t$  with  $P^\infty(\omega_t), P^\infty(\nu_t) > 0$  such that

$$\frac{Q(\omega_t)}{Q(\nu_t)} > \lim_{n \rightarrow \infty} \frac{|\{\omega_n \in [\varphi_n]_n | \omega_n \models \omega_t\}|}{|\{\nu_n \in [\varphi_n]_n | \nu_n \models \nu_t\}|} = \frac{P^\infty(\omega_t)}{P^\infty(\nu_t)} .$$

Define a function  $P \in \mathbb{E}$  which agrees with  $Q$  except for  $\omega_t, \nu_t$  and (at least some of) their extensions. Put for all  $n \geq t$

$$\begin{aligned} P(\omega_n) &:= \frac{Q(\omega_t) + Q(\nu_t)}{P^\infty(\omega_t) + P^\infty(\nu_t)} P^\infty(\omega_n) \text{ for } \omega_n \text{ that extend } \omega_t \\ P(\nu_n) &:= \frac{Q(\omega_t) + Q(\nu_t)}{P^\infty(\omega_t) + P^\infty(\nu_t)} P^\infty(\nu_n) \text{ for } \nu_n \text{ that extend } \nu_t . \end{aligned}$$

So,  $P$  assigns the same joint probability mass to  $\omega_t, \nu_t$  and its extensions as  $Q$ . However,  $P$  does so by adhering to the same ratios as  $P^\infty$ . By Proposition 4 it holds for all large enough  $n \geq t$ :

$$- \sum_{\substack{\lambda_n \in \Omega_n \\ \lambda_n \models (\omega_t \vee \nu_t)}} Q(\lambda_n) \cdot \log(Q(\lambda_n)) < - \sum_{\substack{\lambda_n \in \Omega_n \\ \lambda_n \models (\omega_t \vee \nu_t)}} P(\lambda_n) \cdot \log(P(\lambda_n)) .$$

Since  $Q$  and  $P$  only disagree on  $\omega_t, \nu_t$  and (at least some of) their extensions, this means that  $H_n(Q) < H_n(P)$  for all large enough  $n$ . This entails that  $Q \notin \text{maxent } \mathbb{E}$ .

This means, that we can always improve in the entropy ordering by assigning probability zero to  $t$ -states with vanishingly few extensions compatible with  $\varphi$  (case 1). We can also improve in the entropy ordering by assigning probabilities according to the same ratios as  $P^\infty$ . There is however only one probability function that satisfies both these conditions, which is  $P^\infty$ . Hence,  $P^\infty$  has greater entropy than every other function  $Q \in \mathbb{E} \setminus \{P^\infty\}$ . And so  $P^\dagger = P^\infty$ . ■

This proof leaves two open questions. 1) What is the concrete form of  $P^\infty$ , assuming it exists for  $\varphi \in \Pi_1$ ? 2) Does the existence of a unique maximal entropy function,  $\text{maxent } \mathbb{E} = \{P^\dagger\}$ , entail that the entropy limit exists and that they are equal,  $P^\dagger = P^\infty$ , for all  $\varphi \in \Pi_1$ ?

While we do not know the answers to these questions, we do know that there are premiss sentences  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  of which Theorem 14 holds nontrivially.

**Definition 15** (Slow Premisses, [38], p. 346). A premiss sentence  $\varphi \in \mathcal{SL}$  is called *slow*, if and only if  $|\varphi_n|_n$  is polynomially bounded in  $n$ .

**Proposition 16** (Entropy Limit of Slow Premiss Sentences, [38], Theorem 3). *For all consistent and slow  $\varphi \in \Pi_1$  the entropy limit exists and assigns probability one to the premiss  $\varphi$ ,  $P^\infty \in \mathbb{E}$ .*

**Remark 17.** Theorem 14 requires that the consistent premiss  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  has a well-defined entropy limit. Slow sentences do have such a well-defined entropy limit. Hence, the entropy-limit conjecture holds non-trivially for every consistent slow premiss  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$ .

Furthermore, for all consistent premiss sentences  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  in which  $\theta(\vec{x})$  is a conjunction of literals the entropy limit is well-defined (Proposition 12) and thus the entropy-limit conjecture holds non-trivially for all these sentences, too.

## 4 Non-Categorical Premisses and Jeffrey Updating

In §3, we saw that entropy maximisation on predicate languages for categorical  $\Pi_1$  (and also  $\Sigma_1$ ) premisses amounts to updating the equivocator (the prior probability function in a state of maximal uncertainty, i.e., no evidence is available at all) by conditionalisation. This mirrors the finite case in which entropy maximisation agrees with conditionalisation in case of categorical evidence [44]. We now turn to non-categorical premisses of the form  $\varphi^X$ ,  $X \in (0, 1)$  and show that for  $\Pi_1$  and for  $\Sigma_1$  premisses entropy maximisation amounts to Jeffrey updating (Theorem 21). Again, this mirrors the finite case in which entropy maximisation agrees with Jeffrey updating in case of non-categorical premisses of the form  $\varphi^X$  [44]. Our result is also in line with the literature that shows that MaxEnt updating agrees with Jeffrey updating on infinite domains [13].

## 4.1 Point probabilities

While  $\Pi_k$  and  $\Sigma_k$  categorical constraints require different approaches to maximise entropy, this is no longer so for non-categorical constraints. Every non-categorical  $\Pi_k$  constraint is equivalent to a non-categorical  $\Sigma_k$  constraint, since  $\chi = \{\varphi^X\}$  is equivalent to  $\{\neg\varphi^{1-X}\}$ .

**Lemma 18.** *For all contingent  $\varphi \in S\mathcal{L}$  such that  $P_\varphi^\infty$  and  $P_{\neg\varphi}^\infty$  both exist and all  $X \in (0, 1)$  it holds that*

$$P_{\varphi^X}^\infty = X \cdot P_\varphi^\infty + (1 - X) \cdot P_{\neg\varphi}^\infty .$$

Furthermore, if  $P_\varphi^\infty(\varphi) = 1$  and  $P_{\neg\varphi}^\infty(\neg\varphi) = 1$ , then  $P_{\varphi^X}^\infty(\varphi) = X$ .

So, if the entropy limit exists for both categorical premisses  $\varphi, \neg\varphi$ , then the entropy limit for the non-categorical premiss(es)  $\varphi^X$  (and  $\neg\varphi^{1-X}$ ) exists and is obtained by a weighted average inspired by Jeffrey's conditioning. If  $P_{\neg\varphi}^\infty = P_{\neg\varphi}(\cdot|\neg\varphi)$  and if  $P_\varphi^\infty = P_\varphi(\cdot|\varphi)$ , then the entropy limit for the non-categorical premiss(es)  $\varphi^X$  (and  $\neg\varphi^{1-X}$ ) is indeed given by Jeffrey updating of the equivocator. In many cases where the premisses only involve one type of quantifier (either  $\forall$  or  $\exists$  but not both), this is indeed the case as we showed in §3.

**Proof:** First, note that  $\mathbb{E} = \{P \in \mathbb{P} : P(\varphi) = X \ \& \ P(\neg\varphi) = 1 - X\}$ . Next, we observe that  $Q^n := X \cdot P_\varphi^n + (1 - X) \cdot P_{\neg\varphi}^n$  satisfies all the constraints on  $\mathcal{L}_n$ ,  $Q^n(\varphi_n) = X$  and  $Q^n(\neg\varphi_n) = 1 - X$ .

We now see that for all  $n \in \mathbb{N}$  that  $X \cdot P_\varphi^n$  maximises entropy over the set

$$\{P : \Omega_n \rightarrow [0, 1] : \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi_n}} P(\omega_n) = X \quad \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \neg\varphi_n}} P(\omega_n) = 0\} ,$$

and that  $(1 - X) \cdot P_{\neg\varphi}^n$  maximises entropy over the set

$$\{P : \Omega_n \rightarrow [0, 1] : \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi_n}} P(\omega_n) = 0 \quad \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \neg\varphi_n}} P(\omega_n) = 1 - X\} .$$

We see this by recalling Proposition 4. For all functions  $f : \{1, \dots, N\} \rightarrow \mathbb{R}_{\geq 0}$  it holds that

$$\arg \sup_{f: \sum_{i=1}^N f(i)=X} - \sum_{i=1}^N f(i) \log(f(i)) = \frac{\arg \sup_{f: \sum_{i=1}^N f(i)=1} - \sum_{i=1}^N f(i) \log(f(i))}{X} .$$

Finally, observe that the objective function is additive with respect to  $n$ -states in the following sense:

$$H_n(P) = - \sum_{\omega_n \in \Omega_n} \left( \sum_{\omega_n \models \varphi_n} P(\omega_n) \log(P(\omega_n)) + \sum_{\omega_n \models \neg\varphi_n} P(\omega_n) \log(P(\omega_n)) \right) .$$

Hence, if there exist two sets of  $n$ -states (here:  $[\varphi_n]_n$  and  $[\neg\varphi_n]_n$ ) such that every constraint applies to exactly one of these sets of  $n$ -states, then the maximum entropy function can be found by maximising entropy separately over these two sets. Hence, this shows that  $P_{\varphi^X}^n = Q^n = X \cdot P_\varphi^n + (1 - X) \cdot P_{\neg\varphi}^n$ .

By the assumption that  $P_\varphi^\infty, P_{\neg\varphi}^\infty$  are well-defined, we get that  $P_{\varphi^X}^\infty := \lim_{n \rightarrow \infty} X \cdot P_\varphi^n + (1 - X) \cdot P_{\neg\varphi}^n = X \cdot P_\varphi^\infty + (1 - X) \cdot P_{\neg\varphi}^\infty$  satisfies **P1** and **P2** on  $QFSL$ . By Gaifman's Theorem [17],  $P_{\varphi^X}^\infty$  is (uniquely extendible to) a probability function on  $S\mathcal{L}$ .

Furthermore, if  $P_\varphi^\infty(\varphi) = 1$  and  $P_{\neg\varphi}^\infty(\neg\varphi) = 1$ , then it is obvious that  $P_{\varphi^X}^\infty(\varphi) = X$  and so  $P_{\varphi^X}^\infty \in \mathbb{E}$ . ■

**Lemma 19.** *Under the assumption of Lemma 18, if  $\text{maxent } \mathbb{E}_\varphi = \{P_\varphi^\dagger\} = \{P_\varphi^\infty\}$  and  $\text{maxent } \mathbb{E}_{\neg\varphi} = \{P_{\neg\varphi}^\dagger\} = \{P_{\neg\varphi}^\infty\}$ , then*

$$\text{maxent } \mathbb{E}_{\varphi^X} = \{P_{\varphi^X}^\dagger\} = \{P_{\varphi^X}^\infty\} = \{X \cdot P_\varphi^\infty + (1 - X) \cdot P_{\neg\varphi}^\infty\} .$$

**Proof:** By the above,  $P_{\varphi^X}^\infty \in \mathbb{E}_{\varphi^X}$ . Denote by  $H_{n|S}(P)$  the  $n$ -entropy of  $P$  evaluated on all  $n$ -states in  $S \subseteq \Omega_n$ . So,

$$H_{n|[\varphi_n]_n}(P) := - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi_n}} P(\omega_n) \log(P(\omega_n)) .$$

By assumption  $P_\varphi^\dagger$  exists and is unique, it must hence be in  $\mathbb{E}_\varphi$ . Since  $\text{maxent } \mathbb{E}_\varphi = \{P_\varphi^\dagger\}$ , and since  $P_\varphi^\dagger$  it has greater entropy than every other probability function  $R$  with  $R(\varphi) = 1$ , we have that  $X \cdot P_\varphi^\dagger$  will dominate any probability function  $Q$  with  $Q(\varphi) = X$  in entropy. In the same way  $(1 - X)P_{\neg\varphi}^\dagger$  will dominate any probability function  $Q$  with  $Q(\neg\varphi) = 1 - X$  in entropy. Then by Proposition 4 and the discussion immediately after that,  $X \cdot P_\varphi^\dagger + (1 - X) \cdot P_{\neg\varphi}^\dagger$  will dominate every probability function in  $\mathbb{E}_{\varphi^X}$ . By assumption, however,  $X \cdot P_\varphi^\dagger + (1 - X) \cdot P_{\neg\varphi}^\dagger = X \cdot P_\varphi^\infty + (1 - X) \cdot P_{\neg\varphi}^\infty = P_{\varphi^X}^\infty$ . ■

It is worth noting two points here. First, an application of Lemma 19 requires that  $P^\infty$  and  $P^\dagger$  are defined for both  $\varphi$  and  $\neg\varphi$ . (It is not sufficient that  $P_\varphi^\infty$  and  $P_\varphi^\dagger$  are well defined.) For example, if  $\varphi$  is a slow  $\Pi_1$  sentence, then  $\neg\varphi \in \Sigma_1$  and we know that  $P^\infty$  and  $P^\dagger$  are well-defined for both  $\varphi$  and  $\neg\varphi$ . Lemma 19 also applies non-trivially to consistent  $\varphi = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  in which  $\theta(\vec{x})$  is a conjunction of literals (Remark 17).

Second, we note that nothing in the proof of Lemma 19 hinges on working with a *single* non-categorical premiss. Indeed, all that was needed for that result was that for any satisfiable sentence  $\varphi$ , the pair  $\{\varphi_n, \neg\varphi_n\}$  give a partition of  $n$ -states for all  $n$ . The result can thus be generalised in a straightforward way to any set of premisses that satisfy this condition.

**Definition 20.** A non-empty set of sentences  $\varphi_1, \dots, \varphi_k \in S\mathcal{L}$  is called a *partition on the large*  $\mathcal{L}_n$ , if and only if there exists a  $J \in \mathbb{N}$  such that for all  $n \geq J$ ,

- the  $(\varphi_i)_n$  are satisfiable,  $(\varphi_i)_n \not\models \perp$  for all  $1 \leq i \leq k$ ,
- the  $(\varphi_i)_n$  are mutually exclusive,  $\models ((\varphi_i)_n \wedge (\varphi_j)_n) \rightarrow \perp$  for  $i \neq j$ ,
- the  $(\varphi_i)_n$  exhaust the  $n$ -universe,  $\models \bigvee_{i=1}^k (\varphi_i)_n \leftrightarrow \top$ .

Notice that a set of sentences will trivially fail to satisfy this condition if at least one sentence does not have finite models of size  $n$  for all sufficiently large  $n$ , even if it does have an infinite model. Consider for example the sentence  $\varphi = \forall xzw\exists y(U_1xy \wedge [(U_1xw \wedge U_1wz) \rightarrow U_1xz] \wedge \neg U_1xx)$  which only has infinite models. One may think of  $\{\varphi, \neg\varphi\}$  as partitioning the full language by partitioning the class of models of the language  $\mathcal{L}$ . Since  $P(\varphi_n) = 0 = 1 - P(\neg\varphi_n)$  holds for all  $n \in \mathbb{N}$  and all probability functions  $P \in \mathbb{P}$ ,  $P(\varphi_n) = X$  for all  $X \in (0, 1)$  is unsatisfiable and hence  $\mathbb{E}_n = \emptyset$  for all  $n$ . We hence require partitions on finite sublanguages.

Vice versa, not every partition on the large  $\mathcal{L}_n$  partitions the class of models of  $\mathcal{L}$ .  $\psi_1 := \varphi \vee U_2t_1$  and  $\psi_2 := \neg U_2t_1$  form a partition of all finite sublanguages but  $\psi_1 \wedge \psi_2$  has infinite models characterised by  $\varphi \wedge \neg U_2t_1$ . So,  $\{\psi_1, \psi_2\}$  does not partition the class of models of  $\mathcal{L}$ .

**Theorem 21 (Entropy Maximisation and Jeffrey Updating).** *If  $\chi = \{\varphi_1^{X_1}, \dots, \varphi_k^{X_k}\}$  where  $\varphi_1, \dots, \varphi_k$  is a partition on the large  $\mathcal{L}_n$  and  $X_1, \dots, X_k \geq 0$  such that  $\sum_{i=1}^k X_i = 1$ , and if for all  $1 \leq i \leq k$  such that  $X_i > 0$  it holds that  $\text{maxent } \mathbb{E}_{\varphi_i} = \{P_{\varphi_i}^\dagger\} = \{P_{\varphi_i}^\infty\}$ , then*

$$\text{maxent } \mathbb{E} = \{P_\chi^\dagger\} = \{P_\chi^\infty\} = \left\{ \sum_{\substack{i=1 \\ X_i > 0}}^k X_i \cdot P_{\varphi_i}^\infty \right\} .$$

**Proof:** The proof follows immediately by applying the argument in the proof of Lemma 19 a finite number of times.  $\blacksquare$

## 4.2 Generalisation to probability intervals

We now show how to use the above results to prove that the entropy-limit conjecture holds for certain non-categorical premisses  $\varphi^X$  where  $X$  is a set of probabilities.

**Proposition 22.** *Suppose  $X \subseteq [0, 1]$  is a compact interval,  $\chi = \{\forall \vec{x}\theta(\vec{x})^X\}$  where  $\varphi = \forall \vec{x}\theta(\vec{x}) \in \Pi_1$  and  $P_=(\varphi) = 0$ . If the entropy limits exists for the categorical constraint  $\chi^1 := \{\forall \vec{x}\theta(\vec{x})^1\}$ , then*

$$P_{\varphi^X}^\infty = P_{\varphi^{\inf X}}^\infty = P_{\varphi^X}^\dagger = P_{\varphi^{\inf X}}^\dagger .$$

**Proof:** First note that it follows from Lemma 18 that the entropy limits exist for the non-categorical constraint  $\chi^x := \{\forall \vec{x}\theta(\vec{x})^x\}$  for all  $x \in [0, 1]$ . We use  $P_x^n$  to denote the unique probability function on  $\mathcal{L}_n$  with maximal  $n$ -entropy subject to the constraint  $\chi^x = \{\varphi^x\}$ .

It is easy to check that for all  $x \geq \epsilon > 0$  there exists some  $M \in \mathbb{N}$  such that for all  $n \geq M$ ,  $H_n(P_x^n) < H_n(P_{x-\epsilon}^n)$ . It follows that  $P_{\varphi^x}^n = P_{\varphi^{\inf X}}^n$ . Hence,  $P_{\varphi^x}^\infty = P_{\varphi^{\inf X}}^\infty$ .

Applying Theorem 14 we note that  $P_{\varphi^1}^\dagger = P_{\varphi^1}^\infty$ . In particular,  $P_{\varphi^1}^\dagger$  exists, is unique and it satisfies the constraint  $\chi^1$ . Applying Theorem 21 we obtain  $P_{\varphi^x}^\dagger = P_{\varphi^x}^\infty$  for all  $x \in X$ .

Using the proof technique from Lemma 19 we see that for all  $y \in X \setminus \{\inf X\}$  there exists some  $M \in \mathbb{N}$  such that for all  $n \geq M$  it holds that

$H_n(P_{\varphi^y}^\dagger) < H_n(P_{\varphi^{\inf X}}^\dagger)$ . Hence,  $P_{\varphi^{\inf X}}^\dagger$  has greater entropy than every other probability function in  $\mathbb{E}$  and hence  $P_{\varphi^X}^\dagger = P_{\varphi^{\inf X}}^\dagger$ . ■

**Proposition 23.** *Suppose  $X \subseteq [0, 1]$  is a compact interval,  $\chi = \{\forall \vec{x}\theta(\vec{x})^X\}$  where  $\varphi = \forall \vec{x}\theta(\vec{x}) \in \Pi_1$ ,  $P_=(\varphi) > 0$  and  $\lambda := \arg \inf_{x \in X} |P_=(\varphi) - x|$ . It holds that*

$$P_{\varphi^X}^\infty = P_{\varphi^\lambda}^\infty = P_{\varphi^X}^\dagger = P_{\varphi^\lambda}^\dagger .$$

**Proof:** The proof is an easy adaptation of the proof of the previous proposition replacing  $\inf X$  by  $\lambda$ .

It is easy to check that for all  $x \in X \setminus \{\lambda\}$  there exists some  $M \in \mathbb{N}$  such that for all  $n \geq M$  it holds that  $H_n(P_x^n) < H_n(P_\lambda^n)$ . It follows that  $P_{\varphi^X}^n = P_{\varphi^\lambda}^n$ . Hence,  $P_{\varphi^X}^\infty = P_{\varphi^\lambda}^\infty$ .

Note first that by Theorem 11 the entropy limit exists and is equal to the maximum entropy function for the categorical constraint  $\chi^1$ ,  $P_{\varphi^1}^\dagger = P_{\varphi^1}^\infty$ . In particular,  $P_{\varphi^1}^\dagger$  exists, is unique and it satisfies the constraint  $\chi^1$ . Applying Theorem 21 we obtain  $P_{\varphi^x}^\dagger = P_{\varphi^x}^\infty$  for all  $x \in X$ .

Using the proof technique from Lemma 19 we see that for all  $y \in X \setminus \{\lambda\}$  there exists some  $M \in \mathbb{N}$  such that for all  $n \geq M$ ,  $H_n(P_{\varphi^y}^\dagger) < H_n(P_{\varphi^\lambda}^\dagger)$ . Hence,  $P_{\varphi^\lambda}^\dagger$  has greater entropy than every other probability function in  $\mathbb{E}$  and hence  $P_{\varphi^X}^\dagger = P_{\varphi^\lambda}^\dagger$ . ■

**Theorem 24.** *Suppose  $X \subseteq [0, 1]$  is a compact interval,  $\chi = \{\forall \vec{x}\theta(\vec{x})^X\}$  where  $\varphi = \forall \vec{x}\theta(\vec{x}) \in \Pi_1$ .*

- *If  $\lambda := P_=(\varphi) = 0$  and the entropy limit exists for the categorical constraint  $\chi^1 := \{\forall \vec{x}\theta(\vec{x})^1\}$ , then*

$$P_{\varphi^X}^\infty = P_{\varphi^{\inf X}}^\infty = P_{\varphi^X}^\dagger = P_{\varphi^{\inf X}}^\dagger .$$

- *If  $P_=(\varphi) > 0$  and  $\lambda := \arg \inf_{x \in X} |P_=(\varphi) - x|$ , then*

$$\begin{aligned} P_{\varphi^X}^\infty &= P_{\varphi^\lambda}^\infty = P_{\varphi^X}^\dagger = P_{\varphi^\lambda}^\dagger \\ &= P_=(\varphi) \cdot P_=(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \models \varphi}} \omega_N) + P_=(\neg\varphi) \cdot P_=(\cdot \mid \bigvee_{\substack{\omega_N \in \Omega_N \\ \omega_N \wedge \neg\varphi \text{ is consistent}}} \omega_N) , \end{aligned}$$

where  $N = N_\varphi$ .

**Proof:** The only thing left to prove is the last equality which follows from Theorems 11, 14 and 21. ■

## 5 Convergence in entropy

In this section we show that there are some rather general conditions under which the entropy-limit conjecture is true. We suppose only that

$\mathbb{E}, \mathbb{E}_1, \mathbb{E}_2, \dots$  are convex sets of probability functions generated by some consistent set  $\chi$  of constraints involving probabilities of sentences of  $\mathcal{L}$  and that the  $P^n \stackrel{\text{df}}{=} \arg \max_{P \in \mathbb{E}_n} H_n(P)$  exist for sufficiently large  $n$ . For example, if  $\chi = \{\varphi_1^{X_1}, \dots, \varphi_k^{X_k}\}$  and the  $X_1, \dots, X_k$  are probabilities or closed intervals of probabilities, then the  $\mathbb{E}_n$  are closed and this guarantees the existence of the  $P^n$  for non-empty  $\mathbb{E}_n$ .

The main condition required for the general result is that  $P^n$  converges to  $P^\infty$  in entropy. Thus, we first introduce this kind of convergence and compare it to  $L_1$  convergence, which also plays a role in what follows:

**Definition 25 (Convergence in Entropy).** Suppose  $P$  and  $Q_n$ , for  $n = 1, 2, \dots$ , are probability functions on  $\mathcal{L}$ . The  $(Q_n)_{n \geq 1}$  converge in entropy to  $P$  if  $|H_n(Q_n) - H_n(P)| \rightarrow 0$  as  $n \rightarrow \infty$ .

We define  $L_1$  distance as follows,

$$\|P - Q\|_n \stackrel{\text{df}}{=} \sum_{\omega_n \in \Omega_n} |P(\omega_n) - Q(\omega_n)| = 2 \max_{\varphi \in S\mathcal{L}_n} (P(\varphi) - Q(\varphi)) . \quad (6)$$

The latter equality follows as per Cover and Thomas [16, Equation 11.137].

**Definition 26 (Convergence in  $L_1$ ).** Suppose  $P$  and  $Q_n$ , for  $n = 1, 2, \dots$ , are probability functions on  $\mathcal{L}$ . The  $(Q_n)_{n \geq 1}$  converge in  $L_1$  to  $P$  if  $\|Q_n - P\|_n \rightarrow 0$  as  $n \rightarrow \infty$ .

The entropy function  $H_n$  is not 1-1. Therefore, that the  $Q_n$  converge to  $P$  in entropy does not imply that they converge in  $L_1$  to  $P$ , nor that, if they do additionally converge in  $L_1$  to  $P$ , then  $P$  is the unique function to which they converge in entropy.

**Example 27.** Suppose  $\mathcal{L}$  is a language with a single predicate  $U$  which is unary. Define  $P$  by  $P(Ut_1 \wedge Ut_2 \wedge \dots \wedge Ut_n) = 1$  for all  $n$  and  $R$  by  $R(\neg Ut_1 \wedge \neg Ut_2 \wedge \dots \wedge \neg Ut_n) = 1$  for all  $n$ . For  $k = 0, 1, 2, \dots$  let  $Q_{2k} = P$  and  $Q_{2k+1} = R$ . Since  $H_n(P) = H_n(R) = 0$  for all  $n$ , the  $Q_n$  converge in entropy to both  $P$  and  $R$  but converge in  $L_1$  to neither function.

**Example 28.** As in the previous example, except let  $Q_n = P$  for all  $n$ . Now the  $Q_n$  converge in  $L_1$  to  $P$ , but converge in entropy to both  $P$  and  $R$ , among other functions.

However, it turns out that, under certain conditions, if the  $n$ -entropy maximisers  $P^n$  converge in entropy to  $P \in \mathbb{E}$  then they converge in  $L_1$  to  $P$ . To show this we need two lemmas.

First, a Pythagorean theorem holds for what we call the  $n$ -divergence  $d_n$  [16, Theorem 11.6.1]:

**Definition 29 ( $n$ -divergence).** The  $n$ -divergence of two probability functions  $P$  and  $Q$  is defined as the Kullback-Leibler divergence of  $P$  from  $Q$  on  $\mathcal{L}_n$ :

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

**Lemma 30 (Pythagorean theorem).** For any convex  $\mathbb{F} \subseteq \mathbb{P}$ , if  $P \in \mathbb{F}$  and  $Q \notin \mathbb{F}$ , then

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

where  $R_n = \arg \inf_{S \in \mathbb{F}} d_n(S, Q)$ .

**Corollary 31.** For any convex  $\mathbb{F} \subseteq \mathbb{P}$ , if  $P \in \mathbb{F}$  and  $R_n = \arg \sup_{S \in \mathbb{F}} H_n(S)$ , then

$$H_n(R_n) - H_n(P) \geq d_n(P, R_n) ,$$

**Proof:** If the equivocator function  $P_{=} \notin \mathbb{F}$ , then we can apply the Pythagorean theorem to  $Q = P_{=}$  and simplify.

Otherwise,  $R_n = P_{=|_{\mathcal{L}_n}}$  and the inequality holds with equality:

$$\begin{aligned} d_n(P, P_{=}) &= \sum_{\omega \in \Omega_n} P(\omega) \log \frac{P(\omega)}{1/2^{r_n}} \\ &= \sum_{\omega \in \Omega_n} P(\omega) \log P(\omega) + \sum_{\omega \in \Omega_n} P(\omega) \log 2^{r_n} \\ &= -H_n(P) + r_n \log 2 \\ &= -H_n(P) + H_n(P_{=}) . \end{aligned}$$

■

The second lemma connects the  $L_1$  distance to  $n$ -divergence [see, e.g., 16, Lemma 11.6.1]:

**Lemma 32 (Pinsker's Inequality).**  $d_n(P, Q) \geq \frac{1}{2} \|P - Q\|_n^2$ .

Apart from convergence in entropy, the other key condition invoked by our general entropy-limit theorem is *regularity*:

**Definition 33 (Regularity).** The set  $\chi$  of constraints that generate  $\mathbb{E}, \mathbb{E}_1, \mathbb{E}_2, \dots$  is *regular* if, for sufficiently large  $n$ , the  $P^n \stackrel{\text{def}}{=} \arg \max_{Q \in \mathbb{E}_n} H_n(Q)$  dominate members of  $\mathbb{E}$  in  $n$ -entropy, in the sense that  $P^n = \arg \max_{Q \in \mathbb{F}_n} H_n(Q)$ , where  $\mathbb{F}_n \stackrel{\text{def}}{=} \langle P^n, \mathbb{E}_{|_{\mathcal{L}_n}} \rangle$ , the convex hull of  $P^n$  and  $\mathbb{E}_{|_{\mathcal{L}_n}} \stackrel{\text{def}}{=} \{P_{|_{\mathcal{L}_n}} : P \in \mathbb{E}\}$ .

**Example 34.** If  $\theta = \forall \vec{x} \theta(\vec{x}) \in \Pi_1$  and  $\chi = \{\theta\}$ , then  $\mathbb{E} = \{P \in \mathbb{P} : P(\forall \vec{x} \theta(\vec{x})) = 1\}$  and  $\mathbb{E}_n = \{P \in \mathbb{P}_n : P((\forall \vec{x} \theta(\vec{x}))_n) = 1\}$  for all  $n$ . If  $\forall \vec{x} \theta(\vec{x}) \in \Pi_1$ , then  $\chi$  is regular. This is because  $\mathbb{E}_n \supseteq \mathbb{E}_{n+1|_{\mathcal{L}_n}} \supseteq \dots \supseteq \mathbb{E}_{|_{\mathcal{L}_n}}$ , so  $\mathbb{F}_n \subseteq \mathbb{E}_n$  for all  $n$ .

**Proposition 35.** Suppose  $\chi$  is regular. If the  $P^n$  converge in entropy to  $P \in \mathbb{E}$ , then they converge in  $L_1$  to  $P$ .

**Proof:** By regularity,  $P^n = \arg \max_{Q \in \mathbb{F}_n} H_n(Q)$  for sufficiently large  $n$  and convex  $\mathbb{F}_n$ . So by Corollary 31, for sufficiently large  $n$ ,

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

by Pinsker's inequality. Hence, that the  $P^n$  converge in entropy to  $P$  implies that  $\|P - P^n\|_n^2$  converges to zero, which in turn implies that the  $P^n$  converge

in  $L_1$  to  $P$ . ■

Note that the regularity condition can be dropped if  $P$  is the equivocator function  $P_{=}$ :

**Proposition 36.** *If the  $P^n$  converge in entropy to  $P_{=}$ , then they converge in  $L_1$  to  $P_{=}$ .*

**Proof:** As we saw in the proof of Corollary 31,

$$d_n(P^n, P_{=}) = H_n(P_{=}) - H_n(P^n) .$$

So,

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

by Pinsker's inequality. Hence, that the  $P^n$  converge in entropy to  $P_{=}$  implies that  $\|P_{=} - P^n\|_n^2$  converges to zero, which in turn implies that the  $P^n$  converge in  $L_1$  to  $P_{=}$ . ■

Importantly for our purposes, convergence in entropy guarantees the existence of the pointwise entropy limit  $P^\infty$  in these cases:

**Proposition 37.** *Suppose  $\chi$  is regular or  $P$  is the equivocator function. If the  $P^n$  converge in entropy to  $P \in \mathbb{E}$ , then  $P^\infty$  exists and  $P = P^\infty$ .*

**Proof:** Applying Proposition 35 if  $\chi$  is regular, or Proposition 36 if  $P$  is the equivocator, together with Equation 6, we see that if the  $P^n$  converge in entropy to  $P$ , then  $P(\psi) = \lim_{n \rightarrow \infty} P^n(\psi)$  for every quantifier-free sentence  $\psi$ .  $P$  is the unique such limit of the  $P^n$  because it is in  $\mathbb{E}$  and so a probability function, and hence determined by its values on the quantifier-free sentences of  $\mathcal{L}$  [17].

Now,  $P^\infty$  is defined as the unique extension to  $\mathcal{L}$  of pointwise limit of  $P^n$  on quantifier-free sentences, assuming that this pointwise limit exists and satisfies the axioms of probability on quantifier-free sentences of  $\mathcal{L}$ . This latter assumption holds because, as we have seen,  $\lim_{n \rightarrow \infty} P^n(\psi) = P(\psi)$  for quantifier-free  $\psi$ , where  $P$  is a probability function. Since  $P^\infty$  is the unique extension to  $\mathcal{L}$ , it must agree with  $P$  on  $\mathcal{L}$  as a whole. Therefore  $P^\infty$  exists and  $P = P^\infty$ . ■

We can now progress to the main result of this section:

**Theorem 38 (Entropy-Limit Theorem under convergence in entropy).** *Suppose  $\chi$  is regular or  $P$  is the equivocator function. If the  $P^n$  converge in entropy to  $P \in \mathbb{E}$ , then  $P^\infty$  exists and*

$$\text{maxent } \mathbb{E} = \{P\} = \{P^\infty\} .$$

**Proof:** The existence of  $P^\infty$  and the fact that  $P = P^\infty$  is an application of Proposition 37. So it remains to show that  $\text{maxent } \mathbb{E} = \{P^\infty\}$ .

If  $P$  is the equivocator function, this fact follows straightforwardly.  $P = P^\infty = P_{=} \in \mathbb{E}$  and for any other function  $Q$  in  $\mathbb{E}$ ,  $P$  and  $Q$  must differ on some

$n$ -states for large enough  $n$ .  $P_=_$  has greater  $n$ -entropy than  $Q$  for all such  $n$ ; this holds for any other  $Q \in \mathbb{E}$ , so  $P_=_$  is the unique member of maxent  $\mathbb{E}$ .

We turn next to the case in which  $P$  is not the equivocator function—this is the case in which  $\chi$  is regular. First we shall show that  $P^\infty \in \text{maxent } \mathbb{E}$ ; later we shall see that there is no other member of maxent  $\mathbb{E}$ .

First, then, assume for contradiction that  $P^\infty \notin \text{maxent } \mathbb{E}$ . Then there is some  $Q \in \mathbb{E}$  such that  $Q$  has greater entropy than  $P^\infty$ . I.e., for sufficiently large  $n$ ,  $H_n(P^n) \geq H_n(Q) > H_n(P^\infty)$ . N.b.,  $Q \neq P^\infty$ . Hence, for sufficiently large  $n$ ,

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

where the latter two inequalities hold by Corollary 31 (given regularity) and Pinsker's inequality. Hence, since the  $P^n$  converge in entropy to  $P^\infty$ , they converge pointwise to  $Q$ . By the uniqueness of pointwise limits,  $Q = P^\infty$ : a contradiction. Hence,  $P^\infty \in \text{maxent } \mathbb{E}$ , as required.

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

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

Let  $R \stackrel{\text{def}}{=} \lambda P^\dagger + (1 - \lambda) P^\infty$  for some  $\lambda \in (0, 1)$ . Now by the log-sum inequality [16, Theorem 2.7.1], for all  $n \in J$  large enough that  $P^\dagger(\omega_n) \neq P^\infty(\omega_n)$  for some  $\omega_n \in \Omega_n$ ,

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

Hence,

$$\begin{aligned} H_n(P^n) - H_n(P^\infty) &> H_n(P^n) - H_n(R) \\ &\geq d_n(R, P^n), \end{aligned}$$

for large enough  $n \in J$ , by Corollary 31 and regularity.

Now by Pinsker's inequality (Lemma 32) and the definition of  $R$ ,

$$\begin{aligned} d_n(R, P^n) &\geq \frac{1}{2} \|R - P^n\|_n^2 \\ &= \frac{1}{2} \|P^\infty - P^n + \lambda(P^\dagger - P^\infty)\|_n^2 \\ &= \frac{1}{2} \left( \sum_{\omega_n \in \Omega_n} |P^\infty(\omega_n) - P^n(\omega_n) + \lambda(P^\dagger(\omega_n) - P^\infty(\omega_n))| \right)^2. \end{aligned}$$

Let  $f_n(\varphi) \stackrel{\text{def}}{=} P^\infty(\varphi) - P^n(\varphi) + \lambda(P^\dagger(\varphi) - P^\infty(\varphi))$  and  $\rho_n \stackrel{\text{def}}{=} \bigvee_{f_n(\omega_n) > 0} \omega_n$ . Then,

$$\sum_{\omega_n \in \Omega_n} |f_n(\omega_n)| = \sum_{\omega_n: f_n(\omega_n) > 0} f_n(\omega_n) - \sum_{\omega_n: f_n(\omega_n) \leq 0} f_n(\omega_n)$$

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

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

Let us consider the behaviour of

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

as  $n \rightarrow \infty$ . Now,  $P^\infty(\rho_n) - P^n(\rho_n) \rightarrow 0$  as  $n \rightarrow \infty$ , because  $P^n$  converges in  $L_1$  to  $P^\infty$  (Proposition 35). However,  $\lambda(P^\dagger(\rho_n) - P^\infty(\rho_n)) \not\rightarrow 0$  as  $n \rightarrow \infty$ , as we shall now see.  $P^\dagger \neq P^\infty$  by assumption, so they must differ on some quantifier-free sentence  $\psi$ , a sentence of  $\mathcal{L}_m$ , say. Suppose without loss of generality that  $P^\dagger(\psi) > P^\infty(\psi)$  (otherwise take  $\neg\psi$  instead) and let  $\delta = P^\dagger(\psi) - P^\infty(\psi) > 0$ . 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  $P^n$  converges in  $L_1$  to  $P^\infty$  we can consider  $n > m$  large enough that (see equation 6):

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

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

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

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

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

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

One can use this result to test whether some hypothesised function  $P \in \mathbb{E}$  is both the entropy limit  $P^\infty$  and maximal entropy function  $P^\dagger$ , via the following procedure:

1. Determine  $P^n$  as a function of  $n$ .
2. Determine whether  $P^n$  converges in entropy to  $P$ .

3. Determine whether  $\chi$  is regular or  $P$  is the equivocator function.
4. If these last two conditions hold, then  $P = P^\dagger = P^\infty$ .

With regard to step 2, a rapid form of convergence in  $L_1$  is sufficient (but not necessary) for convergence in entropy. Recall that  $r_n$  is the number of atomic sentences in  $\mathcal{L}_n$ :

**Lemma 39.** *If  $r_n \|Q_n - P\|_n \rightarrow 0$  as  $n \rightarrow \infty$ , then the  $Q_n$  converge in entropy to  $P$ .*

**Proof:** By Cover and Thomas [16, Theorem 17.3.3], for sufficiently large  $n$  we have that:

$$\begin{aligned} H_n(Q_n) - H_n(P) &\leq -\|Q_n - P\|_n \log \frac{\|Q_n - P\|_n}{2^{r_n}} \\ &= r_n \|Q_n - P\|_n \log 2 - \|Q_n - P\|_n \log \|Q_n - P\|_n . \end{aligned}$$

Both these latter terms tend to zero with  $n$ , by the fact that  $r_n \|Q_n - P\|_n$  tends to zero together with (in the case of the second term) the fact that  $x \log x \rightarrow 0$  as  $x \rightarrow 0$ .  $\blacksquare$

In the remainder of this section we provide a range of examples to illustrate the usage of the above algorithm.

**Example 40.** Suppose  $\chi = \{\exists x Ux\}$ , where  $\mathcal{L}$  has a single unary predicate  $U$ . Letting  $\varphi$  be  $\exists x Ux$ ,  $\varphi_n$  is defined as  $Ut_1 \vee \dots \vee Ut_n$ . We have that  $\mathbb{E} = \{P \in \mathcal{P} : P(\varphi) = 1\}$  and  $\mathbb{E}_n = \{P \in \mathbb{P}_n : P(\varphi_n) = 1\}$ . The  $n$ -entropy maximiser gives probability 0 to the  $n$ -state  $\neg Ut_1 \wedge \dots \wedge \neg Ut_n$  and divides probability 1 amongst the  $2^n - 1$  other  $n$ -states:

$$P^n(\omega_n) = \begin{cases} \frac{1}{2^n - 1} & : \omega_n \models \varphi_n \\ 0 & : \omega_n \not\models \varphi_n \end{cases} .$$

We shall use Lemma 39 to show that the  $P^n$  converge in entropy to the equivocator function  $P_=:$

$$\begin{aligned} r_n \|P^n - P_=\|_n &= n \sum_{\omega_n \in \Omega_n} |P^n(\omega_n) - P_=(\omega_n)| \\ &= n \left[ (2^n - 1) \left( \frac{1}{2^n - 1} - \frac{1}{2^n} \right) + \left( \frac{1}{2^n} - 0 \right) \right] \\ &= n \left[ 1 - \frac{2^n - 1}{2^n} + \frac{1}{2^n} \right] \\ &= \frac{2n}{2^n} \rightarrow 0 \text{ as } n \rightarrow \infty . \end{aligned}$$

Note that  $P_=\in \mathbb{E}$ :

$$P_=(\exists x Ux) = \lim_{n \rightarrow \infty} P_=(\bigvee_{i=1}^n Ut_i) = \lim_{n \rightarrow \infty} (1 - P_=(\bigwedge_{i=1}^n \neg Ut_i)) = 1 - \lim_{n \rightarrow \infty} \frac{1}{2^n} = 1 .$$

Now  $\chi$  is not regular. This is because  $\mathbb{E}|_n = \mathbb{P}_n$ , so  $\mathbb{F}_n = \langle P^n, \mathbb{P}_n \rangle$  and  $\arg \max_{Q \in \mathbb{F}_n} H_n(Q)$  is  $P_=:$  rather than  $P^n$ . However, because the  $P^n$  converge in entropy to the equivocator function, the Entropy-Limit Theorem (Theorem 38) nevertheless implies that  $\text{maxent } \mathbb{E} = \{P^\infty\} = \{P_=\}$ .

**Example 41.** Suppose  $\chi = \{\forall x Ux^X\}$ , where  $\mathcal{L}$  has a single unary predicate  $U$  and  $X \in [0, 1]$ . Letting  $\varphi$  be  $\forall x Ux$ ,  $\varphi_n$  is  $Ut_1 \wedge \dots \wedge Ut_n$ . We have that  $\mathbb{E} = \{P \in \mathbb{P} : P(\varphi) = X\}$  and  $\mathbb{E}_n = \{P \in \mathbb{P}_n : P(\varphi_n) = X\}$ . Then the  $n$ -entropy maximiser gives probability  $X$  to the  $n$ -state  $\varphi_n$  and divides probability  $1 - X$  amongst all other  $n$ -states:

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

Let us consider whether the  $P^n$  might converge in entropy to the following function:

$$P(\omega_n) = \begin{cases} X + x_n & : \omega_n = \varphi_n \\ \frac{1-X-x_n}{2^n-1} & : \omega_n \models \neg\varphi_n \end{cases}$$

where  $x_n = \frac{1-X}{2^n}$ . Now,

$$|P^n(\omega_n) - P(\omega_n)| = \begin{cases} x_n & : \omega_n = \varphi_n \\ \frac{x_n}{2^n-1} & : \omega_n \models \neg\varphi_n \end{cases}$$

and,

$$\begin{aligned} r_n \|P^n - P\|_n &= r_n \sum_{\omega_n \in \Omega_n} |P^n(\omega_n) - P(\omega_n)| \\ &= 2r_n x_n \\ &= 2(1-X) \frac{n}{2^n} \\ &\rightarrow 0 \text{ as } n \rightarrow \infty . \end{aligned}$$

Hence the  $P^n$  do indeed converge in entropy to  $P$ . Moreover,  $P \in \mathbb{P}$  because  $\sum_{\omega_n \in \Omega_n} P(\omega_n) = 1$  and for  $\omega_n = \varphi_n$ ,

$$\sum_{\omega_{n+1} \models \varphi_n} P(\omega_{n+1}) = X + \frac{1-X}{2^{n+1}} + \frac{1}{2^{n+1}-1} \left(1 - X - \frac{1-X}{2^{n+1}}\right) = X + \frac{1-X}{2^n} = P(\omega_n)$$

and for  $\omega_n \neq \varphi_n$ ,

$$\sum_{\omega_{n+1} \models \omega_n} P(\omega_{n+1}) = 2 \frac{1}{2^{n+1}-1} \left(1 - X - \frac{1-X}{2^{n+1}}\right) = \frac{1-X}{2^n} = P(\omega_n) .$$

Moreover,  $P \in \mathbb{E}$ :

$$P(\forall x Ux) = \lim_{n \rightarrow \infty} \bigwedge_{i=1}^n P(Ut_i) = \lim_{n \rightarrow \infty} X + x_n = X .$$

In addition, If  $X > 0$  then  $\chi$  is regular. To see this, observe that for any  $Q \in \mathbb{E}$ ,  $X = Q(\varphi) \leq Q(\varphi_n)$  because  $\varphi \models \varphi_n$ . Therefore, when  $n$  is large enough that  $X > \frac{1}{2^n}$  we have that  $P_{=}(\varphi) = \frac{1}{2^n} < X = P^n(\varphi_n) \leq Q(\varphi_n)$  and so, since  $P^n$  spreads the remaining probability  $1 - X$  evenly across the remaining  $n$ -states,  $H_n(P^n) \geq H_n(Q)$ . On the other hand, if  $X = 0$  then  $P$  is the equivocator function. Either way, we can apply Theorem 38 to conclude that  $P = P^\infty = P^\dagger$  in this example.

**Example 42.** Suppose  $\chi = \{\forall x(Ux \rightarrow Vx)^X\} = \{\forall x((Ux \wedge Vx) \vee (\neg Ux \wedge Vx) \vee (\neg Ux \wedge \neg Vx))^X\}$ , where  $\mathcal{L}$  has a two unary predicates  $U$  and  $V$  and  $X \in [0, 1]$ . Let  $\varphi$  be  $\forall x(Ux \rightarrow Vx)$ . In this example we can proceed analogously to Example 41. Here  $\mathbb{E} = \{P \in \mathbb{P} : P(\varphi) = X\}$  and  $\mathbb{E}_n = \{P \in \mathbb{P}_n : P(\varphi_n) = X\}$ . The  $n$ -entropy maximiser shares probability  $X$  equally amongst the  $n$ -states compatible with  $\varphi_n$  and divides probability  $1 - X$  amongst all other  $n$ -states:

$$P^n(\omega_n) = \begin{cases} \frac{X}{|\varphi_n|_n} & : \omega_n \models \varphi_n \\ \frac{1-X}{|\neg\varphi_n|_n} & : \omega_n \models \neg\varphi_n \end{cases} .$$

In this case we have:

$$P(\omega_n) = \begin{cases} \frac{X+x_n}{|\varphi_n|_n} & : \omega_n \models \varphi_n \\ \frac{1-X-x_n}{|\neg\varphi_n|_n} & : \omega_n \models \neg\varphi_n \end{cases} .$$

where  $x_n = (1 - X) \frac{|\varphi_n|_n}{2^{r_n}}$ . (Again, one can check that  $P \in \mathbb{E}$ .) Now,

$$|P^n(\omega_n) - P(\omega_n)| = \begin{cases} \frac{|x_n|}{|\varphi_n|_n} & : \omega_n \models \varphi_n \\ \frac{|-x_n|}{|\neg\varphi_n|_n} & : \omega_n \models \neg\varphi_n \end{cases} .$$

and,

$$\begin{aligned} r_n \|P^n - P\|_n &= r_n \sum_{\omega_n \in \Omega_n} |P^n(\omega_n) - P(\omega_n)| \\ &= 2r_n x_n \\ &= 2r_n (1 - X) \frac{|\varphi_n|_n}{2^{r_n}} . \end{aligned}$$

This tends to zero as  $n \rightarrow \infty$  if either  $X = 1$  or  $\frac{r_n |\varphi_n|_n}{2^{r_n}} \rightarrow 0$  as  $n \rightarrow \infty$ . In the previous example, where  $\varphi$  was  $\forall x Ux$ ,  $r_n = n$ ,  $|\varphi_n|_n = 1$ , and  $\frac{r_n |\varphi_n|_n}{2^{r_n}} = \frac{n}{2^n} \rightarrow 0$  as  $n \rightarrow \infty$ . In the present example,  $r_n = 2n$ ,  $|\varphi_n|_n = 3^n$  and  $\frac{r_n |\varphi_n|_n}{2^{r_n}} = \frac{2n3^n}{4^n} \rightarrow 0$  as  $n \rightarrow \infty$ . In addition,  $\chi$  is regular if  $X > 0$ , as in the previous example, and if  $x = 0$  then  $P$  is the equivocator. So in this example too the entropy limit coincides with the maximal entropy function,  $P = P^\infty = P^\dagger$ .

**Example 43.** Suppose  $\chi = \{\forall x(Ux \rightarrow Ut_3)^X\}$ , where  $\mathcal{L}$  is a unary language with a single unary predicate  $U$  and  $X \in [0, 1]$ . Let  $\varphi$  be  $\forall x(Ux \rightarrow Ut_3)$ . Note that  $\varphi \equiv Ut_3 \vee \forall x \neg Ux$ . Now,

$$[\varphi_n] = \{\pm Ut_1 \wedge \pm Ut_2 \wedge Ut_3 \wedge \pm Ut_4 \wedge \cdots \wedge \pm Ut_n, \neg Ut_1 \wedge \neg Ut_2 \wedge \neg Ut_3 \wedge \cdots \wedge \neg Ut_n\}$$

and

$$|\varphi_n|_n = 2^{n-1} + 1 .$$

The  $n$ -entropy maximiser will give these  $n$ -states the same probability:

$$P^n(\omega_n) = \begin{cases} \frac{X}{2^{n-1}+1} & : \omega_n \models \varphi_n \\ \frac{1-X}{2^{n-1}-1} & : \omega_n \not\models \varphi_n \end{cases} .$$

Let us consider

$$P(\omega_n) = \begin{cases} \frac{X}{2^{n-1}} & : \omega_n \models Ut_3 \\ \frac{1-X}{2^{n-1}} & : \omega_n \not\models Ut_3 \end{cases} .$$

and ask whether the  $P^n$  converge to  $P \in \mathbb{E}$  in entropy:

$$\begin{aligned}
r_n \|P^n - P\|_n &= n \sum_{\omega_n \in \Omega_n} |P^n(\omega_n) - P(\omega_n)| \\
&= n \left( \left| \frac{X}{2^{n-1}+1} - \frac{1-X}{2^{n-1}} \right| + \sum_{\omega \models Ut_3} \left| \frac{X}{2^{n-1}+1} - \frac{X}{2^{n-1}} \right| + \sum_{\omega \models \neg \varphi_n} \left| \frac{1-X}{2^{n-1}-1} - \frac{1-X}{2^{n-1}} \right| \right) \\
&= \frac{n}{2^{n-1}+1} \left| 2X - 1 - \frac{1-X}{2^{n-1}} \right| + n2^{n-1} \left| \frac{-X}{2^{n-1}(2^{n-1}+1)} \right| + n(2^{n-1}-1) \left| \frac{1-X}{2^{n-1}(2^{n-1}-1)} \right| \\
&= \frac{n}{2^{n-1}+1} \left| 2X - 1 - \frac{1-X}{2^{n-1}} \right| + \frac{nX}{2^{n-1}+1} + \frac{n(1-X)}{2^{n-1}} \\
&\longrightarrow 0 \text{ as } n \longrightarrow \infty .
\end{aligned}$$

because each of the three terms tends to zero with  $n$ .  $\chi$  is regular as long as  $X > 1/2$ , for otherwise  $P$  has greater  $n$ -entropy than  $P^n$  for sufficiently large  $n$ . Hence, we can invoke Theorem 38 to conclude that  $P = P^\infty = P^\dagger$  when  $X > 1/2$ .

In cases where the condition of Lemma 39 does not hold, the following lemma can come in useful:

**Lemma 44.** *Probability functions  $(Q_n)_{n \geq 1}$  converge in entropy to  $P$  if and only if*

$$\begin{aligned}
&\sum_{\omega \in \Omega_n: P(\omega) > 0} (P(\omega) - Q_n(\omega)) \left[ 1 + \log P(\omega) - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{Q_n(\omega)}{P(\omega)} \right)^j \right] \\
&\quad - \sum_{\omega \in \Omega_n: P(\omega) = 0} Q_n(\omega) \log Q_n(\omega) \\
&\quad \longrightarrow 0 \text{ as } n \longrightarrow \infty .
\end{aligned}$$

Note that if  $Q_n(\omega)$  is zero whenever  $P(\omega)$  is zero then the second term,  $-\sum_{\omega \in \Omega_n: P(\omega) = 0} Q_n(\omega) \log Q_n(\omega)$ , vanishes.

**Proof:** Let  $x = x_\omega = Q_n(\omega)$  and  $a = a_\omega = P(\omega) > 0$  and consider the Taylor series expansion of  $x \log x$  at  $a$ :

$$\begin{aligned}
x \log x &= a \log a + (1 + \log a)(x - a) + \frac{(x - a)^2}{2!a} - \frac{(x - a)^3}{3!a^2} + \frac{2(x - a)^4}{4!a^3} - \dots \\
&= a \log a + (1 + \log a)(x - a) + \sum_{j=2}^{\infty} \frac{(-1)^j (j-2)! (x - a)^j}{j! a^{j-1}} \\
&= a \log a + (1 + \log a)(x - a) + \sum_{j=2}^{\infty} \frac{(a - x)^j}{j(j-1)a^{j-1}} \\
&= a \log a + (1 + \log a)(x - a) + (a - x) \sum_{j=1}^{\infty} \frac{(a - x)^j}{j(j+1)a^j} \\
&= a \log a + (x - a) \left[ 1 + \log a - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{x}{a} \right)^j \right] .
\end{aligned}$$

So,

$$\begin{aligned}
H_n(Q_n) - H_n(P) &= \sum_{\omega \in \Omega_n} -x_\omega \log x_\omega + a_\omega \log a_\omega \\
&= \sum_{\omega \in \Omega_n: a_\omega > 0} (a_\omega - x_\omega) \left[ 1 + \log a_\omega - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{x_\omega}{a_\omega} \right)^j \right] \\
&\quad - \sum_{\omega \in \Omega_n: a_\omega = 0} x_\omega \log x_\omega .
\end{aligned}$$

The left-hand side converges to zero just when the right-hand side converges to zero.  $\blacksquare$

**Example 45.** Let us return to the case of  $\chi = \{\forall x(Ux \rightarrow Ut_3)^X\}$ . Recall from Example 43 that:

$$P^n(\omega) = \begin{cases} \frac{X}{2^{n-1}+1} & : \omega \models \varphi_n \\ \frac{1-X}{2^{n-1}-1} & : \omega \not\models \varphi_n \end{cases}$$

and

$$P(\omega) = \begin{cases} \frac{X}{2^{n-1}} & : \omega \models Ut_3 \\ \frac{1-X}{2^{n-1}} & : \omega \not\models Ut_3 \end{cases}$$

We shall apply Lemma 44 to see that the  $P^n$  converge in entropy to  $P$ . Let  $\Delta_n \stackrel{\text{df}}{=} H_n(P^n) - H_n(P)$ . If  $X = 1$  then  $P^n$  and  $P$  are not zero on the same  $n$ -states and the second term in Lemma 44 does not vanish. Suppose first, then, that  $X < 1$ :

$$\begin{aligned}
\Delta_n &= \sum_{\omega \in \Omega_n} (P(\omega) - P^n(\omega)) \left[ 1 + \log P(\omega) - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{P^n(\omega)}{P(\omega)} \right)^j \right] \\
&= \sum_{\omega \models Ut_3} \left( \frac{X}{2^{n-1}} - \frac{X}{2^{n-1}+1} \right) \left[ 1 + \log \frac{X}{2^{n-1}} - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{X}{2^{n-1}+1} \frac{2^{n-1}}{X} \right)^j \right] \\
&\quad + \sum_{\omega \models \bigwedge_{i=1}^n \neg Ut_i} \left( \frac{1-X}{2^{n-1}} - \frac{X}{2^{n-1}+1} \right) \left[ 1 + \log \frac{1-X}{2^{n-1}} - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{X}{2^{n-1}+1} \frac{2^{n-1}}{1-X} \right)^j \right] \\
&\quad + \sum_{\omega \not\models \varphi_n} \left( \frac{1-X}{2^{n-1}} - \frac{1-X}{2^{n-1}-1} \right) \left[ 1 + \log \frac{1-X}{2^{n-1}} - \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \left( 1 - \frac{1-X}{2^{n-1}-1} \frac{2^{n-1}}{1-X} \right)^j \right] \\
&= 2^{n-1} \left( \frac{X}{2^{n-1}(2^{n-1}+1)} \right) \left( 1 + \log \frac{X}{2^{n-1}} \right) - 2^{n-1} \sum_{j=2}^{\infty} \frac{X}{2^{n-1}(2^{n-1}+1)^j j(j-1)} \\
&\quad + \left( \frac{1-2X + \frac{1-X}{2^{n-1}}}{2^{n-1}+1} \right) \left( 1 + \log \frac{1-X}{2^{n-1}} \right) - \sum_{j=2}^{\infty} \frac{\left( \frac{1-2X}{1-X} + \frac{1}{2^{n-1}} \right)^j (1-X)}{2^{n-1} \left( 1 + \frac{1}{2^{n-1}} \right)^j j(j-1)} \\
&\quad + (2^{n-1}-1) \left( \frac{X-1}{2^{n-1}(2^{n-1}-1)} \right) \left( 1 + \log \frac{1-X}{2^{n-1}} \right) - (2^{n-1}-1) \sum_{j=2}^{\infty} \frac{(-1)^j (1-X)}{2^{n-1}(2^{n-1}-1)^j j(j-1)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{X(1 + \log X - (n-1)\log 2)}{2^{n-1} + 1} - \sum_{j=2}^{\infty} \frac{X}{(2^{n-1} + 1)^j j(j-1)} \\
&\quad + \frac{(1 + \log(1-X) - (n-1)\log 2)(1 - 2X + \frac{1-X}{2^{n-1}})}{2^{n-1} + 1} - \sum_{j=2}^{\infty} \frac{\left(\frac{1-2X}{1-X} + \frac{1}{2^{n-1}}\right)^j (1-X)}{2^{n-1} \left(1 + \frac{1}{2^{n-1}}\right)^j j(j-1)} \\
&\quad + \frac{(X-1)(1 + \log(1-X) - (n-1)\log 2)}{2^{n-1}} - \sum_{j=2}^{\infty} \frac{(-1)^j (1-X)}{2^{n-1} (2^{n-1} - 1)^{j-1} j(j-1)} \\
&\rightarrow 0 \text{ as } n \rightarrow \infty,
\end{aligned}$$

because each of the component terms individually tend to zero with  $n$ . Hence we do have convergence in entropy for  $X < 1$ . If  $X = 1$ , the third line in the above sum disappears and we have to rewrite the second line, which corresponds to the  $n$ -states  $\bigwedge_{i=1}^n \neg Ut_i$  on which  $P^n$  is positive but  $P$  is zero:

$$\begin{aligned}
\Delta_n &= \frac{1 - (n-1)\log 2}{2^{n-1} + 1} - \sum_{j=2}^{\infty} \frac{1}{(2^{n-1} + 1)^j j(j-1)} \\
&\quad + \frac{1}{2^{n-1} + 1} \log \frac{1}{2^{n-1} + 1} .
\end{aligned}$$

The first line of the sum tends to zero as before; the second line is approximately  $\frac{-(n-1)\log 2}{2^{n-1}}$  which also tends to zero. Hence, we also have convergence in entropy when  $X = 1$ . Recall that  $\chi$  is regular as long as  $X > 1/2$ . Hence, we can again invoke Theorem 38 to conclude that  $P$  coincides with the maximal entropy function  $P^\dagger$  when  $X > 1/2$ .

Finally, here is an example involving a  $\Pi_2$  constraint, which shows that the entropy-limit conjecture holds in cases other than those covered by previous sections of the paper.

**Example 46.**  $\chi = \{\forall x \exists y Uxy\}$ . Let  $\varphi$  be  $\forall x \exists y Uxy$ . Here  $r_n = n^2$ ,  $\varphi_n = \bigwedge_{i=1}^n \bigvee_{j=1}^n Ut_i t_j$ ,  $|\varphi_n|_n = (2^n - 1)^n$  and for all  $\omega \in \Omega_n$

$$P^n(\omega) = \begin{cases} \frac{1}{|\varphi_n|_n} & : \omega \models \varphi_n \\ 0 & : \omega \not\models \varphi_n \end{cases} .$$

Now  $P^n$  converges in entropy to the equivocator  $P_-$ :

$$\begin{aligned}
H_n(P_-) - H_n(P^n) &= - \sum_{\omega \in \Omega_n} \frac{1}{2^{n^2}} \log \frac{1}{2^{n^2}} + \sum_{\omega \models \varphi_n} \frac{1}{(2^n - 1)^n} \log \frac{1}{(2^n - 1)^n} \\
&= 2^{n^2} \frac{1}{2^{n^2}} \log 2^{n^2} - (2^n - 1)^n \frac{1}{(2^n - 1)^n} \log (2^n - 1)^n \\
&= -n \log \frac{2^n - 1}{2^n} \\
&= -\log \left(1 - \frac{1}{2^n}\right)^n .
\end{aligned}$$

So

$$\lim_{n \rightarrow \infty} H_n(P_-) - H_n(P^n) = \lim_{n \rightarrow \infty} -\log \left(1 - \frac{1}{2^n}\right)^n = -\log \lim_{n \rightarrow \infty} \left(1 - \frac{1}{2^n}\right)^n = -\log 1 = 0 .$$

Moreover, the equivocator is in  $\mathbb{E}$ :

$$\begin{aligned}
P_{=}(\forall x \exists y Uxy) &= \lim_{n \rightarrow \infty} P_{=}(\bigwedge_{i=1}^n \exists y Ut_i y) \\
&\geq \lim_{n \rightarrow \infty} (1 - \sum_{i=1}^n P_{=}(\neg \exists y Ut_i y)) \\
&= \lim_{n \rightarrow \infty} (1 - \sum_{i=1}^n P_{=}(\forall y \neg Ut_i y)) \\
&= \lim_{n \rightarrow \infty} (1 - \sum_{i=1}^n 0) \\
&= 1 .
\end{aligned}$$

Theorem 38 then implies that  $\text{maxent } \mathbb{E} = \{P_{=}\} = \{P^{\infty}\}$ .

## 6 Conclusions

We have shown that the entropy-limit conjecture holds in the following scenarios:

**Categorical  $\Pi_1$ .**  $\chi = \{\forall \vec{x} \theta(\vec{x})\}$  and  $\forall \vec{x} \theta(\vec{x}) \in \Pi_1$  (Theorem 11 and Theorem 14).

**Non-categorical partition.**  $\chi = \{\varphi_1^{X_1}, \dots, \varphi_k^{X_k}\}$  where  $\varphi_1, \dots, \varphi_k$  is a partition on the large  $\mathcal{L}_n$  and  $X_1, \dots, X_k \geq 0$  such that  $\sum_{i=1}^k X_i = 1$ , and for all  $1 \leq i \leq k$  such that  $X_i > 0$  it holds that  $\text{maxent } \mathbb{E}_{\varphi_i} = \{P_{\varphi_i}^{\dagger}\} = \{P_{\varphi_i}^{\infty}\}$  (Theorem 21).

**Convergence in Entropy.** The  $P^n$  converge in entropy to  $P \in \mathbb{E}$  and either  $\chi$  is regular or  $P$  is the equivocator function (Theorem 38).

Taking into account previous work (see §2), the entropy-limit conjecture has now been verified in quite a broad range of different scenarios. Future work might proceed in one of two directions. The first is to further extend the range of scenarios in which the conjecture is tested—e.g., to categorical constraints of greater quantifier complexity or to a broader range of non-categorical constraints. The second is to consider inference processes other than the maximum entropy principle, which might be relevant to questions other than the search for a canonical inductive logic or a canonical characterisation of normal models. There are several examples of such inference processes that have been proposed and studied in the literature, for example Centre of Mass, Minimum Distance and the spectrum of inference processes based on generalised Rényi entropies [43]. These inference processes differ in the structural properties that they impose on the probability function that they pick for inference. There are, however, several such properties that are in common between them that allow for a generalisation of some of our results—see [36] for a detailed analysis of these structural properties for different inference processes. Of particular interest is a symmetry property called the Renaming Principle.

The Renaming Principle (RP) is a symmetry axiom that ensures that the choice of the probability function is invariant under a uniform renaming of the set of state descriptions of finite sublanguages.

An inference process  $\iota$ , defined on the finite languages  $\mathcal{L}_n$ , satisfies *Renaming Principle* if for two sets of linear constraints  $\chi$  and  $\chi'$  of the form

$$\chi = \left\{ \sum_{j=1}^{r_n} a_{ji} P(\omega_j) = b_i \mid i = 1, \dots, m \right\}$$

$$\chi' = \left\{ \sum_{j=1}^{r_n} a_{ji} P(\omega'_j) = b_i \mid i = 1, \dots, m \right\} ,$$

where the  $\omega'_1, \dots, \omega'_{r_n}$  are a permutation of the  $n$ -states  $\omega_1, \dots, \omega_{r_n}$  of  $\mathcal{L}_n$ , it holds that:

$$\iota_\chi(\omega_j) = \iota_{\chi'}(\omega'_j).$$

What is special about RP in our context is that almost all the results we have provided (as well as those given in [38, 39, 40]) hold for any inference process that satisfies RP. This is a rather large class of inference processes that includes not only the Maximum Entropy but also the examples given above (Center of Mass, Minimum Distance and those based on generalised Rényi entropies). For a detailed discussion on this point see Rafiee Rad [42].

We give another symmetry result that follows from RP in the Appendix. An immediate question, which we hope to study further in future work, is whether or not the conjecture and the results thereof can be generalised if we take an approach analogous to the maximal-entropy approach for defining these other inference processes on first order languages.

Another promising avenue for further research is the introduction of functions to the underlying language, as recently studied by Howarth and Paris [23].

## A Appendices

### A.1 Defining the entropy limit

As pointed out in the Section 2 there are two ways to define the entropy-limit approach on first order languages. One, the Barnett-Paris definition, is to define the entropy-limit function as the limit of local entropy maximisers directly on all sentences of  $\mathcal{L}$ . That is to take

$$P^\infty(\psi) = \lim_{r \rightarrow \infty} P^r(\psi^r)$$

if the limit exists for all for all  $\psi \in S\mathcal{L}$ , and to take  $P^\infty$  as undefined otherwise. The second approach, the Rad-Paris definition, is to define  $P^\infty$  on quantifier free sentences as the limit of local entropy maximisers and then take its unique extension (by Gaifman's Theorem) to the whole of  $S\mathcal{L}$ .

If the pointwise limit given by the first approach exists and is a probability function, then it agrees with the one obtained from the second approach. To see this let the probability function  $W$  be given by the pointwise limit and let  $P^\infty$  be the one obtained from the Rad-Paris definition. Then for all  $n$  and  $n$ -states  $\omega_n$

$$W(\omega_n) = \lim_{r \rightarrow \infty} P^r((\omega_n)_r) = \lim_{r \rightarrow \infty} P^r(\omega_n) = P^\infty(\omega_n) .$$

Thus,  $W$  agrees with  $P^\infty$  on all  $n$ -states and so on all quantifier free sentences and hence, by the uniqueness criteria in Gaifman's Theorem they agree on all  $S\mathcal{L}$ .

The main issue with the Barnett-Paris approach is that the pointwise limit on the whole of  $S\mathcal{L}$  might exist but not be a probability function. This is obviously circumvented by the second approach: defining  $P^\infty$  on quantifier free sentences as the above limit ensures that axioms P1 and P2 are satisfied, and Gaifman's Theorem guarantees a unique extension of  $P^\infty$  to be a probability function over all  $S\mathcal{L}$ . To see how the first approach can fail in this respect, consider the following example. Let  $\mathcal{L}$  be language with equality and a single binary relation  $U$ , and consider the following set of sentences:  $\varphi_1 = \forall x \neg U(x, x)$ ,  $\varphi_2 = \forall x, y, z ((U(x, y) \wedge U(y, z)) \rightarrow U(x, z))$ ,  $\varphi_3 = \forall x, y (\neg(x = y) \rightarrow (U(x, y) \vee U(y, x)))$  and  $\varphi_4 = \forall x \exists y U(x, y)$ . Note that  $\varphi_1, \varphi_2$  and  $\varphi_3$  are the axioms for a linear strict order and adding  $\varphi_4$  ensures that there are no end points. As noted above, these sentences together have no finite model. Let

$$\varphi = \varphi_1 \wedge \varphi_2 \wedge \varphi_3.$$

First consider  $P^\infty$  defined by Rad-Paris definition. For each  $n$  and  $n$ -state  $\omega_n$

$$P_\varphi^\infty(\omega_n) = \lim_{r \rightarrow \infty} P_\varphi^r(\omega_n) = \lim_{r \rightarrow \infty} \sum_{\omega_r \models \omega_n} P_\varphi^r(\omega_r) .$$

As we have observed already,  $P^r$  assigns the full probability mass equally among those  $r$ -states that are consistent with  $\varphi$ , i.e., those  $r$ -states that characterise a strict linear order over  $t_1, \dots, t_r$ . There are  $r!$  many such  $r$ -states. If  $\omega_n$  is inconsistent with  $\varphi$ , then all its  $r$ -state extensions are inconsistent with it and we have  $P_\varphi^r(\omega_n) = 0$  for all  $r > n$ . If, on the other hand,  $\omega_n$  does characterise a strict linear order over  $t_1, \dots, t_n$ , then it can be extended to a strict linear order over  $t_1, \dots, t_r$  in  $\prod_{i=1}^{r-n} (n+i) = \frac{r!}{n!}$  many ways each receiving the same probability of  $1/r!$  under  $P_\varphi^r$ , thus

$$P_\varphi^\infty(\omega_n) = \lim_{r \rightarrow \infty} \sum_{\omega_r \models \omega_n} P_\varphi^r(\omega_r) = \begin{cases} \frac{(r!/n!)}{r!} = 1/n! & \text{if } \omega_n \text{ is consistent with } \varphi \\ 0 & \text{otherwise .} \end{cases}$$

Then  $P_\varphi^\infty(\varphi_4) = 1$ . To see this first notice that by P3,

$$P_\varphi^\infty(\varphi_4) = \lim_{m \rightarrow \infty} P_\varphi^\infty\left(\bigwedge_{i=1}^m \exists y U(t_i, y)\right) \quad (7)$$

Second notice that by our result for the  $\Sigma_1$  sentences,

$$P_\varphi^\infty(\exists y U(t_i, y)) = \lim_{n \rightarrow \infty} P_\varphi^\infty\left(\bigvee_{k=1}^n U(t_i, t_k)\right) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} P_\varphi^r\left(\bigvee_{k=1}^n U(t_i, t_k)\right).$$

Let  $n > i$ . Then since  $\bigvee_{k=1}^n U(t_i, t_k) \in S\mathcal{L}_n$  there are  $n$ -states  $\omega_1, \dots, \omega_s$  such that  $\models \bigvee_{k=1}^n U(t_i, t_k) \leftrightarrow \bigvee_{i=1}^s \omega_i$ . Thus

$$P_\varphi^\infty(\exists y U(t_i, y)) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} P_\varphi^r\left(\bigvee_{k=1}^n U(t_i, t_k)\right) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \sum_{i=1}^s P_\varphi^r(\omega_i) .$$

The local entropy maximiser  $P_\varphi^r$  assigns probability zero to those  $n$ -states that do not correspond to a strict linear ordering of  $t_1, \dots, t_n$  and assigns probability of  $1/n!$  to each of the rest. Of these, only one does not appear among  $\{\omega_1, \dots, \omega_s\}$ , namely the one which puts  $t_i$  as the final element in the ranking. Hence,

$$P_\varphi^\infty(\exists y U(t_i, y)) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \sum_{i=1}^s P_\varphi^r(\omega_i) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \frac{n! - 1}{n!} = 1 .$$

Then  $P_\varphi^\infty$  assigns probability 1 to each conjunct in (7) and hence will give probability 1 to the whole conjunction,  $P_\varphi^\infty(\bigwedge_{i=1}^m \exists y U(t_i, y)) = 1$ , and we have

$$P_\varphi^\infty(\varphi_4) = \lim_{m \rightarrow \infty} P_\varphi^\infty\left(\bigwedge_{i=1}^m \exists y U(t_i, y)\right) = \lim_{m \rightarrow \infty} 1 = 1 .$$

Now for all  $\psi \in \mathcal{SL}$  let  $W(\psi) = \lim_{r \rightarrow \infty} P_\varphi^r(\psi^r)$  as given by the Barnett-Paris definition and assume the limit is well defined for all  $\psi$  and that  $W$  is a probability function on  $\mathcal{SL}$ . Then by the discussion above  $W$  agrees with  $P_\varphi^\infty$  on all  $\mathcal{SL}$  but then

$$1 = P_\varphi^\infty(\varphi_4) = W(\varphi_4) = \lim_{r \rightarrow \infty} P_\varphi^r((\varphi_4)_r) = \lim_{r \rightarrow \infty} 0 = 0,$$

a contradiction. Notice that the penultimate equality follows from the fact that  $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$  has no finite models. Thus, if  $W$  is well defined on all  $\mathcal{SL}$ , it cannot be a probability function.

We have shown two things: (i) The Barnett-Paris approach might fail to produce a probability function (violating P3); (ii) The Rad-Paris approach does produce a probability function in all cases in which the Barnett-Paris entropy-limit is well-defined on  $QF\mathcal{SL}$ . Note finally that the Rad-Paris entropy-limit function may fail to satisfy the constraints  $\chi$ . The maximal entropy function, if unique, always satisfies the constraints since it is by definition a member of  $\mathbb{E}$ , the set of probability functions satisfying  $\chi$ .

## A.2 Symmetries

We show here that symmetries are deeply engrained in the entropy-limit approach.

**Definition 47** (Symmetrical Treatment, adapted from [46], p. 160). For  $\varphi_1, \dots, \varphi_k \in \mathcal{SL}$  and  $X_1, \dots, X_k \subseteq [0, 1]^k$ , we say that for  $n \geq N$  two  $n$ -states  $\omega_n, \omega'_n \in \Omega_n$  are *treated symmetrically by the premisses*  $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ , if and only if for all  $m \geq n$  there exists a bijection  $f_m : \{\omega_m \in \Omega_m : \omega_m \models \omega_n\} \rightarrow \{\omega'_m \in \Omega_m : \omega'_m \models \omega'_n\}$  such that for all  $1 \leq i \leq k$  it holds that  $\omega_m \in [(\varphi_i)_m]_m \iff f_m(\omega_m) \in [(\varphi_i)_m]_m$ <sup>4</sup>

<sup>4</sup>It is worth noting that the notion of symmetrical treatment by the evidence used here is an strengthening of the one introduced by Williamson in [46]. The notion as introduced by Williamson requires the existence of a bijection  $f$  from the extensions of  $\omega_n$  to extensions of  $\omega'_n$  but the one used here requires a (possibly different) bijection  $f_m$  between these extensions at every  $\mathcal{L}_m$  for  $m > n$ .

**Proposition 48.** *If two  $n$ -states  $\omega_n, \omega'_n \in \Omega$  are treated symmetrically by the premisses and if  $P^\infty$  exists, then  $P^\infty(\omega_n) = P^\infty(\omega'_n)$*

**Proof:** Suppose not and let  $\omega_n, \omega'_n \in \Omega_n$  be treated symmetrically by the premisses such that  $P^\infty(\omega_n) \neq P^\infty(\omega'_n)$ . We now derive a contradiction.

Since  $P^\infty$  is a limit of probability functions  $P^n$  which, eventually, all assign a definite probability to  $\omega_n$  and  $\omega'_n$ , there has to exist some  $m \geq n$  such that, say,  $P^m(\omega_n) > P^m(\omega'_n)$  and so  $\sum_{\omega_m \models \omega_n} P^m(\omega_m) > \sum_{\omega'_m \models \omega'_n} P^m(\omega'_m)$ .

Let us now define a probability function  $Q_m$  on  $S\mathcal{L}_m$  which agrees with  $P^m$  on all  $m$ -states which are neither an extension of  $\omega_n$  nor of  $\omega'_n$ . For those extensions we let

$$\begin{aligned} Q_m(\omega_m) &:= P^m(f_m(\omega_m)) \\ Q_m(\omega'_m) &:= P^m(f_m^{-1}(\omega'_m)) . \end{aligned}$$

Note that  $Q_m$  disagrees with  $P^m$  on some  $m$ -states.

We now show that  $Q_m$  satisfies the constraints reinterpreted on  $\mathcal{L}_m$  by showing that  $Q_m((\varphi_i)_m) = P^m((\varphi_i)_m)$ . Consider an arbitrary  $1 \leq i \leq k$  and observe that

$$\begin{aligned} Q_m((\varphi_i)_m) &= \sum_{\substack{\nu_m \in [(\varphi_i)_m]_m \\ \nu_m \models \neg(\omega_n \vee \omega'_n)}} P^m(\nu_m) \\ &\quad + \sum_{\substack{\nu_m \in [(\varphi_i)_m]_m \\ \nu_m \models \omega_n}} P^m(f_m(\nu_m)) + \sum_{\substack{\nu_m \in [(\varphi_i)_m]_m \\ \nu_m \models \omega'_n}} P^m(f_m^{-1}(\nu_m)) \\ &= P^m((\varphi_i)_m) . \end{aligned}$$

So,  $Q_m$  and  $P^m$  both satisfy the linear constraints on  $\mathcal{L}_m$ . Since the constraints are all linear, the set of probability functions satisfying the constraints is convex and hence  $0.5(Q_m + P^m)$  satisfies the constraints on  $\mathcal{L}_m$  and is different from  $Q_m$  and  $P^m$ . Applying Proposition 4 we note that  $0.5(Q_m + P^m)$  has strictly greater  $m$ -entropy than  $P^m$  which is a contradiction to  $P^m$  being the unique  $m$ -entropy maximiser satisfying the constraints on  $\mathcal{L}_m$ . ■

Hence, all states which are treated symmetrically by the premisses have the same probability under  $P^\infty$ , if  $P^\infty$  exists.

## References

- [1] Adamčík, M. (2014). The Information Geometry of Bregman Divergences and Some Applications in Multi-Expert Reasoning. *Entropy*, 16(12):6338–6381.
- [2] Adamčík, M. (2016). On the Applicability of the ‘Number of Possible States’ Argument in Multi-Expert Reasoning. *Journal of Applied Logic*, 19, Part 1:20–49.
- [3] Adamčík, M. (2017). A logician’s approach to meta-analysis with unexplained heterogeneity. *Journal of Biomedical Informatics*, 71:110–129.

- [4] Adamčík, M. and Wilmers, G. (2014). Probabilistic Merging Operators. *Logique et Analyse*, 57(228):563–590.
- [5] Bacchus, F., Grove, A., Halpern, J., and Koller, D. (1994). Generating new beliefs from old. In *Proceedings of UAI*, pages 37–45. Morgan Kaufmann.
- [6] Bacchus, F., Grove, A., Halpern, J., and Koller, D. (1996). From statistical knowledge to degrees of belief. *Artificial Intelligence*, 87:75–143.
- [7] Barnett, O. and Paris, J. (2008). Maximum entropy inference with quantified knowledge. *Logic Journal of the IGPL*, 16(1):85–98.
- [8] Berger, A., Della Pietra, S., and Della Pietra, V. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.
- [9] Carnap, R. (1945). On inductive logic. *Philosophy of Science*, 12(2):72–97.
- [10] Carnap, R. (1971). A basic system of inductive logic part 1. In Carnap, R. and Jeffrey, R. C., editors, *Studies in inductive logic and probability*, volume 1, pages 33–165. University of California Press, Berkeley CA.
- [11] Carnap, R. (1980). A basic system of inductive logic part 2. In Carnap, R. and Jeffrey, R. C., editors, *Studies in inductive logic and probability*, volume 2, pages 7–155. University of California Press, Berkeley CA.
- [12] Caticha, A. (2014). Entropic Dynamics: an inference approach to quantum theory, time and measurement. *Journal of Physics: Conference Series*, 504(1):012009.
- [13] Caticha, A. and Giffin, A. (2006). Updating Probabilities. In *Proceedings of MaxEnt*, volume 872, pages 31–42.
- [14] Chang, C. and Keisler, H. (2015). *Model Theory, Studies in Logic and the Foundations of Mathematics, vol73*. North Holland Publishing Co.
- [15] Chen, C. H. (1990). Maximum Entropy Analysis for Pattern Recognition. In Fougère, P. F., editor, *Maximum Entropy and Bayesian Methods*, pages 403–408. Springer.
- [16] Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley and Sons, New York, second (2006) edition.
- [17] Gaifman, H. (1964). Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2(1):1–18.
- [18] Grove, A., Halpern, J., and Koller, D. (1994). Random worlds and maximum entropy. *Journal of Artificial Intelligence Research*, 2:33–88.
- [19] Grove, A., Halpern, J., and Koller, D. (1996a). Asymptotic conditional probabilities: the non-unary case. *Journal of Symbolic Logic*, 61(1):250–276.
- [20] Grove, A., Halpern, J., and Koller, D. (1996b). Asymptotic conditional probabilities: the unary case. *SIAM Journal of Computing*, (1):1–51.

- [21] Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic logics and probabilistic networks*. Synthese Library. Springer, Dordrecht.
- [22] Hodges, W. (2015). *Model Theory*. Cambridge University Press.
- [23] Howarth, E. and Paris, J. B. (2019). Pure Inductive Logic with Functions. *Journal of Symbolic Logic*, pages 1–22.
- [24] Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physical Review*, 106:620–630.
- [25] Jaynes, E. T. (1957b). Information theory and statistical mechanics II. *Physical Review*, 108(2):171–190.
- [26] Jaynes, E. T. (1990). Notes on present status and future prospects. *Maximum Entropy and Bayesian Methods*, W.T. Grandy and L.H. Schick, (edt), pages 1–13.
- [27] Jaynes, E. T. (1991). How should we use entropy in economics? available online: [www.leibniz.imag.fr/LAPLACE/Jaynes/prob.html](http://www.leibniz.imag.fr/LAPLACE/Jaynes/prob.html).
- [28] Kapur, J. N. (1983a). Non-additive measures of entropy and distributions of statistical mechanics. *Indian Journal of Pure and Applied Mathematics*, 14(11):1372–1387.
- [29] Kapur, J. N. (1983b). Twenty five years of maximum entropy. *Journal of Mathematical and Physical Sciences*, 17:103–156.
- [30] Kapur, J. N. and Kesavan, H. K. (1992). Entropy Optimization Principles and Their Applications. In Singh, V. P. and Fiorentino, M., editors, *Entropy and Energy Dissipation in Water Resources*, pages 3–20. Springer.
- [31] Kern-Isberner, G. (1998). Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98(1-2):169–208.
- [32] Kern-Isberner, G. and Lukasiewicz, T. (2004). Combining probabilistic logic programming with the power of maximum entropy. *Artificial Intelligence*, 157(1-2):139–202.
- [33] Landes, J. and Masterton, G. (2017). Invariant Equivocation. *Erkenntnis*, 82:141–167.
- [34] Landes, J. and Williamson, J. (2015). Justifying objective Bayesianism on predicate languages. *Entropy*, 17(4):2459–2543.
- [35] Paris, J. and Vencovská, A. (2015). *Pure inductive logic*. Cambridge University Press, Cambridge.
- [36] Paris, J. B. (1994). *The uncertain reasoner’s companion*. Cambridge University Press, Cambridge.
- [37] Paris, J. B. and Rafiee Rad, S. (2008). Inference Processes for Quantified Predicate Knowledge. In Hodges, W. and de Queiroz, R., editors, *Proceedings of WoLLIC*, volume 5110 of *LNCS*, pages 249–259. Springer.

- [38] 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. Springer.
- [39] Rafiee Rad, S. (2009). *Inference Processes for on First Order Probabilistic Languages*. PhD thesis, Manchester Institute for Mathematical Sciences.
- [40] Rafiee Rad, S. (2017). Equivocation axiom on first order languages. *Studia Logica*, 105(1):121–152.
- [41] Rafiee Rad, S. (2018). Maximum Entropy Models for  $\Sigma_1$  Sentences. *Journal of Applied Logics - IfCoLoG Journal of Logics and their Applications*, 5(1):287–300.
- [42] Rafiee Rad, S. (2019). On probabilistic characterisation of models of first order theories. manuscript.
- [43] Rényi, A. (1961). On measures of information and entropy. *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 106:547–561.
- [44] Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, 31:131–144.
- [45] Williamson, J. (2008). Objective Bayesian probabilistic logic. *Journal of Algorithms in Cognition, Informatics and Logic*, 63:167–183.
- [46] Williamson, J. (2017). *Lectures on Inductive Logic*. Oxford University Press, Oxford.
- [47] Zellner, A. (1991). Bayesian Methods and Entropy in Economics and Econometrics. In Grandy, W. T. and Schick, L. H., editors, *Maximum Entropy and Bayesian Methods*, pages 17–31. Springer.