Lecture Notes

Adding Probability to a Formal Logic

Assaf Kfoury

April 16, 2018 (last modified: April 19, 2018)

There are different ways of combining notions of probability with a formal logic – any formal logic: Not only any of the classical formal logics, such as *propositional logic* (PL) and *first-order logic* (FOL), but also any of the formal logics invented by computer scientists for their own needs and applications, such as LTL, CTL, other temporal and modal logics, versions of Hoare Logic, and many others.

An overview can be found on the Web (click here and here or see [5]) and the references therein. This handout illustrates some of the issues in relation to the simplest possible logic, standard *propositional logic* (PL), and again in relation to standard *first-order logic* (FOL). I assume prior knowledge of the basics of PL and FOL, for which there are many good references (*e.g.*, Chapters 1 and 2 in [12] or Chapters 1 and 2 in [7]). In other handouts I discuss issues related to the introduction of probability notions into logics specifically tailored to applications of computer science.

Our goal in this handout is to add probability notions to pre-existing, non-probabilistic formal logics, while preserving their technical apparatus as much as possible. Thus, probability is introduced by adding probability distributions on the domains to which these logics are applied ('variable x is assigned value a from the domain of interest with some probability $p \in [0,1]$'). Our goal is not to invent formal logics for reasoning *about* probabilities; in those latter logics, all formulas are either true or false (which is a subject of much interest in its own right and for other purposes in computer science). In our discussion to follow, formulas have probabilistic truth values, *i.e.*, true with some probability $p \in [0,1]$ and false with probability $(1-p)$.

1 Probability + Propositional Logic

Throughout, let `BVar` be a countable, finite or infinite, set of Boolean variables (or atoms), and let $\mathscr{L}(\mathrm{PL})$ be the language of standard PL, *i.e.*, all the well-formed formulas (WFF's) of standard propositional logic over the variables in `BVar`.

A *truth-value assignment* σ is a map from `BVar` to $\mathbb{B} \triangleq \{\mathtt{true}, \mathtt{false}\}$, $\sigma : \mathtt{BVar} \to \mathbb{B}$. If σ satisfies the wff $\varphi \in \mathscr{L}(\mathrm{PL})$, we write $\sigma \models \varphi$. If $\sigma \models \varphi$ for every truth-value assignment σ, we write $\models \varphi$.

A *probability function* for $\mathscr{L}(\mathrm{PL})$ is a function $\mathrm{Pr} : \mathscr{L}(\mathrm{PL}) \to \mathbb{R}$ satisfying the following conditions:

Non-negativity:
$\mathrm{Pr}(\varphi) \geqslant 0$ for every $\varphi \in \mathscr{L}(\mathrm{PL})$.

Tautologies:
If $\models \varphi$, then $\mathrm{Pr}(\varphi) = 1$.

Finite additivity:
If $\models \neg(\varphi \wedge \psi)$, *i.e.*, there is no truth-value assignment σ such that both $\models \varphi$ and $\models \psi$, then $\mathrm{Pr}(\varphi \vee \psi) = \mathrm{Pr}(\varphi) + \mathrm{Pr}(\psi)$.

Informally, $\mathrm{Pr}(\varphi)$ represents the probability that the wff φ evaluates to `true` when we choose Pr as the probability function for $\mathscr{L}(\mathrm{PL})$.

In general, there are infinitely many probability functions for $\mathscr{L}(\mathrm{PL})$. Let *probabilistic propositional logic* (here abbreviated pPL) be standard PL augmented with probability functions for $\mathscr{L}(\mathrm{PL})$.

Before proceeding further, for those who have not had a course on probability theory, we clarify two related concepts that appear repeatedly in the sequel: *random variable* and *probability distribution*.

Terminology 1. A *random variable* is a function that assigns numerical values to the outcomes of a statistical experiment; *e.g.*, the number of heads obtained by flipping a coin three times. That is, a random variable is a function from the set of possible outcomes of a statistical experiment to the set of real numbers; *e.g.*, in the experiment of flipping the coin three times, the random variable maps each outcome to a value in the set $\{0, 1, 2, 3\}$.

A *probability distribution* is a function that maps each value in the range of a random variable to a probability; *e.g.*, the possible outcomes of flipping a coin three times are eight:

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT (H = heads, T = tails),

and if the coin is fair, then each outcome occurs with probability 1/8. The corresponding probability distribution maps every value in $\{0, 1, 2, 3\}$ to a probability in $[0, 1]$; here, it carries out the following mappings $\{0 \mapsto 1/8,\ 1 \mapsto 3/8,\ 2 \mapsto 3/8,\ 3 \mapsto 1/8\}$. □

We state several basic facts of pPL without their proofs, or leave the proofs as exercises.

Proposition 2. *Let* Pr *be a probability function for* $\mathscr{L}(\mathrm{PL})$. *We then have:*

1. Pr *is a probability distribution, i.e.,* $\mathrm{Pr}(\varphi) \in [0, 1]$ *for every* $\varphi \in \mathscr{L}(\mathrm{PL})$.
2. *For all wff's* $\varphi, \psi \in \mathscr{L}(\mathrm{PL})$, *if* $\models (\varphi \leftrightarrow \psi)$, *i.e.,* φ *and* ψ *are equivalent, then* $\mathrm{Pr}(\varphi) = \mathrm{Pr}(\psi)$.

Let $\mathcal{S}$ be the set of all truth-value assignments. A probability distribution over $\mathcal{S}$ is a function $\theta : \mathcal{S} \to [0, 1]$ which – often, not always[1] – satisfies the discrete additivity condition: $\sum_{\sigma \in \mathcal{S}} \theta(\sigma) = 1$.

[1] We have to be careful here. We can always make this additivity condition, $\sum_{\sigma \in \mathcal{S}} \theta(\sigma) = 1$, to hold if $\mathcal{S}$ is a finite

Proposition 3. *Assume for simplicity that* `BVar` *is finite and let* Pr *be a probability function for* $\mathscr{L}(\mathrm{PL})$*. Then* Pr *uniquely induces a probability distribution* θ *on the set* $\mathcal{S}$ *of truth-value assignments.*

Exercise 4. Prove Proposition 3.

Hint: Start with the simple case when `BVar` $= \{x_1\}$, then `BVar` $= \{x_1, x_2\}$, and then generalize to an arbitrary finite `BVar`. □

The two next exercises show the converse of Proposition 3: How to lift a probability distribution θ on $\mathcal{S}$ to a probability function for $\mathscr{L}(\mathrm{PL})$.

Exercise 5. Let `BVar` $= \{x, y, z\}$ and consider all wff's of $\mathscr{L}(\mathrm{PL})$ written over $\{x, y, z\}$. Since $\mathcal{S}$ is the set of all functions $\sigma : \texttt{BVar} \to \mathbb{B}$ and $\mathbb{B}$ is a two-element set, there is a total of 8 truth-value assignments in $\mathcal{S}$. For every $v \in X$, a probability distribution θ_v assigns a value $p \in [0, 1]$ to the event $\langle v \mapsto \texttt{true}\rangle$ and probability $(1 - p)$ to the event $\langle v \mapsto \texttt{false}\rangle$:[2]

$$\theta_v\langle v \mapsto \texttt{true}\rangle = p \quad \text{and} \quad \theta_v\langle v \mapsto \texttt{false}\rangle = (1 - p).$$

This induces a probability distribution θ over $\mathcal{S}$ defined by:

$$\theta\langle x \mapsto b_1, y \mapsto b_2, z \mapsto b_3\rangle = \theta_x\langle x \mapsto b_1\rangle \cdot \theta_y\langle y \mapsto b_2\rangle \cdot \theta_z\langle z \mapsto b_3\rangle$$

for every $\sigma = \langle b_1, b_2, b_3\rangle$ where, for convenience, we write $\langle b_1, b_2, b_3\rangle$ instead of $\langle x \mapsto b_1, y \mapsto b_2, z \mapsto b_3\rangle$. For simplicity in this exercise, let $p = 1/2$ for every $v \in \texttt{BVar}$, so that $\theta(\sigma) = (1/2)^3 = 1/8$ for every truth-value assignment $\sigma \in \mathcal{S}$.

Your task in this exercise is to show that the probability distribution θ over $\mathcal{S}$ induces, or can be lifted uniquely, to a probability function Pr for $\mathscr{L}(\mathrm{PL})$ satisfying the three conditions: **non-negativity**, **tautologies**, and **finite additivity**.

Hint: Proceed by structual induction on the syntax of wff's in $\mathscr{L}(\mathrm{PL})$. You can limit your answer to wff's written in terms of only three logical connectives $\{\neg, \wedge, \vee\}$ over the variables $\{x, y, z\}$. □

Exercise 6. This continues Exercise 5, now with different probability distributions on the variables in `BVar` $= \{x, y, z\}$, namely:

$$\begin{aligned}
&\theta_x\langle x \mapsto \texttt{true}\rangle = 1/2 \quad \text{and} \quad \theta_x\langle x \mapsto \texttt{false}\rangle = 1/2,\\
&\theta_y\langle y \mapsto \texttt{true}\rangle = 1/3 \quad \text{and} \quad \theta_y\langle y \mapsto \texttt{false}\rangle = 2/3,\\
&\theta_z\langle z \mapsto \texttt{true}\rangle = 1/4 \quad \text{and} \quad \theta_z\langle z \mapsto \texttt{false}\rangle = 3/4.
\end{aligned}$$

---

space, which is the case when `BVar` is finite. But this additivity condition may or may not hold if $\mathcal{S}$ is an infinite space, which is the case when `BVar` is infinite. In all cases when this additivity condition holds, whether $\mathcal{S}$ is finite or infinite, we say that θ is a *discrete* probability distribution.

When $\mathcal{S}$ is an infinite space and θ cannot satisfy the additivity condition above, we need to deal with θ as a *continuous* (not *discrete*) distribution (and consider an additivity condition of a different kind). As a continuous distribution over an infinite $\mathcal{S}$, we cannot express θ in tabular form. Instead, if we want to describe θ graphically, we use an equation, the so-called *probability density function* satisfying two conditions: (1) $\theta(\sigma) \geqslant 0$ for every $\sigma \in \mathcal{S}$, and (2) the total area under the curve of the function θ is equal to one. We can take (2) as replacing the additivity condition of a discrete distribution (though a different additivity condition still holds for the continuous case, see discussion after Example 13 on page 8).

I try to keep to a minimum concepts of probability theory that are required for reading this handout. For those who already know basic concepts of probability and may be puzzled by my skirting many of them, I purposely avoid any mention of *conditional probabilities*, *probability-mass functions* (associated with *discrete* random variables), and other notions, and try to minimize mention of *probability-density functions* (associated with *continuous* random variables) and other notions, hopefully making this handout accessible to those with little knowledge of probability theory. If you want to study this background on your own, there are many good references, for example [3, 4, 6], among several others.

[2] I am abusing the notation a little, but the intention should be clear. The event here corresponds to assigning a truth value to the variable v by flipping a coin: It returns `true` with probability p and `false` with probability $(1 - p)$. If $p = 1/2$, the coin is fair.

Repeat Exercise 5 with these probabilities. □

Exercise 7. Let $\{x, y\}$ be propositional variables, and $\Pr : \mathscr{L}(\mathrm{PL}) \to \mathbb{R}$ an arbitrary probability function for $\mathscr{L}(\mathrm{PL})$. Show that:

$$\Pr(y) = \Pr(x \vee y) + \Pr(x \to y) - 1$$

Hint 1: Try Exercises 5 and 6 before you try this exercise.

Hint 2: Start with the probability function $\Pr : \mathscr{L}(\mathrm{PL}) \to \mathbb{R}$ induced by the following probability distributions on the variables $\{x, y\}$:

$$\theta_x\langle x \mapsto \texttt{true}\rangle \;=\; \theta_y\langle y \mapsto \texttt{true}\rangle \;=\; 1/2$$

and consider the truth-table for $(x \vee y)$ and $(x \to y)$. Generalize to the case when $\theta_x\langle x \mapsto \texttt{true}\rangle = p$ and $\theta_y\langle y \mapsto \texttt{true}\rangle = q$ for arbitrary $p, q \in [0, 1]$. □

Exercise 8. This continues Exercise 7. Let $\texttt{BVar} = \{x_1, x_2, x_3, \ldots\}$, a countably infinite set of variables. Let $\Pr : \mathscr{L}(\mathrm{PL}) \to \mathbb{R}$ be the probability function for $\mathscr{L}(\mathrm{PL})$ induced by the following probability distributions on the variables in $\texttt{BVar}$:

$$\theta_n\langle x_n \mapsto \texttt{true}\rangle = 1/(1+n) \quad \text{and} \quad \theta_n\langle x_n \mapsto \texttt{false}\rangle = n/(1+n),$$

for every $n = 1, 2, 3, \ldots$. There are two parts in this exercise, both relative to the specific probability function Pr which is induced by the probability distributions $\{\, \theta_n \,|\, n \geqslant 1 \,\}$ defined above:

1. Show, for all $i, j \in \{1, 2, 3, \ldots\}$:

$$\Pr(x_j) = \Pr(x_i \vee x_j) + \Pr(x_i \to x_j) - 1$$

2. For every $n \geqslant 1$, define the wff $\varphi_n \triangleq \bigvee_{1 \leqslant i \leqslant n}(x_i \wedge x_{i+1})$. Determine the probability $\Pr(\varphi_n)$ as a function of n.

Hint: Try Exercises 7 before you try this exercise. □

In standard PL, if Γ is a set of wff's in $\mathscr{L}(\mathrm{PL})$, a judgment of the form $\Gamma \vdash \varphi$ is read as 'φ *is formally derivable or deducible from* Γ', which means that the wff φ can be formally derived or deduced from finitely many wff's in Γ as premises. This deduction is carried out using formal proof rules; the approach used in Chapter 1 of the book [12] is called *Natural Deduction*, but there are several alternative approaches for carrying out formal proofs of PL.

By *soundness* and *completeness*, $\Gamma \vdash \varphi$ holds iff $\Gamma \vDash \varphi$ holds, a basic fact shown in Section 1.4 of [12]. The assertion '$\Gamma \vDash \varphi$' is read as 'Γ *semantically entails* φ', which means that every truth-value assignment that makes all the wff's in Γ true will make φ true. The left-to-right implication is *soundness* of standard PL ('what is formally deducible is true'), and the right-to-left implication is its *completeness* ('what is true is formally deducible').

When we pass from PL to pPL, we need to adapt the notion of formal deduction '$\vdash$' and the notion of semantic entailment '$\vDash$'. There is more than one way of doing this. The simplest way of adapting '$\vDash$', which we now denote '$\vDash_p$', is defined by:

> We write $\Gamma \vDash_p \varphi$, and say Γ *probabilistically (semantically) entails* φ,
> iff for every probability function $\Pr : \mathscr{L}(\mathrm{PL}) \to \mathbb{R}$, if $\Pr(\psi) = 1$ for all $\psi \in \Gamma$, then $\Pr(\varphi) = 1$.

From this definition of '$\vDash_p$' the following proposition is a straightforward consequence.

Proposition 9. *For any set Γ of wff's and any wff φ in $\mathscr{L}(\mathrm{PL})$, we have $\Gamma \vDash \varphi$ iff $\Gamma \vDash_p \varphi$.*

Corollary 10. *The formal semantics of* pPL *defined by '$\vDash_p$' are sound and complete relative to the formal proof rules of standard* PL*, i.e., for any set Γ of wff's and any wff φ in $\mathscr{L}(\mathrm{PL})$, it holds that:*

$$\Gamma \vDash_p \varphi \quad \textit{iff} \quad \Gamma \vdash \varphi.$$

By the preceding corollary, there is no need to modify or extend the proof rules of standard PL in the presence of probability functions for $\mathscr{L}(\mathrm{PL})$, provided we use the notion of probabilistic satisfaction represented by '$\vDash_p$'.

However, this notion '$\vDash_p$' is unduly restrictive in many application areas; it requires that all the premises in Γ to have probability = 1 in order to conclude that φ has probability = 1. Here is an example of a meaningful but less restrictive notion of probabilistic semantic entailment: If all the premises in Γ have probability $\geqslant 1/2$ then the conclusion φ has probability $\geqslant 1/2$. Further discussion along these lines is found elsewhere (here for example).

We mention one way of defining a less restrictive notion of probabilistic semantic entailment, denoted '$\vDash_a$' (subscript 'a' is for 'Adams'), which is sound and complete relative to the formal proof rules of standard PL:

> We write $\Gamma \vDash_a \varphi$, and say Γ *probabilistically (semantically) entails φ in the sense of Adams*, iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every probability function Pr for $\mathscr{L}(\mathrm{PL})$, it holds that $\Pr(\psi) > 1 - \delta$ for every $\psi \in \Gamma$ implies $\Pr(\varphi) > 1 - \varepsilon$.

We have the following pleasant result.

Proposition 11. *The formal semantics of* pPL *defined by '$\vDash_a$' are sound and complete relative to the formal proof rules of standard* PL*, i.e., for any set Γ of wff's and any wff φ in $\mathscr{L}(\mathrm{PL})$, it holds that:*

$$\Gamma \vDash_a \varphi \quad \textit{iff} \quad \Gamma \vdash \varphi.$$

Hence, for the probabilistic semantic entailment defined by '$\vDash_a$', the proof rules of standard PL suffice and we do not need to introduce new ones.

For a proof of Proposition 11, consult Section 2.2 in the entry 'Logic and Probability' in the *Stanford Encyclopedia of Philosophy* and the references to Adams therein (click here).

Exercise 12. Let Γ be an arbitrary set of wff's and φ an arbitrary wff of $\mathscr{L}(\mathrm{PL})$. There are two parts in this exercise:

1. Using the definitions of '$\vDash_p$' and '$\vDash_a$', show that if $\Gamma \vDash_p \varphi$ then $\Gamma \vDash_a \varphi$.
2. Does the converse (namely, if $\Gamma \vDash_a \varphi$ then $\Gamma \vDash_p \varphi$) hold? Justify your answer.

Hint for part 2: If your answer is 'no', you need to find a counter-example, *i.e.*, a particular Γ and a particular φ such that $\Gamma \vDash_a \varphi$ but $\Gamma \nvDash_p \varphi$. You can simplify your counter-example further by choosing Γ to be a single wff. So, find a single wff ψ and a single wff φ such that $\psi \vDash_a \varphi$ but $\psi \nvDash_p \varphi$. □

2 Probability + First-Order Logic

Just as it is for PL, there are also different ways of adding notions of probability to FOL. A discussion of these is found on the Web (here for example). We start with the simplest way of adding probability to FOL, and then point out some of its limitations. First, a reminder of the syntax of standard FOL:

$$
\begin{array}{llll}
\textit{terms}: & & & \\
t & ::= & c \mid x \mid f(t_1,\ldots,t_n) & (n \geqslant 1) \\
\textit{formulas}: & & & \\
\varphi & ::= & \texttt{true} \mid \texttt{false} \mid t_1 \doteq t_2 \mid P(t_1,\ldots,t_n) \mid & (\text{atomic formulas, } n \geqslant 0) \\
 & & \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \cdots & (\text{propositional connectives}) \\
 & & \forall x.\varphi \mid \exists x.\varphi & (\text{quantifiers})
\end{array}
$$

where the highlighted part, in the last line, is replaced by:

$$(\mathbb{P}x.\varphi) \geqslant q \qquad \text{for some rational } q \in [0,1],$$

to obtain our simplest version of pFOL. We call '$\mathbb{P}$' a *probability quantifier* or a *probabilistic quantifier*, and say that the variable x is *probabilistically quantified* in the wff $((\mathbb{P}x.\varphi) \geqslant q)$. Let $\mathscr{L}(\text{pFOL})$ be the set of all wff's of pFOL.

Informally now, assuming that variable x occurs free in φ, the formula $((\mathbb{P}x.\varphi) \geqslant q)$ should be understood as saying that: *the probability of selecting a value for x satisfying φ is $\geqslant q$*, or also *the subset of all elements (from the domain of interest) satisfying φ when assigned to x is $\geqslant q$.*[3] With this understanding, we introduce two abbreviations:

- $\big((\mathbb{P}x.\varphi) \leqslant q\big)$ abbreviates $\big((\mathbb{P}x.\neg\varphi) \geqslant 1 - q\big)$, and
- $\big((\mathbb{P}x.\varphi) = q\big)$ abbreviates $\big((\mathbb{P}x.\varphi) \geqslant q\big) \wedge \big((\mathbb{P}x.\varphi) \leqslant q\big)$.

With the introduction of formulas of the form '$((\mathbb{P}x.\varphi) \geqslant q)$' we need to adjust the formal semantics of pFOL, which is now defined relative to a first-order structure (or model) $\mathcal{M} \triangleq (A, =, f^{\mathcal{M}}, P^{\mathcal{M}}, \ldots)$ where the domain A is a non-empty set, together with an assignment $\sigma : \texttt{Var} \to A$ and a probability distribution (discrete, for now) $\theta : A \to [0,1]$ over A, where $\texttt{Var}$ is the set of variables occurring in all the wff's of pFOL. We need the structure $\mathcal{M}$ for interpreting function symbols f and predicate symbols P occurring in wff's, the assignment σ for interpreting variables occurring free in wff's, and the probability distribution θ for interpreting probability quantifiers in wff's.

I define the triple $\mathcal{I}$ (for 'interpretation') by $\mathcal{I} \triangleq (\mathcal{M}, \sigma, \theta)$, and write $[\![t]\!]\,\mathcal{I}$ to denote the interpretation of the term t relative to $\mathcal{I}$, defined by induction on the syntax of terms:

$$
\begin{array}{ll}
[\![c]\!]\,\mathcal{I} & \triangleq c^{\mathcal{M}} \\
[\![x]\!]\,\mathcal{I} & \triangleq \sigma(x) \\
[\![f(t_1,\ldots,t_n]\!]\,\mathcal{I} & \triangleq f^{\mathcal{M}}([\![t_1]\!]\,\mathcal{I},\ldots,[\![t_n]\!]\,\mathcal{I})
\end{array}
$$

[3]While the conventions for writing the usual quantifiers '$\forall$' and '$\exists$' are solidly established, different authors have used different styles for writing probability quantifiers. Where I write '$((\mathbb{P}x.\varphi) \geqslant q)$' others write '$(Px \geqslant q).\varphi$' (in [13] for example), or '$\exists^{\geqslant q}.\varphi$' or '$(\exists^{\geqslant q})\,\varphi$' (in [14]), or '$w_x(\varphi) \geqslant q$' (in [11]) among others. Whenever convenient, I also write '$(\mathbb{P}^{\geqslant q}x.\varphi)$' instead of '$((\mathbb{P}x.\varphi) \geqslant q)$'.

By induction on the syntax of formulas, we define the notion of satisfaction '$\vDash_p$' relative to $\mathcal{I}$:[4]

$$\begin{array}{lll}
\mathcal{I} \vDash_p \texttt{true} & & \\
\mathcal{I} \nvDash_p \texttt{false} & & \\
\mathcal{I} \vDash_p t_1 \doteq t_n & \text{iff} & [\![t_1]\!]\mathcal{I} = [\![t_2]\!]\mathcal{I} \\
\mathcal{I} \vDash_p P(t_1,\dots,t_n) & \text{iff} & \big([\![t_1]\!]\mathcal{I},\dots,[\![t_n]\!]\mathcal{I}\big) \in P^{\mathcal{M}} \\
\mathcal{I} \vDash_p \neg\varphi & \text{iff} & \mathcal{I} \nvDash_p \varphi \\
\mathcal{I} \vDash_p \varphi_1 \wedge \varphi_2 & \text{iff} & \mathcal{I} \vDash_p \varphi_1 \text{ and } \mathcal{I} \vDash_p \varphi_2 \\
\mathcal{I} \vDash_p \cdots & \cdots & \text{(other propositional connectives)}
\end{array}$$

The preceding is straightforward and mimics the way the formal semantics of standard FOL is set up. Note that the probability distribution θ plays no role up to this point; the interpretation of terms and satisfaction of wff's are independent of θ so far.

The question is how to interpret the remaining wff's of the form $((\mathbb{P}x.\varphi) \geqslant q)$. Consider first the case when θ is a *discrete* probability distribution over the domain A of the structure $\mathcal{M}$.[5] As usual, we lift θ to a function on subsets of A by defining $\theta(X) \triangleq \sum_{a\in X}\theta(a)$ for any $X \subseteq A$, and define:[6]

$$(\dagger) \quad (\mathcal{M},\sigma,\theta) \vDash_p (\mathbb{P}x.\varphi) \geqslant q \quad \text{iff} \quad \theta\left(\left\{ a \in A \,\middle|\, (\mathcal{M},\sigma[x\mapsto a],\theta) \vDash_p \varphi \right\}\right) \geqslant q$$

There is a couple of subtle points in this definition. First, the symbol '$\geqslant$' on the left-hand side in $(\dagger)$ is just a symbol in the syntax of wff's, whereas the '$\geqslant$' on the right-hand side in $(\dagger)$ is the interpretation of the former, *i.e.*, the standard relation 'greater than or equal to' on numbers. We denote by $\sigma[x \mapsto a]$ the usual update of the assignment σ at x, *i.e.*,

$$\sigma[x\mapsto a](y) \triangleq \begin{cases} a & \text{if } x = y, \\ \sigma(y) & \text{otherwise.} \end{cases}$$

Hence, the right-hand side of of $(\dagger)$ collects the probabilities of all the elements $a \in A$ such that $(\mathcal{M},\sigma[x\mapsto a],\theta)\vDash_p\varphi$ and then compares their sum to q. This kind of summation cannot work when θ is not a discrete probability distribution – unless the collected elements form what is called a *measurable set* (more on this after Example 13 below).

As in standard FOL, if φ is a *closed* wff, *i.e.*, every variable in φ is bound by a probability quantifier '$\mathbb{P}$', then the assignment $\sigma : \texttt{Var} \to A$ in the interpretation $\mathcal{I} = (\mathcal{M},\sigma,\theta)$ plays no role in the satisfaction relation $\mathcal{I}\vDash_p\varphi$. In such a case, we write $\mathcal{M},\theta\vDash_p\varphi$ to mean that $\mathcal{M},\sigma,\theta\vDash_p\varphi$ for every $\sigma : \texttt{Var} \to A$.

In the same vein, we write $\mathcal{M},\sigma\vDash_p\varphi$ to mean that $\mathcal{M},\sigma,\theta\vDash_p\varphi$ for every $\theta : A \to [0,1]$, and $\mathcal{M}\vDash_p\varphi$ to mean that $\mathcal{M},\sigma,\theta\vDash_p\varphi$ for every $\sigma : \texttt{Var} \to A$ and every $\theta : A \to [0,1]$. Finally, we write $\vDash_p\varphi$ to mean that $\mathcal{M},\sigma,\theta\vDash_p\varphi$ for every $\mathcal{M}$, for every $\sigma : \texttt{Var} \to A$, and for every $\theta : A \to [0,1]$.

Before we consider the case when θ is a *continuous* probability distribution, we give a little example when θ is necessarily *discrete* because the domain is finite.

[4] This '$\vDash_p$' is unrelated to the '$\vDash_p$' in Section 1.

[5] See footnote 1 for a brief discussion between *discrete* distribution and *continuous* distribution.

[6] Note the similarities with the way the formal semantics of '$\forall$' and '$\exists$' are defined in standard FOL. For a first-order wff φ, model $\mathcal{M} = (A,\dots)$ with domain A, and assignment $\sigma : \texttt{Var} \to A$, we have:

$$\mathcal{M},\sigma \vDash \forall x.\,\varphi \;\text{ iff }\; \{\, a \in A \mid \mathcal{M},\sigma[x\mapsto a]\vDash\varphi \,\} = A \quad\text{and}\quad \mathcal{M},\sigma \vDash \exists x.\,\varphi \;\text{ iff }\; \{\, a \in A \mid \mathcal{M},\sigma[x\mapsto a]\vDash\varphi \,\} \neq \varnothing.$$

Example 13. Suppose the wff's of $\mathscr{L}(\text{pFOL})$ contain two unary predicate symbols, B and W, and are interpreted in structures of the form $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$ where A is a finite non-empty set. Because B and W are unary, their interpretations $B^{\mathcal{M}}$ and $W^{\mathcal{M}}$ in $\mathcal{M}$ can be viewed as subsets of A. Consider the wff φ_1 defined by:

$$\varphi_1 \triangleq \Big(\big(\mathbb{P}\, x.\, B(x) \leftrightarrow \neg W(x)\big) = 1 \Big)$$

φ_1 is a closed wff, so an assignment $\sigma : \texttt{Var} \to A$ plays no role in the satisfaction of φ_1 in a structure $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$. A moment of thought shows that, for every structure $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$:

$$\mathcal{M} \vDash_p \varphi_1 \ \text{ iff } \text{ the domain } A \text{ can be partitioned into } A = B^{\mathcal{M}} \cup W^{\mathcal{M}} \text{ with } B^{\mathcal{M}} \cap W^{\mathcal{M}} = \varnothing.$$

We think of φ_1 as enforcing the requirement that A is a set of black and white marbles, and no marble can be both black and white. Consider now another wff φ_2 defined by:

$$\varphi_2 \triangleq \Big(\big(\mathbb{P}\, x.\, B(x)\big) = 1/2 \Big)$$

Suppose we restrict attention to discrete probability distributions $\theta : A \to [0,1]$ that are *uniform*, *i.e.*, if $|A| = n$ then $\theta(a) = 1/n$ for every $a \in A$. In such a case, if $\mathcal{M} \vDash_p \varphi_1 \wedge \varphi_2$, then the domain A must be a finite collection of an equal number of black marbles and white marbles. □

When θ is a *continuous* probability distribution, we proceed differently and there is more than one way of doing it. I choose one way here, alternatives are proposed in [2, 8, 9, 10, 11], among others. The only place where there is a difference is in the semantics of wff's of the form $((\mathbb{P}x.\varphi) \geqslant q)$, now defined relative to $\mathcal{I} \triangleq (\mathcal{M}, \sigma, \theta)$ where – parts 1 and 2 below are as before, while part 3 is different because we use the (standard) definition of a probability distribution in general (whether *continuous* or *discrete*):

1. $\mathcal{M} \triangleq (A, =, f^{\mathcal{M}}, P^{\mathcal{M}}, \ldots)$ is a first-order structure (or model),
2. $\sigma : \texttt{Var} \to A$ is an assignment of values in A to variables in $\texttt{Var}$,
3. $\theta : \mathscr{A} \to [0,1]$ is a probability distribution satisfying the following conditions: [7]
   (a) $\mathscr{A} \subseteq 2^A$, *i.e.*, $\mathscr{A}$ is a collection of subsets of A, with $\varnothing \in \mathscr{A}$ and $A \in \mathscr{A}$,
   (b) $\mathscr{A}$ is closed under countable unions, countable intersections, and complementation.
   (c) $0 \leqslant \theta(X) \leqslant 1$ for every $X \in \mathscr{A}$, with $\theta(\varnothing) = 0$ and $\theta(A) = 1$,
   (d) if $\{X_1, X_2, X_3, \ldots\} \subseteq \mathscr{A}$ is a countable set of sets in $\mathscr{A}$ that are pairwise disjoint, then:

$$\theta\big(\bigcup_{i \geqslant 1} X_i\big) = \sum_{i \geqslant 1} \theta(X_i).$$

[7] If $\{X_1, X_2, X_3, \ldots\} \subseteq \mathscr{A}$ is a countable collection, finite or infinite, of sets in $\mathscr{A}$, their countable union is $(\bigcup_{i\geqslant 1} X_i) \in \mathscr{A}$, countable intersection is $(\bigcap_{i\geqslant 1} X_i) \in \mathscr{A}$, and complements are $(A - X_i) \in \mathscr{A}$ for every $i \geqslant 1$. Technically, conditions (3.a) and (3.b) define what is called a σ-algebra or a σ-field. (Warning: The 'σ' in 'σ-algebra' has nothing to do with our $\sigma : \texttt{Var} \to A$.) So, $\mathscr{A}$ is called a *σ-algebra* in A, the pair $(A, \mathscr{A})$ is called a *measurable space*, and the members of $\mathscr{A}$ are called *measurable sets*. If we add the probability distribution θ, the resulting triple $(A, \mathscr{A}, \theta)$ is called a *probability space*. (Our notation is somewhat different from the traditional one in probability theory, *e.g.*, click here: A probability space is often written as $(\Omega, \mathcal{F}, \mathbf{P})$ where Ω is called the *sample space*, the σ-algebra $\mathcal{F} \subseteq 2^\Omega$ is known as the *event space*, which is the set of all possible events $E \in \mathcal{F}$ that one can measure, and the measure $\mathbf{P}(E)$ is the probability of that event.)

Note that *discrete* distributions can be made to satisfy the same four conditions $\{$(3.a),(3.b),(3.c),(3.d)$\}$, which therefore cannot be used to distinguish *discrete* from *continuous*. Indeed, discrete distributions as defined in this handout can be viewed as a special case of distributions satisfying those conditions. Specifically, if $\theta : A \to [0,1]$ is a discrete distribution on A, we can take the measurable sets to be $\mathscr{A} = 2^A$, which is a special case of $\mathscr{A} \subseteq 2^A$, and then extend θ to a probability measure $\theta' : 2^A \to [0,1]$ satisfying the conditions $\{$(3.a),(3.b),(3.c),(3.d)$\}$, with $\theta'(B) \triangleq \sum_{b \in B} \theta(b)$ for every $B \in 2^A$.

The difference now is the result of replacing a discrete distribution θ, which we can take in the form $\theta : A \to [0,1]$, by a continuous distribution θ, which we must take in the form $\theta : \mathscr{A} \to [0,1]$. For the semantics of $((\mathbb{P}x.\varphi) \geqslant q)$, we repeat the earlier definition (†), but now identified differently with '(‡)' because of the added proviso:

$$(‡) \quad (\mathcal{M},\sigma,\theta) \vDash_p (\mathbb{P}x.\varphi) \geqslant q \quad \text{iff} \quad \theta\left(\{\, a \in A \,\middle|\, (\mathcal{M},\sigma[x \mapsto a],\theta) \vDash_p \varphi \,\}\right) \geqslant q,$$
$$\text{provided the set } \{\, a \in A \,|\, (\mathcal{M},\sigma[x \mapsto a],\theta) \vDash_p \varphi \,\} \text{ is in } \mathscr{A}.$$

The proviso guarantees that θ is defined on its argument, because θ is only defined on sets in $\mathscr{A}$.[8]

Example 14. As in Example 13, we consider wff's of $\mathscr{L}(\text{pFOL})$ containing the unary predicate symbols, B and W, and interpreted in structures of the form $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$. The domain A is finite or infinite. We define infinitely many wff's ψ_k, one for every $k \geqslant 1$:

$$\begin{aligned}
\psi_k \;\triangleq\;& \mathbb{P}^{\geqslant p_k}\; x_k.\;\; B(x_k) \;\;\wedge \\
&\Big(\mathbb{P}^{\geqslant p_{k-1}}\, x_{k-1}.\; B(x_{k-1}) \wedge \\
&\Big(\mathbb{P}^{\geqslant p_{k-2}}\, x_{k-2}.\; B(x_{k-2}) \wedge \\
&\Big(\qquad \cdots \\
&\Big(\mathbb{P}^{\geqslant p_2}\;\; x_2.\;\; B(x_2) \;\;\wedge \\
&\Big(\mathbb{P}^{\geqslant p_1}\;\; x_1.\;\; B(x_1) \;\;\wedge \bigwedge_{1 \leqslant i<j \leqslant k} \neg(x_i \doteq x_j)\,\Big)\Big)\cdots\Big)\Big)\Big)
\end{aligned}$$

where we write, for succinctness, $(\mathbb{P}^{\geqslant q}\, x.\,\varphi)$ instead $((\mathbb{P}\,x.\,\varphi) \geqslant q)$, and $\big(\bigwedge_{1 \leqslant i<j \leqslant k} \neg(x_i \doteq x_j)\big)$ is the conjunction of all wff's $\neg(x_i \doteq x_j)$ such that $1 \leqslant i < j \leqslant k$. Another collection of infinitely many wff's is $\{\psi'_k\}_{k \geqslant 1}$ where every ψ'_k is obtained from ψ_k by replacing every occurrence of 'B' by 'W' and every occurrence of 'p_i' by 'q_i'. In particular, the wff's ψ_1, ψ'_1, ψ_2 and ψ'_2 are:

$$\begin{aligned}
\psi_1 &\triangleq \mathbb{P}^{\geqslant p_1}\, x_1.\; B(x_1) \\
\psi'_1 &\triangleq \mathbb{P}^{\geqslant q_1}\, x_1.\; W(x_1) \\
\psi_2 &\triangleq \mathbb{P}^{\geqslant p_2}\, x_2.\; B(x_2) \wedge \Big(\mathbb{P}^{\geqslant p_1}\, x_1.\; B(x_1) \wedge \neg(x_1 \doteq x_2)\Big) \\
\psi'_2 &\triangleq \mathbb{P}^{\geqslant q_2}\, x_2.\; W(x_2) \wedge \Big(\mathbb{P}^{\geqslant q_1}\, x_1.\; W(x_1) \wedge \neg(x_1 \doteq x_2)\Big)
\end{aligned}$$

In addition to $\{\psi_k\}_{k \geqslant 1}$ and $\{\psi'_k\}_{k \geqslant 1}$, we consider two other wff's, φ and φ':

$$\varphi \;\triangleq\; \mathbb{P}^{\geqslant 1}\, x.\big(B(x) \vee W(x)\big) \quad \text{and} \quad \varphi' \;\triangleq\; \mathbb{P}^{\geqslant 1}\, x.\neg\big(B(x) \wedge W(x)\big)$$

All the wff's in $\{\psi_k\}_{k \geqslant 1} \cup \{\psi'_k\}_{k \geqslant 1} \cup \{\varphi, \varphi'\}$ are closed. Hence, when any of them, say φ, is interpreted relative to $\mathcal{I} \triangleq (\mathcal{M},\sigma,\theta)$, we can dispense with the assignment σ and write $(\mathcal{M},\theta) \vDash_p \varphi$ instead of $(\mathcal{M},\sigma,\theta) \vDash_p \varphi$. And if we write $\mathcal{M} \vDash_p \varphi$, it means $(\mathcal{M},\theta) \vDash_p \varphi$ for every probability distribution θ. We continue this example in Exercises 15 and 16 below. □

[8] As a follow-up to footnote 7, the proviso in (‡) requires that *definable sets* be *measurable* in the probability space $(A, \mathscr{A}, \theta)$.

Exercise 15. This is a continuation of Example 14. There are several parts:

1. For every structure $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$, show that $\mathcal{M} \vDash_p \varphi \wedge \varphi'$ iff the domain A is partitioned into two disjoint set by $B^{\mathcal{M}}$ and $W^{\mathcal{M}}$, *i.e.*,

$$\{a \in A \mid B^{\mathcal{M}}(a)\} \cap \{a \in A \mid W^{\mathcal{M}}(a)\} = \varnothing \quad \text{and} \quad \{a \in A \mid B^{\mathcal{M}}(a)\} \cup \{a \in A \mid W^{\mathcal{M}}(a)\} = A.$$

   You have to do this part for all cases, whether A is *finite* or *infinite*, and whether you consider *discrete* or *continuous* probability distributions.

2. Let $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$ be such that $\mathcal{M} \vDash_p \varphi \wedge \varphi'$ and assume the conclusion of part 1 holds. Let $p_1 = q_1 = p_2 = q_2 = 1/2$. There are two subparts, (a) and (b) in this part:

   (a) $\mathcal{M} \vDash_p \psi_1 \wedge \psi_1'$.

   (b) $\mathcal{M} \vDash_p \psi_1 \wedge \psi_1' \wedge \psi_2 \wedge \psi_2'$.

   For each of the two subparts, (a) and (b), what is the most precise statement you can make about the cardinality of the domain A? Can it be finite and even? Can it be finite and odd? Can it be infinite? Justify.

3. Repeat part 2 with its two subparts, now with probabilities: $p_1 = q_1 = 1/2$ and $p_2 = q_2 = 1/4$.

4. Let $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$ be such that $\mathcal{M} \vDash_p \varphi \wedge \varphi'$ and assume the conclusion of part 1 holds. Let $\mathcal{M} \vDash_p \psi_k \wedge \psi_k'$ for every $k \geqslant 1$ and $p_i = q_i = 1/2^i$ for every $i \geqslant 1$. What is the most precise statement you can make about the cardinality of the domain A? Can it be finite? Can it be infinite? Justify.

Hint: The four parts are in increasing difficulty. Do them in sequence. □

Exercise 16. This is a continuation of Example 14 and Exercise 15. Define infinitely many wff's $\widetilde{\psi}_k$, one for every $k \geqslant 1$:

$$\widetilde{\psi}_k \;\triangleq\; \mathbb{P}^{\geqslant p_k} x_k.\ \mathbb{P}^{\geqslant p_{k-1}} x_{k-1}.\ \cdots\ .\ \mathbb{P}^{\geqslant p_2} x_2.\ \mathbb{P}^{\geqslant p_1} x_1.\ B(x_k) \wedge \cdots \wedge B(x_1) \wedge \bigwedge_{1 \leqslant i < j \leqslant k} \neg(x_i \doteq x_j)$$

In words, $\widetilde{\psi}_k$ is obtained from the earlier ψ_k by moving probability quantifiers to outermost position on the left.[9] Clearly, $\widetilde{\psi}_1$ and ψ_1 are the same wff, but not $\widetilde{\psi}_k$ and ψ_k for $k \geqslant 2$. There are two parts:

1. Prove or disprove that $\widetilde{\psi}_2$ and ψ_2 are equivalent wff of pFOL.
2. Prove or disprove that $\widetilde{\psi}_k$ and ψ_k are equivalent wff of pFOL, for every $k \geqslant 2$.

Hint: If you know how to do part 1, part 2 is easy. □

It turns out that pFOL as presented so far is unduly restrictive. In particular, it does not allow for the selection of several objects from the domain in sequence or simultaneously, and independently of each other, as illustrated by the next example and exercise.

Example 17. This is a continuation of Example 13. Suppose we limit attention to uniform discrete distributions $\theta : A \to [0, 1]$. As pointed out earlier, if $\mathcal{M}$ is a finite structure such that $\mathcal{M} \vDash_p \varphi_1 \wedge \varphi_2$, then the domain A of $\mathcal{M}$ is a finite set (or bag) of an equal number, say $m \geqslant 1$, of black marbles and white marbles. The size of A is $|A| = 2m$.

Suppose we want to write a wff expressing the following action (with eyes closed!):

[9] We can also define infinitely many wff's $\widetilde{\psi}_k'$, one for every $k \geqslant 1$, where $\widetilde{\psi}_k'$ is obtained from $\widetilde{\psi}_k$ by replacing every occurrence of 'B' by 'W' and every occurrence of 'p_i' by 'q_i', but there is no need for them in this exercise.

- *pick a black marble, put it back in the bag, and then pick a white marble from the bag.*

We say this action succeeds if the first pick is indeed a black marble and the second pick is indeed a white marble. Hence, the expected probability that the action succeeds is $\geqslant (1/2)\cdot(1/2) = 1/4$, but not more. Is there a wff ψ describing this action that will be satisfied by $\mathcal{M}$, *i.e.*, such that $\mathcal{M} \vDash_p \psi$? One 'reasonable' attempt is to define ψ by:

$$\psi \triangleq \big(\mathbb{P}^{\geqslant 1/2}\, y.\, (\,\psi_1 \wedge \psi_2\,)\big) \quad \text{where } \psi_1 \triangleq \big(\mathbb{P}^{\geqslant 1/2}\, x.\, B(x)\big) \text{ and } \psi_2 \triangleq W(y),$$

where variables x and y correspond to picking a black marble and a white marble, respectively. Note that, out of the two probabilities occurring in our proposed ψ, the outer one is $1/2$, purposely contradicting our informal expectation that the action will succeed with probability only $\geqslant 1/4$. So, should we hope to show that $\mathcal{M} \nvDash_p \psi$? Unfortunately, our formal apparatus so far confirms just the opposite, namely, that $\mathcal{M} \vDash_p \psi$ does hold, as shown next.

Under the assumption that the bag A contains the same number m of black and white marbles, and that the distribution θ is uniform, it is easy to see that $\mathcal{M} \vDash_p \psi_1$. The wff $\psi_2 \triangleq W(y)$ contains free variable y; hence, for any assignment $\sigma : \mathtt{Var} \to A$, the set of elements for which ψ_2 is true is such that:

$$\theta\Big(\Big\{ a \in A \,\Big|\, (\mathcal{M}, \sigma[y \mapsto a], \theta) \vDash_p W(y) \Big\}\Big) \;=\; 1/2$$

under the assumption that θ is the uniform discrete distribution. Hence, under the same assumption:

$$\theta\Big(\Big\{ a \in A \,\Big|\, (\mathcal{M}, \sigma[y \mapsto a], \theta) \vDash_p \psi_1 \wedge \psi_2 \Big\}\Big) \;=\; 1/2$$

so that also $(\mathcal{M}, \sigma, \theta) \vDash_p \psi$ or, because is ψ is closed and we limit attention to uniform distributions, $\mathcal{M} \vDash_p \psi$, which contradicts our informal expectation. *Conclusion*: Our proposed wff ψ does not describe, or model, the action correctly; specifically, it does not reflect that the action succeeds with probability only $\geqslant 1/4$. □

Exercise 18. This is a follow-up to Examples 13 and 17. For a finite structure $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$ such that $\mathcal{M} \vDash_p \varphi_1 \wedge \varphi_2$, and for a uniform discrete distribution $\theta : A \to [0,1]$, show that in fact $\mathcal{M} \vDash_p \psi'$, where ψ' is obtained from ψ by replacing the two occurrences of '$\leqslant$' by '$=$'. That ψ' is satisfied in $\mathcal{M}$ is again in contradiction with the expected probability $1/4$ for the described action to succeed. □

Example 17 and follow-up Exercise 18 are not a *proof* that there is no wff in $\mathscr{L}\big(\text{pFOL}\big)$ that models the action described in Example 17 with its expected probability $= 1/4$.[10] Nevertheless, it points to a limitation of using probability quantifiers that bind a single variable at a time. We remedy this situation by allowing wff's of the form $((\mathbb{P}\vec{x}.\varphi) \geqslant q)$, where $\vec{x}$ is a finite non-empty sequence of variables, $\vec{x} = x_1 x_2 \cdots x_n$ with $n \geqslant 1$. In doing so, we extend the syntax of wff's; call $\mathscr{L}\big(\text{pFOL}'\big)$ the set of wff's so extended.

Example 19. Consider the action described in Example 17 with expected probability $= 1/4$. We can now define a wff $\psi^{(q)}$ of $\mathscr{L}\big(\text{pFOL}'\big)$, parametrized with a probability $q \in [0,1]$, to model it:

$$\psi^{(q)} \;\triangleq\; \mathbb{P}^{\geqslant q}\, x\, y.\, B(x) \wedge \neg B(y)$$

If $\mathcal{M} \triangleq (A, =, B^{\mathcal{M}}, W^{\mathcal{M}})$ is a structure satisfying $\varphi_1 \wedge \varphi_2$ as in Example 17, and θ is the uniform discrete probability on the domain A of $\mathcal{M}$, it is easy to check that $\mathcal{M}, \theta \vDash_p \psi^{(1/4)}$, while $\mathcal{M}, \theta \nvDash_p \psi^{(1/2)}$. □

[10] It is far more difficult to prove that *no* such wff in $\mathscr{L}\big(\text{pFOL}\big)$ exists.

Given a structure $\mathcal{M}$ with domain A, and a probability distribution θ on A, discrete or continuous, we need to extend θ to a probability distribution on elements of A^n (if θ is discrete) or on subsets of A^n (if θ is continuous), for every $n \geqslant 1$.

Exercise 20. There are two parts in relation to the extended probabilistic logic pFOL$'$:

1. Provide the details of the formal syntax of the wff's in $\mathscr{L}(\text{pFOL}')$.
2. Provide the details of the formal semantics of the wff's in $\mathscr{L}(\text{pFOL}')$.

Hint 1: Start with part 1, which is very easy.

Hint 2: For part 2, consider separately the two cases, *discrete* probability distributions and *continuous* probability distributions. For each of the two, you will have to take a product of probability distributions. The counterpart for $\mathscr{L}(\text{pFOL}')$ of (†) on page 7 is now:

$$(\dagger') \quad (\mathcal{M},\sigma,\theta) \vDash_p (\mathbb{P}\vec{x}.\varphi) \geqslant q \quad \text{iff} \quad \theta'\left(\{\,\vec{a} \in A^n \,\middle|\, (\mathcal{M},\sigma[\vec{x} \mapsto \vec{a}],\theta') \vDash_p \varphi\,\}\right) \;\geqslant\; q$$

where $n \geqslant 1$ is the number of variables in the sequence $\vec{x}$ and $\theta'(a_1,\ldots,a_n) \triangleq \theta(a_1) \times \cdots \times \theta(a_n)$, because for a discrete probability distribution 'the product of the distributions on the components is the distribution on the product of the components'. If θ is a continuous probability distribution, what is the counterpart (‡$'$) for $\mathscr{L}(\text{pFOL}')$ of (‡) on page 9? □

What Next? So far this section has presented only the semantics of pFOL and pFOL$'$. There are at least three major topics, all related to an axiomatization of pFOL$'$, which need still to be discussed to round up our presentation:

1. Axioms and proof rules for formal derivation of valid wff's in $\mathscr{L}(\text{pFOL}')$.
2. Soundness and completeness of these axioms and proof rules, w.r.t. the semantics in this section.
3. Complexity results (tractable, decidable, undecidable) of validity proofs in pFOL$'$.

Some coverage of these issues is found in [1, 11, 13] and the references therein.

References

[1] Martin Abadi and Joseph Y. Halpern. Decidability and Expressiveness for First-Order Logics of Probability. *Information and Computation*, 112(1):1 – 36, 1994.

[2] Fahiem Bacchus. *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach to Probabilities*. MIT Press, Cambridge, MA, USA, 1990.

[3] Patrick Billingsley. *Probability and Measure (Third Edition)*. Wiley-Interscience, 1995.

[4] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics*. Springer, 2005.

[5] Lorenz Demey, Barteld Kooi, and Joshua Sack. Logic and Probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition, 2017.

[6] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.

[7] Herbert B. Enderton. *A Mathematical Introduction to Logic (2nd Edition)*. Harcourt/Academic Press, 2001.

[8] Ronald Fagin and Joseph Y. Halpern. Uncertainty, Belief, and Probability. *Computational Intelligence*, 7(3):160–173, August 1991.

[9] Ronald Fagin and Joseph Y. Halpern. Reasoning About Knowledge and Probability. *Journal of the ACM*, 41(2):340–367, March 1994.

[10] Ronald Fagin, Joseph Y. Halpern, and Nimrod Megiddo. A Logic for Reasoning about Probabilities. *Information and Computation*, 87(1):78 – 128, 1990. Special Issue: Selections from 1988 IEEE Symposium on Logic in Computer Science.

[11] Joseph Y. Halpern. An Analysis of First-Order Logics of Probability. *Artificial Intelligence*, 46(3):311 – 350, 1990.

[12] Michael Huth and Mark Ryan. *Logic in Computer Science*. Cambridge University Press, 2011.

[13] H. Jerome Keisler. Probability Quantifiers. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, chapter XIV, pages 509–556. Springer-Verlag, Berlin, 1985.

[14] H. Jerome Keisler and Tawfik Boulos Lotfallah. Almost Everywhere Elimination of Probability Quantifiers. *Journal of Symbolic Logic*, 74(4):1121–1142, December 2009.