Reasoning with Probabilities
Basic Probability Logics

Joshua Sack

August 6, 2013

Probabilistic Propositional Logic

Probability language (with linear combinations)

Let AP be a set of proposition letters.

Propositional formulas:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi$$

Terms:

$$t ::= aP(\varphi) \mid t + t$$

Probability formulas (denote the set of these by $\mathcal{L}_{\mathrm{LC}}$):

$$f ::= t \geq a \mid \neg f \mid f \wedge f$$

where $p \in AP$ and $a \in \mathbb{Q}$.

Example: $2P(q) + 5P(r) \geq 1 \wedge P(q \wedge r) - P(q) + P(r) \geq 0$.

This language is from:

R. Fagin, J. Halpern, N. Megiddo. Reasoning about Probabilities. *Information and Computation* (1990).

Language without linear combinations

Let AP be a set of proposition letters.
Propositional formulas (denote the set of these by $\mathcal{L}_{\mathrm{PL}}(AP)$):

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi$$

Probability formulas (denote the set of these by $\mathcal{L}_{\mathrm{NC}}$):

$$f ::= P(\varphi) \geq a \mid \neg f \mid f \wedge f$$

where $p \in AP$ and $a \in \mathbb{Q}$.

Example: $P(q) \geq 1 \wedge \neg P(q \wedge r) \geq 0$.

Probability models and semantics

Let AP be a set of proposition letters.
$M = (X, \mathcal{A}, \mu, \|\cdot\|)$, where

- $(X, \mathcal{A}, \mu)$ is a probability space
- $\|\cdot\| : AP \to \mathcal{A}$

Define function $[\![\cdot]\!]$ from propositional formulas to $\mathcal{A}$:

$$
\begin{aligned}
[\![\top]\!] &= X \\
[\![p]\!] &= \|p\| \\
[\![\neg\varphi]\!] &= X - [\![\varphi]\!] \\
[\![\varphi \wedge \psi]\!] &= [\![\varphi]\!] \cap [\![\psi]\!]
\end{aligned}
$$

Note: $[\![\varphi]\!] \in \mathcal{A}$ for every φ.

Define relation $\models$ between models and probability formulas:

$$
\begin{aligned}
M \models a_1 P(\varphi_1) + \cdots + a_n P(\varphi_n) \geq r \text{ iff} \\
a_1 \mu([\![\varphi_1]\!]) + \cdots + a_n \mu([\![\varphi_n]\!]) \geq r.
\end{aligned}
$$

A note about σ-algebras

Here are two examples of measure spaces that are used.

- (discrete) $(X, \mathcal{A}, \mu)$, where
  - $\mathcal{A} = \mathcal{P}(X)$ (the power set of X)
  - μ is such that
    - $\{a \in X \mid \mu(\{a\}) > 0\}$ is countable, and
    - $\sum_{a \in X} \mu(\{a\}) = 1$

  In such cases, we often focus on the mass function of μ whose domain is X rather than the set function μ itself.
- (continuous) $(X, \mathcal{A}, \mu)$, where
  - $X = [0, 1]$,
  - $\mathcal{A}$ is the set of Lebesgue measurable subsets of $[0, 1]$,
  - μ is the uniform distribution.

  Recall from the discussion of Vitalli sets that $\mathcal{A}$ *cannot* be $\mathcal{P}(X)$ if we want μ to remain a uniform probability distribution.

Abbreviations

Let

$$\sum_{k=1}^{n} a_k P(\varphi_k) \equiv a_1 P(\varphi_1) + \cdots + a_n P(\varphi_n)$$

Then if $t = \sum_{k=1}^{n} a_k P(\varphi_k)$, let $bt = \sum_{k=1}^{n} b a_k P(\varphi_k)$

$$\begin{aligned} t \leq r &\equiv -t \geq -r \\ t = r &\equiv (t \leq r) \wedge (t \geq r) \\ t > r &\equiv \neg(t \leq r) \end{aligned} \qquad \begin{aligned} t_1 \geq t_2 &\equiv t_1 - t_2 \geq 0 \\ t_1 \leq t_2 &\equiv t_1 - t_2 \leq 0 \\ t_1 = t_2 &\equiv t_1 - t_2 = 0 \end{aligned}$$

Without linear combinations:

$$\begin{aligned} P(\varphi) \leq r &\equiv P(\neg\varphi) \geq 1 - r \\ P(\varphi) = r &\equiv (P(\varphi) \leq r) \wedge (P(\varphi) \geq r) \\ P(\varphi) > r &\equiv \neg(P(\varphi) \leq r) \end{aligned}$$

Expressing Finite Additivity

With linear combinations

If $\neg(\varphi \wedge \psi)$ is a tautology, then

$$P(\varphi) + P(\psi) = P(\varphi \vee \psi)$$

In general (for *any* φ and ψ),

$$P(\varphi \wedge \psi) + P(\varphi \wedge \neg\psi) = P(\varphi)$$

Without linear combinations

$$(P(\varphi \wedge \psi) = r \wedge P(\varphi \wedge \neg\psi) = s) \rightarrow P(\varphi) = r + s$$

For a given φ and ψ, expressing additivity without linear combinations as given above involves infinitely many formulas (ranging over r and s).

Expressivity of linear combinations

Theorem

The class $\mathcal{C}$ of probability models $(X, \mathcal{A}, \mu, \|\cdot\|)$, such that $\|p\| \geq \|q\|$, is definable (among all probability models) by a formula in $\mathcal{L}_{\mathrm{LC}}$, but not by any formula in $\mathcal{L}_{NC}$.

Proof idea:

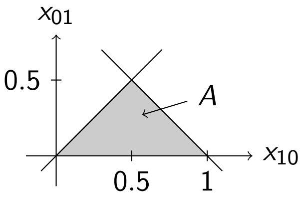
- Note $P(p) \geq P(q) \in \mathcal{L}_{\mathrm{LC}}$ characterizes $\mathcal{C}$.
- To show no such formula is in $\mathcal{L}_{\mathrm{NC}}$, focus on atoms:
  - Let $AP = \{p_1, \ldots, p_n\}$ s.t. $p = p_i$ and $q = p_j$ for some i, j
  - Let $At_{AP} = \{\bigwedge_{i=1}^{n} \ell(p) \mid \ell(p) \in \{p, \neg p\}\}_{\ell \in 2^{AP}}$
- $P(p) \geq P(q)$ is equivalent to $P(p \wedge \neg q) - P(\neg p \wedge q) \geq 0$ (here $AP = \{p, q\}$)

Visualizing the solution set

- Let $S = \mathbb{R}^4$, where each axis corresponds to the probability value of an atom in $At_{\{p,q\}}$.
- Denote these axis by x_{00}, x_{01}, x_{10}, and x_{11}.
- Identify $P(p \wedge \neg q) - P(\neg p \wedge q) \geq 0$ with the inequality $x_{10} - x_{01} \geq 0$ (setting $x_{10} = \mu([\![p \wedge \neg q]\!])$ etc).

Then the projection of the solution set of $x_{10} - x_{01} \geq 0$ in S onto the x_{10}-x_{01} plane is then the area A enclosed by the equations:

$$\begin{aligned} x_{10} - x_{01} &\geq 0, \\ x_{10} + x_{01} &\leq 1, \\ x_{01} &\geq 0, \end{aligned}$$



Lemma

Lemma

Suppose that $p, q \in AP$, $\varphi \in \mathcal{L}_{\mathrm{PL}}(AP)$ *is a propositional formula, and* $c \in 2^{At_{AP}}$ *is such that for each* $\chi \in At_{AP}$,

$$c_\chi = \begin{cases} 1 & \chi \to \varphi \text{ is a tautology} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\models P(\varphi) \geq r \leftrightarrow \sum_{\chi \in At_{AP}} c_\chi P(\chi) \geq r.$$

Remaining steps

Suppose toward a contradiction that $f \in \mathcal{L}_{NC}(AP)$ is such that $\models f \leftrightarrow P(p) \geq P(q)$.

- Place f into disjunctive normal form, and pick some disjunct d. Then

$$\models d \rightarrow P(p) \geq P(q).$$

- Let B be the set of values that x_{10} can attain given d.
- Let $\theta : B \rightarrow \mathbb{R}$ map each a to the supremum of the values that x_{10} can attain when $x_{01} = a$ given d.

Then θ must be non-increasing, as each constraint in d is

$$\sum_{\chi \in At_{AP}} c_\chi P(\chi) \geq r \text{ or } \sum_{\chi \in At_{AP}} c_\chi P(\chi) < r$$

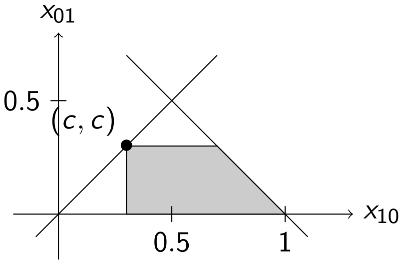
with $c_\chi \in \{0, 1\}$ (non-negative!)

Visualizing final steps

Let c be the infimum of values x_{10} can obtain given d. Then

$$\vDash d \rightarrow P(p \wedge \neg q) \geq c \wedge P(\neg p \wedge q) \leq c.$$

Thus the models that satisfy d must be contained in regions that we depict as follows:



No finite set of regions subject to such constraints has a union equal to A.

Proof system

- All propositional tautologies
- Equality: $P(\varphi) = P(\psi)$ whenever $\varphi \leftrightarrow \psi$ is a propositional tautology
- Kolmogorov axioms of probability:
  - $P(\varphi) \geq 0$
  - $P(\top) = 1$
  - $P(\varphi \wedge \psi) + P(\varphi \wedge \neg\psi) = P(\varphi)$
- Modus ponens: If $\vdash \varphi$ and $\vdash \varphi \rightarrow \psi$, then $\vdash \psi$.
- Inequality axioms (next slide)

Inequality axioms

- (permutation)
  $a_1 P(\varphi_1) + \cdots + a_n P(\varphi_n) \geq r \rightarrow$
  $a_{j_1} P(\varphi_{j_1}) + \cdots + a_{j_n} P(\varphi_{j_n}) \geq r$
- (adding coefficients)
  $(\sum_{k=1}^{n} a_k P(\varphi_k) \geq r) \wedge (\sum_{k=1}^{n} b_k P(\varphi_k) \geq s) \rightarrow$
  $(\sum_{k=1}^{n} (a_k + b_k) P(\varphi_k) \geq (r + s))$
- (adding and deleting 0 terms)
  $(t \geq r) \leftrightarrow (t + 0P(\varphi) \geq r)$
- (multiplying by non-zero coefficient)
  $t \geq r \leftrightarrow at \geq ar$ whenever $a > 0$.
- (dichotomy)
  $t \geq r \vee t \leq r$
- (monotonicity)
  $t \geq r \rightarrow t > s$, whenever $r > s$.

Lemma for Completeness

- $AP = \{p_1, \dots, p_n\}$ is a set of proposition letters,
- $At(AP) = \{\bigwedge_{i=1}^{n} q_i \mid q_i \in \{p_i, \neg p_i\}\}$ is set of atoms.

Lemma

Let $t \geq r$ be a probability formula, and AP a set of proposition letters containing all letters occurring in t. Let $At(AP) = \{\alpha_1, \dots, \alpha_{2^n}\}$. Then there are rationals $a_1, \dots, a_{2^n}$ such that $t \geq r$ is equivalent to $a_1 P(\alpha_1) + \cdots + a_{2^n} P(\alpha_{2^n}) \geq r$.

Let $At(AP, \varphi) = \{\alpha \in At(AP) \mid \vdash \alpha \to \varphi\}$. Then

$$P(\varphi) \equiv \sum_{\alpha \in At(AP,\varphi)} P(\varphi \wedge \alpha) \equiv \sum_{\alpha \in At(AP,\varphi)} P(\alpha).$$

The first equivalence comes from multiple applications of additivity proposition letter by proposition letter.

Completeness of Halpern's Probability Logic

Let f be a probability formula. It is a Boolean combination of atomic probability formulas.

- Transform f into disjunctive normal form: a disjunction of conjunctions of probability formulas.
- Consider a disjunct

$$g = (t_1 \geq r_1) \wedge \cdots \wedge (t_k \geq r_k)$$
$$\wedge \neg(t_{k+1} \geq r_{k+1}) \wedge \cdots \wedge \neg(t_m \geq r_m).$$

- Let $AP = \{p_1, \ldots, p_n\}$ be the set of proposition letters occurring in g
- Let $At = \{\delta_1, \ldots, \delta_{2^n}\}$ be the set of all atoms: conjunctions of n literals from AP
- Each conjunct $t_i \geq r_i$ of g is equivalent to $a_{i,1}P(\delta_1) + \cdots + a_{i,2^n}P(\delta_{2^n}) \geq r_i$

System of inequalities

The disjunct g is equivalent to the following system of inequalities:

$$
\begin{array}{|rcrcl|}
\hline
a_{1,1}P(\delta_1)+ & \cdots & +a_{1,2^n}P(\delta_{2^n}) & \geq & r_1 \\
 & & & \vdots & \\
a_{k,1}P(\delta_1)+ & \cdots & +a_{k,2^n}P(\delta_{2^n}) & \geq & r_k \\
\hline
a_{k+1,1}P(\delta_1)+ & \cdots & +a_{k+1,2^n}P(\delta_{2^n}) & < & r_{k+1} \\
 & & & \vdots & \\
a_{m,1}P(\delta_1)+ & \cdots & +a_{m,2^n}P(\delta_{2^n}) & < & r_m \\
\hline
P(\delta_1)+ & \cdots & +P(\delta_{2^n}) & \geq & 1 \\
-P(\delta_1)- & \cdots & -P(\delta_{2^n}) & \geq & -1 \\
\hline
 & & P(\delta_1) & \geq & 0 \\
 & & & \vdots & \\
 & & P(\delta_{2^n}) & \geq & 0 \\
\hline
\end{array}
$$

Final step

Completeness follows from the fact that the logic can follow the along with the steps of a mathematical algorithm (e.g. Fourier-Motzkin elimination) that checks whether a solution to the system of inequalities exists. If there were no solution, then the logic would prove false.

Small model theorem (towards complexity)

Given a probability formula f, let

- $|f|$ be its length (number of symbols).
- $||f||$ be length of the longest coefficient occurring in f

Theorem (Small model theorem)

If a probability formula f is satisfiable, then it is satisfiable in a model with the following properties

1. *there are at most $|f|$ states,*
2. *every set of states is measurable, and*
3. *the probability of each singleton is a rational number of size $O(|f|||f|| + |f|\log(|f|))$.*

Helpful lemma for small model theorem

From completeness, we have a model for f with at most 2^n (where n is the size of AP) states. We want to bound the model by the size of f (or the number of inequalities in f).

Lemma

If a system of r linear inequalities (or equalities) with integer coefficients each of length at most ℓ has a nonnegative solution, then it has a nonnegative solution with

- *at most r entries positive, and*
- *where the size of each number of the solution is $O(r\ell + r\log(r))$.*

Another lemma for the small model theorem

Lemma

Let f be a probability formula, and let $(X, \mathcal{A}, \mu)$ be a probability space and $\|\cdot\|_1$ and $\|\cdot\|_2$ valuation functions that agree on all atomic propositions occurring in f, then $(X, \mathcal{A}, \mu, \|\cdot\|_1) \models f$ iff $(X, \mathcal{A}, \mu, \|\cdot\|_2) \models f$

Here $\|\cdot\|_1$ and $\|\cdot\|_2$ can have different domains (but both containing the proposition letters in f). Thus f is satisfiable if and only if it is satisfiable in a model whose propositions are just those in f.

Satisfiability problem: NP-complete

- lower bound: probability logic satisfiability at least as hard as the boolean satisfiability problem (known to be NP complete): φ is satisfiable iff $P(\varphi) > 0$ is.
- upper bound: Non-deterministically select a small model. Then check (polynomial time):
  - for each expression $P(\varphi)$
    - determine $[\![\varphi]\!]$ by checking the truth at each state in the model
      (at most $|f|$ such expressions and $|f|$ states to check).
    - determine the probability value of $P(\varphi)$ by adding the probability values of each state in $[\![\varphi]\!]$.
      (each value has size $O(|f|||f|| + |f|\log(|f|))$ and at most $|f|$ states in $[\![\varphi]\!]$).
  - for each atomic probability formula $t \geq a$, perform the arithmetic to determine the truth value.
  - what remains is checking a given valuation (given by the truth of the atomic probability formulas) in a Boolean formula.

Modal Probability Logic

Simple Modal Probabilistic Language

Let AP be a set of proposition letters and I a set of labels. Modal Probability Formulas (denote the set of these by $\mathcal{L}_{\mathrm{MP}}$):

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid P_i(\varphi) \geq r$$

where $p \in AP$, $i \in I$, and $r \in \mathbb{Q}$.

Example: $P_i(P_k(q) \geq 0.5) \geq 1 \wedge \neg P_k(q \wedge r) \geq 0$.

Alternative notation that is often used:

- $L^i_r\varphi$ for $P_i(\varphi) \geq r$ and $M^i_r\varphi$ for $P_i(\varphi) \leq r$ (Suggested by Aumann 1995)
- $\langle i \rangle_r \varphi$ for $P_i(\varphi) \geq r$ (Larsen and Skou)

Models and semantics

Definition

Let AP be a set of proposition letters and I a set of labels. A *Probabilistic Modal Model* is $M = (X, \|\cdot\|, \{\mathbb{P}_i\}_{i\in I})$, where

- X is a set
- $\|\cdot\| : AP \to \mathcal{P}(X)$ is a valuation function
- $\mathbb{P}_i$ is a map from X to probability spaces $(S_{i,x}, \mathcal{A}_{i,x}, \mu_{i,x})$, such that $S_{i,x} \subseteq X$.

The semantics of formulas is defined by a function $[\![\cdot]\!]$ from formulas to subsets of X.

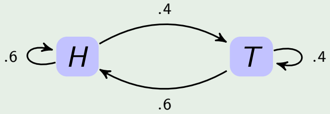
$$
\begin{aligned}
[\![\top]\!] &= X \\
[\![p]\!] &= \|p\| \\
[\![\neg\varphi]\!] &= X - [\![\varphi]\!] \\
[\![\varphi \wedge \psi]\!] &= [\![\varphi]\!] \cap [\![\psi]\!] \\
[\![P_i(\varphi) \geq r]\!] &= \{x \mid \mu^*_{i,x}([\![\varphi]\!] \cap S_{i,x}) \geq r\}
\end{aligned}
$$

Intuition about semantics

When X is finite and all $\mathcal{A}_{i,x} = \mathcal{P}(X)$, then depict a probability function as a directed graph labelled with probabilities:

Example

We represent the uncertainty of *one* agent about the result of flipping a weighted coin:

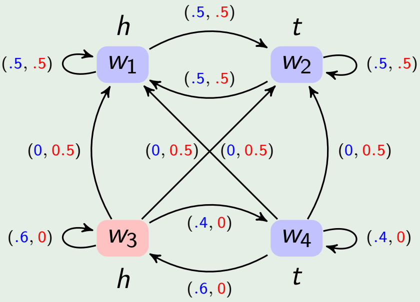


Notice that the sum of the numbers on arrows leaving a state is 1.

Multi-agent example

Example

Player 1 knows the coin is weighted, but player 2 does not:



$w_3 \models P_1(h) \geq .6 \wedge P_1(P_2(h) \geq .5) \geq 1.$
$w_3 \models L^1_{.6} h \wedge L^1{}_1 L^2{}_{.5} h.$

Ensuring measurability of formulas

A probabilistic modal model $(X, \|\cdot\|, \{\mathbb{P}_i\}_{i\in I})$ satisfies meas if there exists a sigma algebra $\mathcal{A} \subseteq \mathcal{P}(X)$ (intuitively $\mathcal{A}$ contains $[\![\varphi]\!]$ for all φ), such that the following conditions hold for each i.

- $\{A \cap S_{i,x} \mid A \in \mathcal{A}\} \subseteq \mathcal{A}_{i,x}$ (for each $x \in X$)
- $\mathbb{P}_i$ is a measurable function from $(X, \mathcal{A})$ to $(\text{spaces}(X), \mathcal{B})$, where
  - spaces(X) is the set of all probability spaces $(S, \mathcal{C}, \nu)$ such that $S \subseteq X$ and $\{A \cap S \mid A \in \mathcal{A}\} \subseteq \mathcal{C}$,
  - $\mathcal{B}$ is the σ-algebra generated from the set

$$\{(S, \mathcal{C}, \nu) \mid \sum_{k=1}^{n} a_k \nu(A_k \cap S) \geq r\}$$

    for each $n \geq 1$, $A_k \in \mathcal{A}$, and $a_k, r \in \mathbb{Q}$ $(1 \leq k \leq n)$
- $\|\cdot\| : AP \to \mathcal{A}$ (for the base case)

Harsanyi Types

Harsanyi Types are used in economics to model probabilities one player may have about the probabilities of others. They can be modeled using probabilistic modal models as follows

Definition

A *Harsanyi type model* is a probabilistic modal model $(X, \|\cdot\|, \{\mathbb{P}_i\}_{i\in I})$ that satisfy meas and where there is a σ-algebra $\mathcal{A}$ over X, such that for each x, $\mathbb{P}_{i,x} = (X, \mathcal{A}, \mu)$ for some probability measure μ.

Definition

The two components $(X, \{\mathbb{P}_i\})$ of a Harsanyi type model is called a *Hansanyi type space*

Proof system for Harsanyi models

Using Aumann's notation, but with only one agent:

- All propositional tautologies
- $L_0(\varphi)$, for all formulas φ
- $L_r(\top)$, for all $r \in \mathbb{Q} \cap [0,1]$
- $L_r\varphi \to \neg L_s \neg\varphi$, for $r + s > 1$
- $L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg\psi) \to L_{r+s}(\varphi)$, for $r + s \leq 1$
- $\neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \to \neg L_{r+s}(\varphi)$, for $r + s \leq 1$
- If $\vdash \varphi \leftrightarrow \psi$, then $\vdash L_r\varphi \leftrightarrow L_r\psi$
- If $\vdash \gamma \to L_s\varphi$ for all $s < r$, then $\vdash \gamma \to L_r\varphi$
- If $\vdash \varphi$ and $\vdash \varphi \to \psi$, then $\vdash \psi$.

This system is sound and weakly complete with respect to the one agent Harsanyi type models.

C. Zhou. A complete deductive system for probability logic. *Logic and Computation*. 2009

Computational interpretation

- Often discrete: $\mathcal{P}_i = (X, \mathcal{P}(X), \mu)$ is such that $\mu(\{x\}) > 0$ or countably many $x \in X$.
- Interpret I as a set of actions (not agents)

When X is finite, a discrete probabilistic modal model $(X, \|\cdot\|, \{\mathbb{P}_i\})$ is can be pictured as a labelled directed graph (relational structure) with

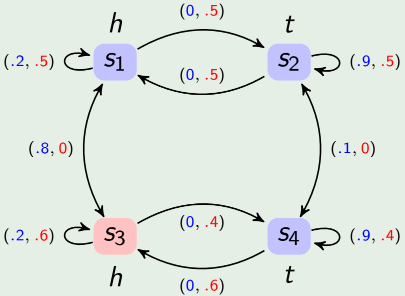
- nodes labelled by subsets of AP and
- relational connections labelled by pairs (i, r), where i is an action, and r is a probability value (the sum of the values of all arrows leaving a state x labeled with i is 1).

- Interpret $P_i(\varphi) \geq r$ to be "The probability that action i results in φ is at least r."

Computational example

Example

Action a can change the chance that action b results in the property h or t.



$s_3 \models P_a(P_b t \geq .5) \geq .8$
$s_3 \models \langle a \rangle_{.8} \langle b \rangle_{.5} t$

Bisimulation on probabilistic modal structures

Definition

Given a discrete probabilistic probabilistic model $M = (X, \|\cdot\|, \{\mathbb{P}_i\})$, a bisimulation on M is an equivalence relation R, such that whenever xRy, then for all labels $i \in I$, all equivalence classes $C \in X/R$, $\mu_{i,x}(C) = \mu_{i,y}(C)$.

A sight generalization of this for probabilistic transition systems where each $\mathbb{P}_i$ is a *partial function* is given in

- K. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, (1991).

Theorem (Adapted from Larsen and Skou Thm. 6.4)

Given a discrete probabilistic model $(X, \|\cdot\|, \{\mathbb{P}_i\})$, such that there exists an ϵ, such that for all $i \in I$ and $x, y \in X$, $\mu_{i,x}(y) = n\epsilon$ for some integer n. Then two states $x, y \in X$ are bisimilar if and only if x and y satisfy exactly the same formulas in $\mathcal{L}_{\mathrm{MP}}$.