On the complexity of first-order logic of probability of type 1

Stanislav O. Speranski

Steklov Mathematical Institute of RAS

Moscow 2025

1. Probabilities on the domain

Consider a (first-order) signature ς. Then, following [1], by an $\mathscr{L}_1(\varsigma)$-structure we mean a triple $\langle D, \pi, \mathsf{p}\rangle$ where:

- D is a non-empty set;
- π is a ς-structure, as defined in first-order logic, with domain D;
- p is a discrete probability distribution on D, i.e. a function from D to $[0,1]$ such that

$$|\{d \in D \mid \mathsf{p}(d) \neq 0\}| \leqslant \aleph_0 \quad \text{and} \quad \sum_{d \in D} \mathsf{p}(d) = 1,$$

which generates the probability measure P on the powerset of D as follows:

$$\mathsf{P}(A) \;:=\; \sum_{d \in A} \mathsf{p}(d).$$

Note, in passing, that given p as above and a non-zero $k \in \mathbb{N}$, we can define a discrete distribution p^k on D^k by

$$\mathsf{p}^k(d_1, \ldots, d_k) \;:=\; \mathsf{p}(d_1) \cdot \ldots \cdot \mathsf{p}(d_k),$$

which generates the measure P^k on the powerset of D^k, of course. Evidently, if $A \subseteq D^k$, and A' is obtained from A by permuting some of the coordinates, then $\mathsf{P}^k(A')$ coincides with $\mathsf{P}^k(A)$.

As for the syntax of $\mathscr{L}_1$, its alphabet includes two disjoint countable sets

$$\mathrm{Var} := \{x, y, z, \ldots\} \quad \text{and} \quad \mathbf{Var} := \{\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \ldots\},$$

whose elements are called individual variables and field variables respectively. Of course, the latter are intended to range over reals. In addition, we have:

- the logical symbols $\top$, $\bot$, $\wedge$, $\vee$ and $\neg$;
- the quantifier symbols $\forall$ and $\exists$;
- the symbols 0, 1, $+$, $-$, $\cdot$, $=$ and $\leqslant$ of the language of ordered fields;
- a special symbol μ, which will be interpreted using probability measures.

Given a signature ς, let $\mu\text{-Form}^1_\varsigma$ and $\mu\text{-Term}^1_\varsigma$ be the sets defined simultaneously by the following conditions:

1. $\mu\text{-Form}^1_\varsigma$ contains all atomic fist-order ς-formulas, including $\top$ and $\bot$;
2. $\mu\text{-Term}^1_\varsigma$ contains 0 and 1;
3. $\mu\text{-Term}^1_\varsigma$ contains all field variables;
4. $\mu\text{-Form}^1_\varsigma$ is closed under $\wedge$, $\vee$ and $\neg$;
5. $\mu\text{-Form}^1_\varsigma$ is closed under $\mathrm{Q}x$, for all $\mathrm{Q} \in \{\forall, \exists\}$ and $x \in \mathrm{Var}$;
6. $\mu\text{-Form}^1_\varsigma$ is closed under $\mathrm{Q}\mathfrak{a}$, for all $\mathrm{Q} \in \{\forall, \exists\}$ and $\mathfrak{a} \in \mathbf{Var}$;
7. if ϕ belongs to $\mu\text{-Form}^1_\varsigma$, and $\vec{x} \in \mathrm{Var}^+$, then $\mu_{\vec{x}}(\phi)$ belongs to $\mu\text{-Term}^1_\varsigma$;
8. $\mu\text{-Term}^1_\varsigma$ is closed under $+$, $-$ and $\cdot$;
9. if t_1 and t_2 belong to $\mu\text{-Term}^1_\varsigma$, then $t_1 = t_2$ and $t_1 \leqslant t_2$ belong to $\mu\text{-Form}^1_\varsigma$.

Elements of these sets are called $\mathscr{L}_1(\varsigma)$-formulas and $\mathscr{L}_1(\varsigma)$-terms respectively. By the depth of an $\mathscr{L}_1(\varsigma)$-formula ϕ, denoted by $\mathrm{dp}(\phi)$, we mean the largest number of nested occurrences of μ in ϕ; similarly for $\mathscr{L}_1(\varsigma)$-terms.

An $\mathscr{L}_1(\varsigma)$-formula ϕ is:

- basic if ϕ has the form $t_1 = t_2$ or $t_1 \leqslant t_2$ where t_1 and t_2 are $\mathscr{L}_1(\varsigma)$-terms;
- regular if all occurrences of (atomic) first-order ς-formulas in ϕ are in the scope of μ.

An $\mathscr{L}_1(\varsigma)$-sentence is an $\mathscr{L}_1(\varsigma)$-formula with no free variable occurrences.

Consider an $\mathscr{L}_1(\varsigma)$-structure $\mathcal{M} = \langle D, \pi, \mathsf{p} \rangle$. Hence the individual variables are intended to range over D. By a valuation in $\mathcal{M}$ we mean a pair $\langle \zeta, \gamma \rangle$ where ζ and γ are functions from Var and $\mathbf{Var}$ to D and $\mathbb{R}$ respectively. Then

$$\mathcal{M} \Vdash \phi[\zeta, \gamma]$$

read as 'ϕ is true in $\mathcal{M}$ under $\langle \zeta, \gamma \rangle$', can be defined by induction on the depth of ϕ. Of course, in case ϕ is a first-order ς-formula, we employ the ς-structure π, viz.

$$\mathcal{M} \Vdash \phi[\zeta, \gamma] \quad :\Longleftrightarrow \quad \pi \vDash \phi[\zeta].$$

Assuming $\mathrm{dp}(\phi) > 0$, the idea is that given an arbitrary valuation $\langle \eta, \delta \rangle$ in $\mathcal{M}$, we interpret each $\mu_{(x_1,\ldots,x_k)}(\psi)$ with $\mathrm{dp}(\psi) < \mathrm{dp}(\phi)$ as

$$\mathsf{P}^k\left(\left\{(d_1,\ldots,d_k) \in D^k \mid \mathcal{M} \Vdash \psi\left[\eta^{\vec{x}}_{\vec{d}}, \delta\right]\right\}\right)$$

where $\eta^{\vec{x}}_{\vec{d}}$ is the function from Var to D such that

$$\eta^{\vec{x}}_{\vec{d}}(u) = \begin{cases} d_i & \text{if } u = x_i \text{ with } i \in \{1,\ldots,k\} \\ \eta(u) & \text{otherwise.} \end{cases}$$

We call an $\mathscr{L}_1(\varsigma)$-sentence valid if it is true in all $\mathscr{L}_1(\varsigma)$-structures.

Theorem 1.1 (see [1])

Let ς *be* $\langle P^2 \rangle$ *where* P *is a binary predicate symbol. Then the validity problem for* $\mathscr{L}_1(\varsigma)$*-sentences is* Π^2_1*-complete. However, if we limit ourselves to at most countable domains, the corresponding problem becomes* Π^1_∞*-complete.*

Let $\mathscr{L}_1^\natural$ be the sublanguage of $\mathscr{L}_1$ obtained by excluding field variables, and hence quantifiers over reals. So the definitions of $\mathscr{L}_1^\natural(\varsigma)$-formula and $\mathscr{L}_1^\natural(\varsigma)$-term are like those for $\mathscr{L}_1$ except that items 3 and 6 are removed.

Theorem 1.2 (see [1])

Let ς *be as before. Then the validity problem for* $\mathscr{L}_1^\natural(\varsigma)$*-sentences is* Π^1_1*-hard, even if we confine ourselves to at most countable domains.*

2. Concerning higher-order arithmetic

In second-order arithmetic, in addition to individual variables x, y, z, . . . , which are intended to range over $\mathbb{N}$, we have k-ary set variables

$$X^k,\ Y^k,\ Z^k,\ \ldots,$$

intended to range over the powerset of $\mathbb{N}^k$, for each positive k. Hence the atomic second-order formulas additionally include all expressions of the form

$$X^k\,(t_1, \ldots, t_k)$$

where t_1, . . . , t_k are terms. In what follows we shall write X instead of X^1.

Let $\mathfrak{N}$ be the standard model of Peano arithmetic presented in the signature $\langle 0, \mathsf{s}, +, \cdot ; = \rangle$. We write σ_s for the much smaller signature $\langle 0, \mathsf{s}; = \rangle$ and $\mathfrak{N}_\mathsf{s}$ for the σ_s-reduct of $\mathfrak{N}$. Take

$$\sigma_\mathsf{s}^\sharp \ := \ \langle 0, \mathsf{s}; =, Y^2 \rangle$$

where Y^2 is treated as a binary predicate symbol. For each $R \subseteq \mathbb{N}^2$, denote by $\langle \mathfrak{N}_\mathsf{s}, R \rangle$ the $\sigma_\mathsf{s}^\sharp$-expansion of $\mathfrak{N}_\mathsf{s}$ in which Y^2 is interpreted as R.

Lemma 2.1 (see [9, Section 5])

There exist first-order $\sigma_\mathsf{s}^\sharp$-formulas $\Psi_+(x, y, z)$, $\Psi_\cdot(x, y, z)$ and a first-order $\sigma_\mathsf{s}^\sharp$-sentence Δ such that for every $R \subseteq \mathbb{N}^2$,

$$\langle \mathfrak{N}_\mathsf{s}, R \rangle \models \Delta \quad \Longleftrightarrow \quad \begin{array}{c} \Psi_+(x, y, z) \text{ and } \Psi_\cdot(x, y, z) \text{ define} \\ \text{addition and multiplication respectively in } \langle \mathfrak{N}_\mathsf{s}, R \rangle. \end{array}$$

Corollary 2.2

Let S^1_1 *denote the collection of all second-order* σ_{s}*-sentences of the form* $\forall Y^2\, \Psi$ *where* Y^2 *is a binary set variable, and* Ψ *contains no set quantifiers. Then*

$$\left\{\Phi \in \mathrm{S}^1_1 \mid \mathfrak{N} \vDash \Phi\right\}$$

is Π^1_1*-complete.*

Corollary 2.3

Let S^1_∞ *denote the collection of all second-order* σ_{s}*-sentences of the form* $\forall Y^2\, \Psi$ *where* Y^2 *is a binary set variable, and* Ψ *contains only unary set quantifiers. Then*

$$\left\{\Phi \in \mathrm{S}^1_\infty \mid \mathfrak{N} \vDash \Phi\right\}$$

is Π^1_∞*-complete.*

In third-order arithmetic we also have class variables

$$\mathcal{X},\ \mathcal{Y},\ \mathcal{Z},\ \ldots,$$

intended to range over the powerset of the powerset of $\mathbb{N}$. It would be more accurate to call these unary class variables, but we shall not deal with class variables of greater arities. Hence the atomic third-order formulas additionally include all expressions of the form $\mathcal{X}(X)$.

Corollary 2.4

Let S^2_1 denote the collection of all third-order σ_{s}-sentences of the form

$$\forall \mathcal{X}\, \forall Y^2\, \Psi$$

where $\mathcal{X}$ is a class variable, Y^2 is a binary set variable, and Ψ contains only unary set quantifiers and no class quantifiers. Then $\left\{\Phi \in \mathrm{S}^2_1 \mid \mathfrak{N} \models \Phi\right\}$ is Π^2_1-complete.

3. The case of structures of type 2

We write $\mathrm{Form}^{\circ}_{\varsigma}$ for the set of all quantifier-free first-order ς-formulas.

Fix a special individual variable $\underline{u}$. Call a regular $\mathscr{L}_1(\varsigma)$-formula flat if each of its basic subformulas is of the form

$$\mu_{\underline{u}}(\phi) \;=\; \mu_{\underline{u}}(\psi) \quad \text{or} \quad \mu_{\underline{u}}(\phi) \;\leqslant\; \mathfrak{a}$$

where ϕ and ψ belong to $\mathrm{Form}^{\circ}_{\varsigma}$, and $\mathfrak{a}$ is a field variable. Obviously, '$\mu_{\underline{u}}(\phi) \;\leqslant\; \mathfrak{a}$' must be omitted in the case of $\mathscr{L}_1^{\natural}$, i.e. if we exclude field variables.

Now Corollary 2.2 can be utilized to get:

Theorem 3.1

Let ς be $\langle P^2 \rangle$ where P is a binary predicate symbol. Then the validity problem for flat $\mathscr{L}_1^\natural(\varsigma)$-sentences is Π_1^1-hard, even if we confine ourselves to at most countable domains.

Proof. Consider an arbitrary $\mathscr{L}_1(\varsigma)$-structure $\mathcal{M} = \langle D, \pi, \mathsf{p} \rangle$. With each $d \in D$, associate the corresponding event

$$[\![d]\!] \;:=\; \{e \in D \mid \pi \models P(d, e)\}.$$

Denote by $\mathscr{D}$ the collection of all such events. If x is a variable distinct from $\underline{u}$, let us write $[x]$ for $P(x, \underline{u})$. Here $P(x, \underline{u})$ may be read as 'x satisfies P at $\underline{u}$'; so $\underline{u}$ is viewed as ranging over 'worlds'. Then the (flat) formula

$$x \approx y \;:=\; \mu\,(([x] \wedge \neg[y]) \vee ([y] \wedge \neg[x])) = 0$$

says 'the symmetric difference of $[\![x]\!]$ and $[\![y]\!]$ has measure zero'. For expository purposes, assume that $\mathsf{p}(d) > 0$ for all $d \in D$. While this restriction is not necessary, it will make some descriptions below simpler. Thus $x \approx y$ means that $[\![x]\!]$ equals $[\![y]\!]$. So

$$x \preccurlyeq y \;:=\; \mu([x] \wedge \neg[y]) = 0$$

says '$[\![x]\!]$ is a subset of $[\![y]\!]$'. For convenience, take

$$\underline{\mathscr{D}} \;:=\; \text{the closure of } \mathscr{D} \text{ under finite intersection and complementation.}$$

Naturally, it can be viewed as a Boolean algebra. Observe that the formula

$$\mathrm{At}(x) \;:=\; \mu([x]) \neq 0 \wedge \forall y\,(\mu([x] \wedge [y]) \neq 0 \rightarrow \mu([x] \wedge [y]) = \mu([x]))$$

holds iff $[\![x]\!]$ is an atom of $\underline{\mathscr{D}}$, i.e. a minimal non-empty event in $\underline{\mathscr{D}}$.

We shall also need the following formulas:

$$\begin{aligned}
\mathrm{Disj}_2\,(x,y) \;&:=\; \mu\,([x]\wedge[y]) = 0;\\
\mathrm{Disj}_3\,(x,y,z) \;&:=\; \mu\,([x]\wedge[y]) = \mu\,([x]\wedge[z]) = \mu\,([y]\wedge[z]) = 0;\\
\mathrm{DEq}_2\,(x,y) \;&:=\; \mathrm{Disj}_2\,(x,y)\wedge\mu\,([x]) = \mu\,([y]);\\
\mathrm{DEq}_3\,(x,y,z) \;&:=\; \mathrm{Disj}_3\,(x,y,z)\wedge\mu\,([x]) = \mu\,([y]) = \mu\,([z]);\\
\mathrm{Step}_2\,(x,y) \;&:=\; \exists y_1\,\exists y_2\,(\mathrm{DEq}_2\,(y_1,y_2)\;\wedge\\
&\qquad\qquad \mu\,([x]) = \mu\,([y_1]\vee[y_2])\wedge\mu\,([y]) = \mu\,([y_1]));\\
\mathrm{Step}_3\,(x,y) \;&:=\; \exists y_1\,\exists y_2\,\exists y_3\,(\mathrm{DEq}_3\,(y_1,y_2,y_3)\;\wedge\\
&\qquad\qquad \mu\,([x]) = \mu\,([y_1]\vee[y_2]\vee[y_3])\wedge\mu\,([y]) = \mu\,([y_1])).
\end{aligned}$$

Their meanings are clear. For technical reasons, suppose that $\mathcal{M}$ satisfies

$$\mathrm{Tech} \;:=\; \forall u\,(\mathrm{At}\,(u)\rightarrow\exists v\,(\mathrm{At}\,(v)\wedge\mathrm{Step}_2\,(u,v))\wedge\exists v\,(\mathrm{At}\,(v)\wedge\mathrm{Step}_3\,(u,v))).$$

With Tech in mind, the formula

$$\mathrm{Ind}_2(x) \ := \ \forall u\,(\mathrm{At}(u) \wedge u \preccurlyeq x \rightarrow \exists v\,(\mathrm{At}(v) \wedge v \preccurlyeq x \wedge \mathrm{Step}_2(u,v)))$$

holds iff for every atom $[\![u]\!]$ (of $\mathscr{D}$) below $[\![x]\!]$ there exists an atom $[\![v]\!]$ below $[\![x]\!]$ whose measure is two times smaller than that of $[\![u]\!]$. Then

$$\begin{aligned}
&\mathrm{Seq}_2(u,x) \ := \\
&\quad \mathrm{At}(u) \wedge u \preccurlyeq x \wedge \mathrm{Ind}_2(x) \ \wedge \\
&\quad \forall v_1\,\forall v_2\,(\mathrm{At}(v_1) \wedge \mathrm{At}(v_2) \wedge v_1 \preccurlyeq x \wedge v_2 \preccurlyeq x \rightarrow \neg\mathrm{DEq}_2(v_1,v_2)) \wedge \\
&\quad \forall v\,(\mathrm{At}(v) \wedge v \preccurlyeq x \wedge \mu([v]) \neq \mu([u]) \rightarrow \exists w\,(\mathrm{At}(w) \wedge w \preccurlyeq x \wedge \mathrm{Step}_2(w,v)))
\end{aligned}$$

means that $[\![u]\!]$ is an atom, and $[\![x]\!]$ is a minimal event above $[\![u]\!]$ satisfying $\mathrm{Ind}_2(x)$. Similarly, we can obtain $\mathrm{Ind}_3(x)$ and $\mathrm{Seq}_3(u,x)$ using $\mathrm{Step}_3(x,y)$, or $\mathrm{Ind}_6(x)$ and $\mathrm{Seq}_6(u,x)$ via

$$\mathrm{Step}_6(x,y) \ := \ \exists z\,(\mathrm{Step}_2(x,z) \wedge \mathrm{Step}_3(z,y)).$$

Finally, we need the formula

$$\begin{aligned}\mathrm{Base}(x_a, x_b, x_c) \;:=\; & \mathrm{Disj}_3(x_a, x_b, x_c) \wedge \mu([x_a] \vee [x_b] \vee [x_c]) = 1 \wedge \\ & \qquad\qquad \mu([x_a] \vee [x_b]) = \mu([x_c]) \wedge \mu([x_a]) = \mu([x_b]) \wedge \\ & \exists u\,(\mathrm{At}(u) \wedge \mathrm{Step}_2(x_a, u) \wedge u \preccurlyeq x_a) \wedge \mathrm{Ind}_2(x_a) \wedge \\ & \exists u\,(\mathrm{At}(u) \wedge \mathrm{Step}_2(x_b, u) \wedge u \preccurlyeq x_b) \wedge \mathrm{Ind}_2(x_b) \wedge \\ & \exists u\,(\mathrm{At}(u) \wedge \mathrm{Step}_3(x_c, u) \wedge u \preccurlyeq x_c) \wedge \mathrm{Ind}_2(x_c) \wedge \mathrm{Ind}_3(x_c)\,.\end{aligned}$$

It guarantees that:

- $[\![x_a]\!]$, $[\![x_b]\!]$ and $[\![x_c]\!]$ are pairwise disjoint;
- the measures of $[\![x_a]\!]$, $[\![x_b]\!]$ and $[\![x_c]\!]$ are equal to $1/4$, $1/4$ and $1/2$;
- $[\![x_a]\!]$ and $[\![x_b]\!]$ can be represented as

$$[\![x_a]\!] \;=\; \bigcup\nolimits_{i \in \mathbb{N}} [\![a_i]\!] \quad \text{and} \quad [\![x_b]\!] \;=\; \bigcup\nolimits_{i \in \mathbb{N}} [\![b_i]\!]$$

where each $[\![a_i]\!]$ and $[\![b_i]\!]$ is an atom and has measure $1/2^{i+3}$;

- $[\![x_c]\!]$ can be represented as

$$[\![x_c]\!] \;=\; \bigcup_{i,j\in\mathbb{N}} [\![c_{ij}]\!]$$

where each $[\![c_{ij}]\!]$ is an atom and has measure $1/\left(2^{i+1}\cdot 3^{j+1}\right)$.

Clearly, in case $\mathrm{Base}\,(x_a, x_b, x_c)$ holds, every atom has the form $[\![a_i]\!]$ or $[\![b_i]\!]$ or $[\![c_{ij}]\!]$, since

$$2\cdot\sum_{i\in\mathbb{N}}\frac{1}{2^{i+3}}+\sum_{i,j\in\mathbb{N}}\frac{1}{2^{i+1}\cdot 3^{j+1}} \;=\; \frac{1}{2}+\frac{1}{2} \;=\; 1.$$

Moreover, each of the $[\![c_{ij}]\!]$'s is uniquely determined by its measure. In particular, $[\![c_{00}]\!]$ can be captured by

$$\mathrm{Start}\,(x) \;:=\; \mathrm{At}\,(x)\wedge\exists y\,(\mu\,([y])=\mu\,(\neg[y])\wedge\mathrm{Step}_3\,(y,x)).$$

In fact, the atoms below $[\![x_a]\!]$ and $[\![x_b]\!]$ will play supporting roles. For instance, $\mathrm{Step}_3\,([\![c_{ij}]\!], [\![c_{ij+1}]\!])$ can be justified by finding $S \subseteq \mathbb{N}$ such that

$$\frac{1}{2^{i+1} \cdot 3^{j+2}} \;=\; \sum_{k\in S} \frac{1}{2^{k+3}}$$

and extending $\mathscr{D}$ to contain both $\bigcup_{k\in S}[\![a_i]\!]$ and $\bigcup_{k\in S}[\![b_i]\!]$. However, we shall be mainly concerned with $[\![x_c]\!]$, which will conveniently be viewed as an infinite matrix: for any $i, j \in \mathbb{N}$,

$$C_i \;:=\; \bigcup\{[\![c_{ij}]\!] \mid j \in \mathbb{N}\} \quad \text{and} \quad C^*_j \;:=\; \bigcup\{[\![c_{ij}]\!] \mid i \in \mathbb{N}\}$$

correspond to the ith row and jth column respectively; thus the diagonal is

$$E \;:=\; \bigcup\{[\![c_{ii}]\!] \mid i \in \mathbb{N}\}.$$

To make sure that all the rows, the columns and the diagonal belong to $\mathscr{D}$, one can add

$$\begin{aligned} \mathrm{Aux} \;:=\; & \exists u\, \exists y\, (\mathrm{Start}\,(u) \wedge \mathrm{Seq}_2\,(u, y) \wedge \forall v\, (\mathrm{At}\,(v) \wedge v \preccurlyeq y \rightarrow \exists z\, \mathrm{Seq}_3\,(v, z))) \wedge \\ & \exists u\, \exists y\, (\mathrm{Start}\,(u) \wedge \mathrm{Seq}_3\,(u, y) \wedge \forall v\, (\mathrm{At}\,(v) \wedge v \preccurlyeq y \rightarrow \exists z\, \mathrm{Seq}_2\,(v, z))) \wedge \\ & \exists u\, \exists y\, (\mathrm{Start}\,(u) \wedge \mathrm{Seq}_6\,(u, y)). \end{aligned}$$

which guarantees, in particular, that for some c_0, c_1, . . . and c_0^*, c_1^*, . . . ,

$$[\![c_0]\!] = C_0, \;\; [\![c_1]\!] = C_1, \;\; \ldots \quad \text{and} \quad [\![c_0^*]\!] = C_0^*, \;\; [\![c_1^*]\!] = C_1^*, \;\; \ldots$$

Thus we are going to deal with $\mathscr{L}_1\,(\varsigma)$-structures that satisfy the sentence

$$\mathrm{Req} \;:=\; \mathrm{Tech} \wedge \exists x_a\, \exists x_b\, \exists x_c\, \mathrm{Base}\,(x_a, x_b, x_c) \wedge \mathrm{Aux}.$$

It is straightforward to check that such structures do exist; we shall call them admissible. Further, for every $S \subseteq \mathbb{N}^2$ there exists an admissible $\mathcal{M}$ such that

$$\bigcup_{(i,j)\in S} [\![c_{ij}]\!] \in \mathscr{D}.$$

This will allow us to interpret a free binary predicate on the natural numbers.

Now consider the following formulas:

$$\begin{aligned}
\mathrm{Row}^0(x) &:= \exists u\,(\mathrm{Start}(u) \wedge \mathrm{Seq}_3(u,x));\\
\mathrm{Col}^0(x) &:= \exists u\,(\mathrm{Start}(u) \wedge \mathrm{Seq}_2(u,x));\\
\mathrm{Row}(x) &:= \exists y\,\exists u\,\big(\mathrm{Col}^0(y) \wedge \mathrm{At}(u) \wedge u \preccurlyeq y \wedge \mathrm{Seq}_3(u,x)\big);\\
\mathrm{Col}(x) &:= \exists y\,\exists u\,\big(\mathrm{Row}^0(y) \wedge \mathrm{At}(u) \wedge u \preccurlyeq y \wedge \mathrm{Seq}_2(u,x)\big);\\
\mathrm{Diag}(x) &:= \exists u\,(\mathrm{Start}(u) \wedge \mathrm{Seq}_6(u,x));\\
\mathrm{Match}(x,y) &:= \exists z\,(\mathrm{Diag}(z) \wedge \mu([x] \wedge [y] \wedge [z]) \neq 0).
\end{aligned}$$

Their meanings are clear. Note that $\mathrm{Match}\,(x,y)$ can be used to switch from rows to columns, and vice versa: if $\mathcal{M}$ is an admissible $\mathscr{L}_1\,(\varsigma)$-structure, then for any $i,j\in\mathbb{N}$,

$$\mathcal{M}\Vdash\mathrm{Match}\left(c_i,c_j^*\right)\quad\Longleftrightarrow\quad i\;=\;j.$$

Let us think of natural numbers as rows. Hence the successor function is captured by $\mathrm{Step}_2\,(x,y)$. To interpret a binary set variable, we introduce

$$\Gamma\,(x,y,z)\;:=\;\exists y^*\,(\mathrm{Col}\,(y^*)\wedge\mathrm{Match}\,(y,y^*)\wedge\mu\,([x]\wedge[y^*]\wedge[z])\neq 0).$$

To see how it works, observe that for every $S\subseteq\mathbb{N}^2$,

$$S\;=\;\left\{(i,j)\in\mathbb{N}^2\mid\mathcal{M}\Vdash\Gamma\,(c_i,c_j,s)\right\},$$

provided that $\mathcal{M}$ is admissible, $\bigcup_{(i,j)\in S}[\![c_{ij}]\!]$ belongs to $\mathscr{D}$ and equals $[\![s]\!]$. Thus elements of $\mathscr{D}$ may be treated as binary relations on $\mathbb{N}$.

We are ready to show the Π^1_1-hardness of the validity problem for flat $\mathscr{L}^{\natural}_1(\varsigma)$-sentences. Let Φ be a σ_{s}-sentence in S^1_1; so it has the form $\forall Y^2\, \Psi$ where Ψ contains no set variables. Without loss of generality, we may assume that:

- each atomic subformula of Ψ has the form

$$x = y \quad \text{or} \quad x = 0 \quad \text{or} \quad \mathsf{s}(x) = y \quad \text{or} \quad Y^2(x, y);$$

- $\vee$ and $\exists$ do not occur in Ψ, although $\wedge$, $\neg$ and $\forall$ may occur in it.

For convenience, the set variable Y^2 will also be treated as distinguished individual variable. Now define $\tau(\Psi)$ recursively:

$$
\begin{aligned}
\tau\left(x=y\right) &:= \mu\left([x]\right)=\mu\left([y]\right);\\
\tau\left(x=0\right) &:= \mathrm{Row}^{0}\left(x\right);\\
\tau\left(\mathsf{s}\left(x\right)=y\right) &:= \mathrm{Step}_{2}\left(x,y\right);\\
\tau\left(Y^{2}\left(x,y\right)\right) &:= \Gamma\left(x,y,Y^{2}\right);\\
\tau\left(\Theta\wedge\Xi\right) &:= \tau\left(\Theta\right)\wedge\tau\left(\Xi\right);\\
\tau\left(\neg\Theta\right) &:= \neg\tau\left(\Theta\right);\\
\tau\left(\forall x\,\Theta\right) &:= \forall x\left(\mathrm{Row}\left(x\right)\rightarrow\tau\left(\Theta\right)\right).
\end{aligned}
$$

By construction, $\tau\left(\Psi\right)$ is always flat. And it is straightforward to verify that

$$\mathfrak{N}\vDash\Phi \quad\Longleftrightarrow\quad \mathrm{Req}\rightarrow\forall Y^{2}\,\tau\left(\Psi\right)\ \text{is valid.}$$

Finally, apply Corollary 2.2. □

If we allow quantifiers over reals, then Corollary 2.3 can be used to obtain:

Theorem 3.2

Let ς be as before. Then the validity problem for flat $\mathscr{L}_1(\varsigma)$-sentences is Π^1_∞-hard, even if we confine ourselves to at most countable universes.

Proof. . . . □

In effect, Corollary 2.4 allows us to get a bit more:

Theorem 3.3

Let ς be as before. Then the validity problem for flat $\mathscr{L}_1(\varsigma)$-sentences is Π^2_1-hard.

Proof. . . . □

References

[1] M. Abadi, J. Y. Halpern. Decidability and expressiveness for first-order logics of probability. Information and Computation 112(1), 1–36, 1994.

[2] J. Y. Halpern. An analysis of first-order logics of probability. Artificial Intelligence 46(3), 311—350, 1990.

[3] D. N. Hoover. Probability logic. Annals of Mathematical Logic 14(3), 287–313, 1978.

[4] D. Ibeling, T. Icard, K. Mierzewski, M. Mossé. Probing the quantitative–qualitative divide in probabilistic reasoning. Annals of Pure and Applied Logic 175(9), 103339, 2024.

[5] H. J. Keisler. Probability quantifiers. In J. Barwise, S. Feferman (eds.), Model-Theoretic Logics, pp. 509–556. Springer, 1985.

[6] Z. Ognjanović, M. Rašković, Z. Marković. Probability Logics. Springer, 2016.

[7] Z. Ognjanović, A. Ilić-Stepić. Logics with probability operators. In Z. Ognjanović (ed.), Probabilistic Extensions of Various Logical Systems, pp. 1–35. Springer, 2020.

[8] S. O. Speranski. An 'elementary' perspective on reasoning about probability spaces. Logic Journal of the IGPL, jzae042, 2024. Published online.

[9] S. O. Speranski. Sharpening complexity results in quantified probability logic. Logic Journal of the IGPL, jzae114, 2024. Published online.

[10] S. O. Speranski. On the decision problem for quantified probability logics. Izvestiya: Mathematics 89(3), 10.4213/im9652, 2025.