ENTROPY OF FORMULAS

VERA KOPONEN

Abstract. A probability distribution can be given to the set of isomorphism classes of models with universe $\{1, \ldots, n\}$ of a sentence in first-order logic. We study the entropy of this distribution and derive a result from the 0-1 law for first-order sentences.
Keywords: first-order logic, finite models, entropy, 0-1 law.

Introduction

We will study the entropy of a probability distribution on the set of isomorphism classes of models with universe $\{1, \ldots, n\}$ of a first-order sentence (i.e. closed formula). Recall that, for a finite probability distribution $\mathbf{p} = (p_1, \ldots, p_k)$, the entropy of $\mathbf{p}$ is $H(\mathbf{p}) = -\sum_{i=1}^{k} p_i \ln p_i$ (where we adopt the convention that $0 \ln 0 = 0$). For any probability distribution $\mathbf{p} = (p_1, \ldots, p_k)$ we have (see [6], Theorems 3.7 and 3.10, for instance) $0 \leq H(\mathbf{p}) \leq \ln k$ and

(a) $H(\mathbf{p}) = \ln k$ if and only if $p_i = 1/k$ for every $i = 1, \ldots, k$, and
(b) $H(\mathbf{p}) = 0$ if and only if $p_i = 1$ for some i.

Let L be a (first-order) language with finitely many relation, function and constant symbols. If φ is an L-sentence which has at least one model with exactly n elements, then let $A_1, \ldots, A_{k_n}$ be an enumeration of mutually non-isomorphic L-structures with universe $\{1, \ldots, n\}$, such that each A_i is a model of φ and any model of φ with exactly n elements is isomorphic to some A_i. Let $[A_i]$ be the set of all L-structures with universe $\{1, \ldots, n\}$ which are isomorphic to A_i. If m_n is the number of L-structures A with universe $\{1, \ldots, n\}$ such that φ is true in A, then $\mathbf{p} = (p_1, \ldots, p_{k_n})$, where $p_i = |[A_i]|/m_n$ for $i = 1, \ldots, k_n$, is a probability distribution. Hence we can consider the entropy $H(\mathbf{p})$ which in this case we denote by $H_n(\varphi)$, and we call it 'the entropy of φ for n-element models'. If φ has no model with exactly n elements then we let $H_n(\varphi) = 0$. It follows that if $\mathbf{p}$ is as defined above, then $0 \leq H_n(\varphi) \leq \ln k_n$ and from (a) and (b) we get:

(a)' $H_n(\varphi) = \ln k_n$ if and only if $[A_i]$ and $[A_j]$ contain the same number of structures for any i and any j, and
(b)' $H_n(\varphi) = 0$ implies that any two models of φ with exactly n elements are isomorphic.

The entropy of a formula is not particularly well-behaved with respect to the relation '$\vdash$', where, for L-sentences φ and ψ, $\varphi \vdash \psi$ means that any L-structure which is a model of φ is also a model of ψ. We may have $\varphi_1 \vdash \varphi_2$ and $H_n(\varphi_1) < H_n(\varphi_2)$, but we may also have $\psi_1 \vdash \psi_2$ and $H_n(\psi_1) > H_n(\psi_2)$; examples showing this are given at the end of the paper.

However, from the 0-1 law of (first-order) formulas we may draw a conclusion about the entropy $H_n(\varphi)$. The 0-1 law says that, under the assumption that L has only finitely many relation symbols and no function or constant symbols, for any L-formula φ, the proportion of L-structures with universe $\{1, \ldots, n\}$ in which φ is true approaches either 0 or 1, as n approaches ∞. Under the additional condition that *not all* relation symbols of L are unary, we will prove that if the above mentioned proportion approaches 1 then $H_n(\varphi)$ is asymptotic to $\ln k_n$ (where k_n is as above). By being asymptotic to $\ln k_n$ we mean that $H_n(\varphi)/\ln k_n \to 1$ as $n \to \infty$. Intuitively this means that, if the proportion of

L-structures with universe $\{1, \ldots, n\}$ in which φ is true approaches 1 as $n \to \infty$, then the entropy $H_n(\varphi)$ approaches maximal entropy as $n \to \infty$.

In the case that the proportion of L-structures with universe $\{1, \ldots, n\}$ in which φ is true approaches 0 as $n \to \infty$, we cannot conclude anything particular about the asymptotic behaviour of $H_n(\varphi)$. For example, we may have $H_n(\varphi) = 0$ for every n, but we may also have $H_n(\varphi) = \ln k_n$ for every n, and it may be the case that $\lim_{n\to\infty} H_n(\varphi)$ and $\lim_{n\to\infty} H_n(\varphi)/\ln k_n$ don't exist; examples illustrating these possibilities are given at the end.

Acknowledgements. The idea to consider the questions dealt with in this paper was suggested to me by Erik Palmgren. I also thank the anonymous referee for suggesting simplifications of the proof of Theorem 3.

Notation and terminology. For definitions of, and elementary results about (first-order) languages and structures, see [5] or [1] for instance; the notation and terminology used here, for structures and languages, follows [5]. We always assume, even when not explicitly mentioned, that the symbol '$=$' is part of the language and is interpreted in structures as the identity relation. We say that a language is *finite and relational* if it has only finitely many relation (also called predicate) symbols and no constant or function symbols. A language is said to be *monadic* if every relation symbol of it, except for $=$, is unary. If A and B are L-structures then $A \cong B$ means that A is *isomorphic* to B. We may, as usual, identify a structure with its universe (or domain) notationally. For a k-ary relation symbol R of the language L and an L-structure A, R^A denotes the interpretation of R in A. For an L-structure A and an L-sentence φ (i.e. closed L-formula), $A \models \varphi$ means that φ is true (or satisfied) in A, or in other words, that A is a model of φ. If X is a set then $|X|$ denotes its cardinality. With $k, m, n, n_1, n_2, \ldots$ we will denote positive integers.

Entropy of formulas

Throughout this paper we will assume that L is a finite and relational language, although we will occasionally repeat this assumption.

Definition 1. Let $\mathcal{S}_n$ be the set of all L-structures with universe $\{1, \ldots, n\}$. Since L is finite, each $\mathcal{S}_n$ is finite. If $A \in \mathcal{S}_n$ then let $[A] = \{B \in \mathcal{S}_n : B \cong A\}$. Let $\mathcal{S}'_n = \{[A] : A \in \mathcal{S}_n\}$. If φ is an L-sentence then let $\mathcal{M}_n(\varphi) = \{A \in \mathcal{S}_n : A \models \varphi\}$ and let $\mathcal{M}'_n(\varphi) = \{[A] : A \in \mathcal{M}_n(\varphi)\}$

For any L-sentence φ we can consider a probability distribution on $\mathcal{M}'_n(\varphi)$ by letting each $[A] \in \mathcal{M}'_n(\varphi)$ have probability $|[A]|/|\mathcal{M}(\varphi)|$. So if $A \in \mathcal{M}_n(\varphi)$, and supposing that each structure in $\mathcal{S}_n$ is equally probable, $|[A]|/|\mathcal{M}(\varphi)|$ is the probability that a model of φ in $\mathcal{S}_n$ is isomorphic to A.

Definition 2. Let L be a finite and relational language. For an L-sentence φ, we define the *entropy of φ for n-element models*, denoted $H_n(\varphi)$, by

$$H_n(\varphi) = -\sum_{i=1}^{k} \frac{|[A_i]|}{|\mathcal{M}(\varphi)|} \ln \frac{|[A_i]|}{|\mathcal{M}(\varphi)|},$$

where $[A_1], \ldots, [A_k]$ is an enumeration of $\mathcal{M}'_n(\varphi)$ without repetitions, if $\mathcal{M}_n(\varphi) \neq \emptyset$. If $\mathcal{M}_n(\varphi) = \emptyset$ then define $H_n(\varphi) = 0$.

The so-called 0-1 law ([2], [4], [1] Theorem 4.1.5, [5] Theorem 7.4.7) states that, for any L-sentence φ,

$$\text{the limit} \quad \lim_{n\to\infty} \frac{|\mathcal{M}_n(\varphi)|}{|\mathcal{S}_n|} \quad \text{exists and is either 0 or 1.}$$

Theorem 3. *Let L be a finite and relational language which is not monadic and let φ be an L-sentence.*

$$\textit{If} \quad \lim_{n\to\infty} \frac{|\mathcal{M}_n(\varphi)|}{|\mathcal{S}_n|} = 1 \quad \textit{then} \quad \lim_{n\to\infty} \frac{H_n(\varphi)}{\ln |\mathcal{M}'_n(\varphi)|} = 1.$$

Remark 4. (i) If $|\mathcal{M}_n(\varphi)|/|\mathcal{S}_n| \to 0$ as $n \to \infty$, then it may or may not be the case that $H_n(\varphi)/\ln |\mathcal{M}'_n(\varphi)| \to 1$ as $n \to \infty$. Examples 6, 7 and 8 show this.
(ii) The theorem does not hold for monadic L. Example 9 shows this.

In order to prove Theorem 3 we will use the following lemma which should occur in the literature in one form or another, but for the sake of completeness a (short) proof is nevertheless given in the appendix.

Lemma 5. *Suppose that a_n and b_n are two sequences such that $a_n \geq b_n > 0$, for every n, $\lim_{n\to\infty} b_n = \infty$ and $\lim_{n\to\infty} a_n/b_n = 1$. Then $\lim_{n\to\infty}(\ln a_n - \ln b_n) = 0$, and consequently $\lim_{n\to\infty} \ln a_n/\ln b_n = \lim_{n\to\infty} \ln 2a_n/\ln b_n = 1$.*

We now prove Theorem 3. Suppose that L is a finite and relational language which is not monadic and suppose that φ is a formula in L such that

$$\lim_{n\to\infty} \frac{|\mathcal{M}_n(\varphi)|}{|\mathcal{S}_n|} = 1.$$

We introduce some simpler notation. For every n, let

$$s_n = |\mathcal{S}_n|, \quad s'_n = |\mathcal{S}'_n|, \quad m_n = |\mathcal{M}_n(\varphi)|, \quad m'_n = |\mathcal{M}'_n(\varphi)|.$$

With the new notation we have

$$\lim_{n\to\infty} \frac{m_n}{s_n} = 1 \tag{1}$$

and we want to prove that $H_n(\varphi)/\ln m'_n \to 1$ as $n \to \infty$.

Since L is not monadic, Theorem 8 in [2] says that

$$\lim_{n\to\infty} \frac{s_n}{s'_n \cdot n!} = 1. \tag{2}$$

For every $[A] \in \mathcal{S}'_n$, $|[A]| = n!/k$ where k is the order of the group of automorphisms of A. So if $|[A]| < n!$ then $|[A]| \leq n!/2$. A structure $A \in \mathcal{S}_n$ is *rigid* if A has only one automorphism. It follows that A is rigid if and only if $|[A]| = n!$.

Let

$$\begin{aligned}
r_n &= \big|\{A \in \mathcal{S}_n : A \text{ is rigid}\}\big|, \\
f_n &= \big|\{A \in \mathcal{M}_n(\varphi) : A \text{ is rigid}\}\big|, \\
\bar{f}_n &= r_n - f_n = \big|\{A \in \mathcal{S}_n - \mathcal{M}_n(\varphi) : A \text{ is rigid}\}\big|, \\
f'_n &= \big|\{[A] \in \mathcal{M}'_n(\varphi) : A \text{ is rigid}\}\big|.
\end{aligned}$$

Observe that $f_n = n! f'_n$ and, by (1), that $\lim_{n\to\infty} \bar{f}_n/s_n = 0$. From (2) together with Lemma 4.3.2 and Proposition 4.3.3 in [1] we get

$$\lim_{n\to\infty} \frac{r_n}{s_n} = 1$$

and from this and (1) we get

$$\frac{f_n}{m_n} = \frac{f_n}{s_n} \cdot \frac{s_n}{m_n} = \left(\frac{r_n}{s_n} - \frac{\bar{f}_n}{s_n}\right) \cdot \frac{s_n}{m_n} \to 1 \text{ as } n \to \infty. \tag{3}$$

Since L is not monadic it has at least one relation symbol R of arity k where $k \geq 2$. For each $A \in \mathcal{S}_n$ and each k-tuple $\bar{a}$ of elements from $\{1, \ldots, n\}$ we have $\bar{a} \in R^A$ or $\bar{a} \notin R^A$. As there are n^k such k-tuples, there are 2^{n^k} possibilities for R^A. Since $2^{n^k} \geq 2^{n^2}$, there are at least 2^{n^2} different structures in $\mathcal{S}_n$, so $s_n \geq 2^{n^2}$, which gives

$$\frac{\ln s_n}{\ln(n!)} \geq \frac{\ln(2^{n^2})}{\ln(n!)} \geq \frac{\ln(2^{n^2})}{\ln(n^n)} = \frac{n \ln 2}{\ln n} \to \infty \text{ as } n \to \infty. \tag{4}$$

By Lemma 5 and (1) we have $\lim_{n\to\infty}(\ln m_n - \ln s_n) = 0$, which together with (4) implies that

$$\frac{\ln m_n}{\ln(n!)} = \frac{\ln m_n - \ln s_n}{\ln(n!)} + \frac{\ln s_n}{\ln(n!)} \to \infty \text{ as } n \to \infty. \tag{5}$$

And (5) in turn gives

$$\frac{\ln \frac{m_n}{n!}}{\ln m_n} = 1 - \frac{\ln(n!)}{\ln m_n} \to 1 \text{ as } n \to \infty. \tag{6}$$

As $\frac{m_n}{n!} \leq m'_n \leq m_n$ we have $\ln \frac{m_n}{n!} \leq \ln m'_n \leq \ln m_n$ which together with (6) implies that

$$\lim_{n\to\infty} \frac{\ln m'_n}{\ln m_n} = 1. \tag{7}$$

Since

$$\frac{H_n(\varphi)}{\ln |\mathcal{M}'_n(\varphi)|} = \frac{H_n(\varphi)}{\ln m_n} \cdot \frac{\ln m_n}{\ln m'_n}$$

it suffices, by (7), to prove that $H_n(\varphi)/\ln m_n \to 1$ as $n \to \infty$. From the definitions of f_n and f'_n it follows that $f_n = n! f'_n$ and that

$$H_n(\varphi) \geq -f'_n \frac{n!}{m_n} \ln \frac{n!}{m_n} = -\frac{f_n}{m_n} \ln \frac{n!}{m_n}. \tag{8}$$

By (8), (6) and (3) we get

$$\frac{H_n(\varphi)}{\ln m_n} \geq \frac{-\frac{f_n}{m_n} \ln \frac{n!}{m_n}}{\ln m_n} = \frac{f_n}{m_n} \cdot \frac{\ln \frac{m_n}{n!}}{\ln m_n} \to 1 \quad \text{as} \quad n \to \infty. \tag{9}$$

Since for every probability distribution $\mathbf{p} = (p_1, \ldots, p_k)$ $H_n(\mathbf{p}) \leq \ln k$, we have $H_n(\varphi) \leq \ln m'_n \leq \ln m_n$ and hence $H_n(\varphi)/\ln m_n \leq 1$, for all sufficiently large n. Together with (9) this implies that

$$\lim_{n\to\infty} \frac{H_n(\varphi)}{\ln m_n} = 1$$

and, as shown above, Theorem 3 follows from this.

Examples

Example 6. This example shows that the conclusion of Theorem 3 may hold even if $|\mathcal{M}_n(\varphi)|/|\mathcal{S}_n| \to 0$ as $n \to \infty$. Let L have one binary relation symbol R and no other relation symbols (exept for =). Let ψ the following L-sentence

$$\forall x, y R(x, y) \vee \forall x, y \neg R(x, y).$$

For any n, $\mathcal{M}'_n(\psi)$ has two elements and each of them contains exactly one structure. It follows that $|\mathcal{M}_n(\psi)| = 2$, for every n. In the proof of Theorem 3 we showed that

$|\mathcal{S}_n| \geq 2^{n^2}$, so we have $|\mathcal{M}_n(\varphi)|/|\mathcal{S}_n| \to 0$ as $n \to \infty$. Since for any n, $|\mathcal{M}'_n(\psi)| = 2$ and $H_n(\psi) = -1/2\ln(1/2) - 1/2\ln(1/2) = \ln 2$, we get $H_n(\psi)/\ln|\mathcal{M}'_n(\psi)| = 1$ for every n.

Example 7. This example shows that the conclusion of Theorem 3 may fail if $|\mathcal{M}_n(\varphi)|/|\mathcal{S}_n| \to 0$ as $n \to \infty$. It also shows that for certain formulas ψ and θ we have $\psi \vdash \theta$ and $H_n(\theta) < H_n(\psi)$ for all sufficiently large n. Let L and ψ be as in the previous example. Let χ be an L-sentence which expresses that

> R is an equivalence relation such that R has exactly two equivalence classes and one of them contains exactly one element.

Finally let θ be $\psi \vee \chi$. For any n, $\mathcal{M}'_n(\theta)$ has three elements: The first contains the unique structure in $\mathcal{S}_n$ which satisfies $\forall x, y R(x,y)$; the second contains the unique strucure in $\mathcal{S}_n$ which satisfies $\forall x, y \neg R(x,y)$; the third element of $\mathcal{M}'_n(\theta)$ contains the precisely n different structures in $\mathcal{S}_n$ in which χ is true. It follows that $|\mathcal{M}_n(\theta)| = n+2$ and

$$\begin{aligned} H_n(\theta) &= -2\Big(\frac{1}{n+2}\ln\frac{1}{n+2}\Big) - \frac{n}{n+2}\ln\frac{n}{n+2} \\ &= 2\cdot\frac{\ln(n+2)}{n+2} + \frac{n}{n+2}\ln\frac{n+2}{n} \to 0 \quad \text{as} \quad n \to \infty, \\ &\text{because} \quad \frac{\ln(n+2)}{n+2} \to 0, \quad \frac{n}{n+2} \to 1 \quad \text{and} \quad \ln\frac{n+2}{n} \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Therefore, $H_n(\theta)/\ln|\mathcal{M}'_n(\theta)| = H_n(\theta)/\ln 3 \to 0$ as $n \to \infty$. We clearly have $\psi \vdash \theta$. Since $H_n(\psi) = \ln 2$ for all n and $\lim_{n\to\infty} H_n(\theta) = 0$ it follows that $H_n(\psi) > H_n(\theta)$ for all sufficiently large n.

Example 8. This example shows that if $|\mathcal{M}_n(\varphi)|/|\mathcal{S}_n| \to 0$ as $n \to \infty$ then $\lim_{n\to\infty} H_n(\varphi)$ may not exist. It also shows that we may have $\varphi \vdash \psi$ and $H_n(\varphi) < H_n(\psi)$. Let L have two relation symbols R, P (except for $=$) where R is binary and P is unary. Let σ_1 be a sentence which expresses that

> R is symmetric and irreflexive,
> for every x there exists a unique y such that $R(x,y)$, and
> either $\forall x P(x)$ or $\forall x \neg P(x)$.

Let σ_2 be the sentence $\forall x, y\big(\neg R(x,y) \wedge \neg P(x)\big)$ and let σ be the sentence $\sigma_1 \vee \sigma_2$.

Then, for every n, $\mathcal{M}'_{2n+1}(\sigma)$ has exactly one element which contains exactly one structure. And, for every n, $\mathcal{M}'_{2n}(\sigma)$ has exactly three elements; one of them contains exactly one structure and each of the other two contains exactly $a_n = (2n)!/2^n n!$ structures; consequently $\big|\mathcal{M}_{2n}(\sigma)\big| = 2a_n + 1$. It follows that $H_{2n+1}(\sigma) = -\ln 1 = 0$, for every n. For every n we also have

$$\begin{aligned} H_{2n}(\sigma) &= -\frac{1}{2a_n+1}\ln\frac{1}{2a_n+1} - 2\Big(\frac{a_n}{2a_n+1}\ln\frac{a_n}{2a_n+1}\Big) \\ &= \frac{\ln(2a_n+1)}{2a_n+1} + \frac{2a_n}{2a_n+1}\ln\frac{2a_n+1}{a_n} \;\to\; \ln 2 \quad \text{as} \quad n \to \infty, \\ &\text{because} \quad \lim_{n\to\infty} a_n = \infty \quad \text{and} \quad \lim_{x\to\infty}\frac{\ln x}{x} = 0. \end{aligned}$$

Hence $\lim_{n\to\infty} H_n(\sigma)$ does not exist; and neither does $\lim_{n\to\infty} H_n(\sigma)/|\mathcal{M}'_n(\sigma)|$ exist since $|\mathcal{M}'_n(\sigma)|$ is always 1 or 3. Clearly, $\sigma_2 \vdash \sigma$ and $H_n(\sigma_2) = 0$ for all n. Hence $H_n(\sigma_2) < H_n(\sigma)$ for all sufficiently large even n.

Example 9. The following example shows that the assumption about non-monadic language L in Theorem 3 is necessary. Let L have only one unary relation symbol P and no other relation symbols (in addition to $=$). Let φ be any sentence which is true in every L-structure; for instance, we can let φ be $\forall x(x = x)$. Then $\mathcal{M}_n(\varphi) = \mathcal{S}_n$. We will show that

$$\lim_{n\to\infty} \frac{H_n(\varphi)}{\ln|\mathcal{M}'_n(\varphi)|} = \frac{1}{2}.$$

First note that for any $A, B \in \mathcal{S}_n$, $A \cong B$ if and only if $|P^A| = |P^B|$, so $|\mathcal{M}'_n(\varphi)| = |\mathcal{S}'_n| = n$. Hence it suffices to prove that $H_n(\varphi)/\ln n \to 1/2$ as $n \to \infty$. For any n and $1 \leq i \leq n$, let $p_{n,i} = \binom{n}{i}/2^n$, so $H_n(\varphi) = -\sum_{i=1}^n p_{n,i} \ln p_{n,i}$. Let $H^*_n(\varphi) = -\sum_{i=1}^n p_{n,i} \log p_{n,i}$, where log is the logarithm with base 2. From the identity $\ln a = \log a/\log e$ it follows that $H_n(\varphi) = H^*_n(\varphi)/\log e$. By [3] (Theorem 3) we have

$$H^*_n(\varphi) = \log\sqrt{\frac{\pi e n}{2}} + \mathrm{O}\big((4n)^{-2}\big).$$

Therefore

$$\begin{aligned}\frac{H_n(\varphi)}{\ln n} &= \frac{\log e}{\log n} H_n(\varphi) = \frac{H^*_n(\varphi)}{\log n} \\ &= \frac{\log\sqrt{\frac{\pi e n}{2}} + \mathrm{O}\big((4n)^{-2}\big)}{\log n} \\ &= \frac{1}{2}\cdot\frac{\log n + \log\frac{\pi e}{2} + 2\cdot\mathrm{O}\big((4n)^{-2}\big)}{\log n} \\ &\to \frac{1}{2} \text{ as } n \to \infty.\end{aligned}$$

Appendix

Proof of Lemma 5: Suppose that a_n and b_n are two sequences such that $a_n \geq b_n > 0$, for every n, $\lim_{n\to\infty} b_n = \infty$ and $\lim_{n\to\infty} a_n/b_n = 1$. By the continuity of ln we have $\lim_{n\to\infty}(\ln a_n - \ln b_n) = \lim_{n\to\infty} \ln\frac{a_n}{b_n} = 0$, and consequently

$$\frac{\ln a_n}{\ln b_n} = \frac{\ln a_n - \ln b_n}{\ln b_n} + 1 \to 1 \text{ as } n \to \infty \text{ (because } \lim_{n\to\infty} b_n = \infty).$$

Since $\ln 2a_n = \ln 2 + \ln a_n$ it follows that $\ln 2a_n/\ln b_n \to 1$ as $n \to \infty$.

References

[1] H-D. Ebbinghaus, J. Flum, *Finite Model Theory*, Second Edition, Springer-Verlag, 1999.
[2] R. Fagin, Probabilities on finite models, *The Journal of Symbolic Logic* (41) 1976, 50-58.
[3] O. Frank, J. Öhrvik, Entropy of sums of random digits, *Computational Statistics & Data Analysis*, (17) 1994, 177-184.
[4] Y. V Glebskii, D. I. Kogan, M. I. Liogonkii, V. A. Talanov, Volume and fraction of satisfiability of formulas of the lower predicate calculus, *Kibernetyka* (2) 1969, 17-27.
[5] W. Hodges, *Model theory*, Cambridge University Press, 1993.
[6] G. A. Jones, J. M. Jones, *Information and Coding Theory*, Springer-Verlag, 2000.

Dept. of Mathematics, Uppsala University, Box 480, 75106 Uppsala, Sweden
E-mail address: vera@math.uu.se