Logic Programs with Annotated Disjunctions
Summary

- Logic Programs with Annotated Disjunctions
- Semantics of LPADs
- Expressiveness of LPADs
- LPADs compared to other formalisms
- Properties of LPADs
- Open problems in reasoning
- Definition of the learning problem
- Learning algorithm
- Open problems in learning
A disjunctive logic program consists of a set of formulas of the form

\[ h_1 \lor h_2 \lor \ldots \lor h_n \leftarrow b_1, b_2, \ldots, b_m \]

where \( h_i \) are logical atoms and \( b_i \) are logical literals.

Herbrand universe of \( P = H_U(P) \)
- if there are no function symbols \( H_U(P) \) is finite
- otherwise \( H_U(P) \) is countably infinite

A set \( S \) is countably infinite if there is a bijective function from \( S \) to \( \mathbb{N} \)

Herbrand base of \( P = H_B(P) \)
- if there are no function symbols \( H_B(P) \) is finite
- otherwise \( H_B(P) \) is countably infinite
Grounding

- A grounding of a clause is the clause where all the variables have been replaced by terms from $H_{U}(P)$

- Given a program $P$, its grounding $P'$ is obtained by replacing each clause with all its possible groundings

- If there are no function symbols, $P'$ is finite

- otherwise it is countably infinite
Interpretations

- **A Herbrand interpretation** is a subset of $H_B(P)$.
  - it can be finite or countably infinite
- The set of all Herbrand interpretations is $\mathcal{I}_P$.
- A disjunctive clause $C$ is **true in an interpretation** $I$ if for all grounding substitutions $\theta$ of $C$:
  $\text{body}(C)\theta \subset (I \cup \neg(H_B(P) \setminus I)) \rightarrow \text{head}(C)\theta \cap I \neq \emptyset$.
- The truth of a range restricted clause $C$ in a finite interpretation $I$ can be tested by running the query $\neg \text{body}(C), \text{not head}(C)$ on a database containing $I$. If the query succeeds $C$ is false in $I$. If the query fails $C$ is true in $I$ [De Raedt, Dehaspe, 1996].
Interpretations

Given

- a definite program $B$ where all the clauses are range restricted
- a finite interpretation $I$

The truth of a range restricted clause $C$ in $M(B \cup I)$ can be tested by running the query $\text{not head}(C), \text{body}(C)$ on a database containing $B$ and $I$. If the query succeeds, $C$ is false in $I$. If the query finitely fails, $C$ is true in $I$ [De Raedt, Dehaspe, 1996].
First Order Logics of Probability

- Semantics: definition of an admissible probability distribution over interpretations
- An admissible probability distribution \( \pi \) over interpretations is a function \( \mathcal{I}_P \rightarrow [0, 1] \) such that

\[
\sum_{I \in \mathcal{I}_P} \pi(I) = 1
\]
A Logic Program with Annotated Disjunctions (LPAD) consists of a set of formulas of the form

\[(h_1 : p_1) \lor (h_2 : p_2) \lor \ldots \lor (h_n : p_n) \leftarrow b_1, b_2, \ldots b_m\]

The \(h_i\) are logical atoms, the \(b_i\) are logical literals and the \(p_i\) are real numbers in the interval \([0, 1]\) such that \(\sum_{i=1}^{n} p_i = 1\).

\[\text{head}(C') = \{(h_i : p_i)|1 \leq i \leq n\} \quad \text{body}(C') = \{b_i|1 \leq i \leq m\}\]
Let $P$ be:

\[
C_1 = \text{heads}(\text{Coin}) : 0.5 \lor \text{tails}(\text{Coin}) : 0.5 \leftarrow \text{toss}(\text{Coin}), \neg \text{biased}(\text{Coin}).
\]

\[
C_2 = \text{heads}(\text{Coin}) : 0.6 \lor \text{tails}(\text{Coin}) : 0.4 \leftarrow \text{toss}(\text{Coin}), \text{biased}(\text{Coin}).
\]

\[
C_3 = \text{fair}(\text{Coin}) : 0.9 \lor \text{biased}(\text{Coin}) : 0.1.
\]

\[
C_4 = \text{toss}(\text{Coin}).
\]
Semantics of LPADs

\[ P' = \]

\[
\begin{align*}
\text{heads}(\text{coin}) : 0.5 \lor \text{tails}(\text{coin}) : 0.5 & \leftarrow \text{toss}(\text{coin}), \neg \text{biased}(\text{coin}). \\
n\text{heads}(\text{coin}) : 0.6 \lor \text{tails}(\text{coin}) : 0.4 & \leftarrow \text{toss}(\text{coin}), \text{biased}(\text{coin}). \\
n\text{fair}(\text{coin}) : 0.9 \lor \text{biased}(\text{coin}) : 0.1. \\
n\text{toss}(\text{coin}).
\end{align*}
\]
Selections and Instances

Let $P$ be a LPAD and $P'$ its grounding.

A selection $\sigma$ is a function which selects one pair $(h : \alpha)$ from each rule of $P'$, i.e. $\sigma : P' \rightarrow (H_B(P) \times [0, 1])$ such that, for each $R$ in $P'$, $\sigma(R) \in \text{head}(R)$.

For each rule $R$, we denote the atom $h$ selected from this rule by $\sigma_{\text{atom}}(R)$ and the probability $\alpha$ selected by $\sigma_{\text{prob}}(R)$.

We denote the set of all selections $\sigma$ by $S_P$.

The instance $P_\sigma$ chosen by $\sigma$ is obtained by keeping only the atom selected for $r$ in the head of each rule $r \in P'$, i.e. $P_\sigma = \{ \text{“} \sigma_{\text{atom}}(R) \leftarrow \text{body}(R) \text{”}| R \in P' \}$. 
Selections and Instances

\[ P' = \]

\[ C_1 = (heads(coin) : 0.5) \lor (tails(coin) : 0.5) \leftarrow toss(coin), \neg biased(coin). \]
\[ C_2 = (heads(coin) : 0.6) \lor (tails(coin) : 0.4) \leftarrow toss(coin), biased(coin). \]
\[ C_3 = (fair(coin) : 0.9) \lor (biased(coin) : 0.1). \]
\[ C_4 = toss(coin). \]

Selection \( \sigma \): \( \sigma(C_1) = (heads(coin), 0.5) \), \( \sigma(C_2) = (heads(coin), 0.6) \),
\( \sigma(C_3) = (fair(coin), 0.9) \), \( \sigma(C_4) = (toss(coin), 1) \),
\[ P_\sigma = \]

\[ heads(coin) \leftarrow toss(coin), \neg biased(coin). \]
\[ heads(coin) \leftarrow toss(coin), biased(coin). \]
\[ fair(coin). \]
\[ toss(coin). \]
Let $P$ be an LPAD. The probability of a selection $\sigma$ in $S_P^P$ is the product of the probability of the individual choices made by that selection, i.e.

$$C_\sigma = \prod_{R \in P} \sigma_{\text{prob}}(R)$$
Probability of a Selection and of an Instance

\[ P' = \]

\[ C_1 = (\text{heads}(\text{coin}) : 0.5) \lor (\text{tails}(\text{coin}) : 0.5) \leftarrow \text{toss}(\text{coin}), \neg \text{biased}(\text{coin}). \]
\[ C_2 = (\text{heads}(\text{coin}) : 0.6) \lor (\text{tails}(\text{coin}) : 0.4) \leftarrow \text{toss}(\text{coin}), \text{biased}(\text{coin}). \]
\[ C_3 = (\text{fair}(\text{coin}) : 0.9) \lor (\text{biased}(\text{coin}) : 0.1). \]
\[ C_4 = \text{toss}(\text{coin}). \]

\[ P_\sigma = \]
\[ \text{heads}(\text{coin}) \leftarrow \text{toss}(\text{coin}), \neg \text{biased}(\text{coin}). \]
\[ \text{heads}(\text{coin}) \leftarrow \text{toss}(\text{coin}), \text{biased}(\text{coin}). \]
\[ \text{fair}(\text{coin}). \]
\[ \text{toss}(\text{coin}). \]

Probability of the selection \( C_\sigma = 0.5 \cdot 0.6 \cdot 0.9 \cdot 1 = 0.27. \)
How large is $S_P$?

- If $P$ is function-free, then $P'$ is finite and $S_P$ is finite as well.
- If $P$ contains function symbols, then $P'$ is countably infinite and $S_P$ is uncountably infinite. Proof using Cantor’s diagonal argument.

$$p(s(X)) : a \lor q(s(X)) : b \leftarrow t(X)$$

1 → a a a a a a a . . .
2 → a a b a a b . . .
3 → a b b a a a b . . .

... b b a . . . is not present in the list.
Semantics of Instances

- Instances = normal logic programs
- Their semantics can be given by any of the semantics defined for normal logic programs (e.g. Clark’s completion, Fitting semantics, stable models, well founded semantics)
- We consider only the well founded semantics, the most skeptical one
  - If $P$ has a well-founded total model then that model is the unique stable model
  - The well founded partial model of $P$ is a subset of every stable model of $P$
Semantics of Instances

Since in LPAD the uncertainty is modeled by means of the annotated disjunctions, the instances of an LPAD should contain no uncertainty, i.e. they should have a single two-valued model (total model)

- stratified programs,
- acyclic programs,
- locally stratified programs,
- weakly stratified programs,
- modularly stratified programs,
- dynamically stratified programs
Local Stratification

A normal program $P$ is locally stratified if every ground atom can be assigned a countable ordinal rank such that, for any rule $C = h \leftarrow B$ in $P'$

- the rank of $h$ is higher than the rank of every atom in a negative literal in $B$
- the rank of $h$ is higher or equal than the rank of every atom in a positive literal in $B$

\[
\begin{array}{cccc}
  p \leftarrow \neg p & p \leftarrow \neg q & p \leftarrow q & p \leftarrow q \\
  q \leftarrow \neg p & q \leftarrow \neg p & q \leftarrow p & q \leftarrow p \\
  \text{n.l.s.} & \text{n.l.s.} & \text{n.l.s.} & \text{l.s.} \\
\end{array}
\]
Semantics of Instances

- Given an instance $P_\sigma$, its semantics is given by its well founded model $WFM(P_\sigma)$ and we require that it is two-valued:
  - An LPAD $P$ is called **sound** iff for each selection $\sigma$ in $S_P$, the well founded model $WFM(P_\sigma)$ of the program $P_\sigma$ chosen by $\sigma$ is two-valued.

- For example, if all the instances of an LPAD are locally stratified then the LPAD is sound.

- We denote with $P_\sigma \models_{WFM} F$ the fact that the formula $F$ is true in the interpretation that represents the well founded model of $P_\sigma$, i.e. $WFM(P_\sigma) \models F$. 
Local Stratification

An LPAD $P$ is locally stratified if every ground atom can be assigned a countable ordinal rank such that, for any rule $C = h_1 : p_1 \lor \ldots \lor h_n : p_n \leftarrow B$ in $P'$, for any $h_i$

- the rank of $h_i$ is higher than the rank of every atom in a negative literal in $B$
- the rank of $h_i$ is higher or equal than the rank of every atom in a positive literal in $B$

An LPAD is locally stratified $\iff$ all its instances are locally stratified

A locally stratified LPAD is sound
Let $P$ be a sound function-free LPAD.

For each of its interpretations $I$ in $\mathcal{L}_P$, the probability $\pi_P^*(I)$ assigned by $P$ to $I$ is the sum of the probabilities of all selections which lead to $I$, i.e. with $S(I)$ being the set of all selection $\sigma$ for which $WFM(P_\sigma) = I$:

$$\pi_P^*(I) = \sum_{WFM(P_\sigma) = I} C_\sigma = \sum_{\sigma \in S(I)} C_\sigma$$
Probability of an Interpretation

For example, consider the interpretation:

\[ I = \{\text{toss(coin)}, \text{fair(coin)}, \text{heads(coin)}\}. \]

\( I \) is the well founded model of two instance of the LPAD \( P \), one is the instance seen before:

\[
\text{heads(coin)} \leftarrow \text{toss(coin)}, \neg \text{biased(coin)}. \\
\text{heads(coin)} \leftarrow \text{toss(coin)}, \text{biased(coin)}. \\
\text{fair(coin)}. \\
\text{toss(coin)}. 
\]
Probability of an Interpretation

- \( I = \{toss(\text{coin}), \text{fair}(\text{coin}), \text{heads}(\text{coin})\} \)
- The other is the instance

\[
\begin{align*}
\text{heads}(\text{coin}) & \leftarrow toss(\text{coin}), \neg\text{biased}(\text{coin}). \\
\text{tails}(\text{coin}) & \leftarrow toss(\text{coin}), \text{biased}(\text{coin}). \\
\text{fair}(\text{coin}). \\
toss(\text{coin}).
\end{align*}
\]

- The probability of this instance is \(0.5 \cdot 0.4 \cdot 0.9 \cdot 1 = 0.18\).
- Therefore, the probability of the interpretation above is

\[
0.5 \cdot 0.4 \cdot 0.9 \cdot 1 + 0.5 \cdot 0.6 \cdot 0.9 \cdot 1 = 0.5 \cdot (0.4+0.6) \cdot 0.9 \cdot 1 = 0.45.
\]
Is $\pi_P^*$ an Admissible Prob. Distribution?

- Let $P$ be a sound function-free LPAD
- Let $S = \sum_{\sigma \in S_P} C_\sigma = \sum_{\sigma \in S_P} \prod_{R \in P} \sigma_{prob}(R)$
- $S = 1$
- Since $P$ is sound, for each $\sigma$ in $S_P$ then $WFM(P_\sigma)$ is in $\mathcal{I}_P$, thus it exists an $I \in \mathcal{I}_P$ such that $\sigma \in S(I)$
-Moreover $S(I) \cap S(J) = \emptyset$ for $I \neq J$, $I, J \in \mathcal{I}_P$ since $WFM(P_\sigma)$ is unique
- Thus $\{S(I) | I \in \mathcal{I}_P\}$ forms a partition of $S_P$
- $\sum_{I \in \mathcal{I}_P} \pi_P^*(I) = \sum_{I \in \mathcal{I}_P} \sum_{\sigma \in S(I)} C_\sigma =$
- $= \sum_{\sigma \in S_P} C_\sigma = S = 1$
- Thus $\pi_P^*$ is an admissible probability distribution
Proof of $S = 1$

- Let $P' = \{C_1, C_2, ..., C_m\}$
- Let $P_n = \{C_1, C_2, ..., C_n\}$ for $n = 1, \ldots, m$
- Let $S_{P_n}$ be the set of all selections for $P_n$
- Let $S_n = \sum_{\sigma \in S_{P_n}} C_{\sigma}$
- then $P_m = P'$, $S_{P_m} = S$ and $S_m = S$
- We will prove by induction that $S_n = 1 \ \forall n$

Case $n = 1$:
- Let $C_1$ be $h_1 : p_1 \lor h_2 : p_2 \lor \ldots \lor h_s : p_s \leftarrow B$
- Let $\sigma_i(C_1) = (h_i, p_i)$ $i = 1, \ldots, s$ then
  $S_{P_1} = \{\sigma_i|i = 1, \ldots, s\}$
- $C_{\sigma_i} = p_i$
- $S_1 = \sum_{i=1}^{s} p_i = 1$
Case \( n \): Let us suppose \( S_{n-1} = 1 \)

Let \( C_n \) be \( h_1 : p_1 \lor h_2 : p_2 \lor \ldots h_s : p_s \leftarrow B \).

Let \( S_{P_n}^i = \{ \sigma' | \sigma \in S_{P_{n-1}} \text{ and } \sigma' \text{ extends } \sigma \text{ by selecting } (h_i, p_i) \text{ on } C_n \} \)

Then \( S_{P_n}^i \cap S_{P_n}^j = \emptyset \) for \( i \neq j \) and \( S_{P_n} = \bigcup_{i=1}^{s} S_{P_n}^i \)

\[
S_n = \sum_{i=1}^{s} \sum_{\sigma \in S_{P_n}^i} \prod_{j=1}^{n} \sigma_{prob}(C_j) = \\
= \sum_{i=1}^{s} \sum_{\sigma \in S_{P_n}^i} \prod_{j=1}^{n-1} \sigma_{prob}(C_j) \sigma_{prob}(C_n) = \\
= \sum_{i=1}^{s} \sum_{\sigma \in S_{P_n}^i} \prod_{j=1}^{n-1} \sigma_{prob}(C_j) p_i = \\
= \sum_{i=1}^{s} p_i \sum_{\sigma \in S_{P_{n-1}}} \prod_{j=1}^{n-1} \sigma_{prob}(C_j) = \\
= \sum_{i=1}^{s} p_i S_{n-1} = \sum_{i=1}^{s} p_i 1 = 1
\]
Value of $S$ in the case of Function Symbols

- If $P$ has function symbols, what is the value of $S$?
- $P'$ is $\{C_1, C_2, \ldots, C_m, \ldots\}$
- $S = \sum_{\sigma \in S_P} \prod_{R \in P'} \sigma_{\text{prob}}(R)$
- The proof can be used as well to show that $S_n = 1$ for all $n$
- Two possibilities
  - Since $S = \lim_{n \to \infty} S_n$ then $S = 1$
  - Since $C_\sigma = \prod_{R \in P'} \sigma_{\text{prob}}(R) = 0$ for all $\sigma \in S_P$ then $S = 0$
Probability of a Formula

Let $P$ be a sound LPAD and let $\phi$ be a first order formula. The probability of $\phi$ is given by

$$\pi_P^*(\phi) = \sum_{I \in \mathcal{I}_P, I \models \phi} \pi_P^*(I)$$

For the program above $\pi_P^*$ is

- $I_1 = \{ \text{heads(coin), toss(coin), fair(coin)} \}$, $\pi_P^*(I_1) = 0.45$
- $I_2 = \{ \text{tails(coin), toss(coin), fair(coin)} \}$, $\pi_P^*(I_2) = 0.45$
- $I_3 = \{ \text{heads(coin), toss(coin), biased(coin)} \}$, $\pi_P^*(I_3) = 0.06$
- $I_4 = \{ \text{tails(coin), toss(coin), biased(coin)} \}$, $\pi_P^*(I_4) = 0.04$

The probability of $\text{heads(coin)}$ is

$$\pi_P^*(\text{heads(coin)}) = \pi_P^*(I_1) + \pi_P^*(I_3) = 0.45 + 0.06 = 0.51$$
Conditional Probability

Let $P$ be a sound LPAD and let $\phi$ and $\psi$ be two first order formulas. The conditional probability of $\phi$ given $\psi$ can be computed with Bayes theorem:

$$\pi^*_P(\phi|\psi) = \pi^*_P(\phi \land \psi)/\pi^*_P(\psi)$$

$I_1 = \{\text{heads(coin), toss(coin), fair(coin)}\}$ \hspace{1cm} $\pi^*_P(I_1) = 0.45$
$I_2 = \{\text{tails(coin), toss(coin), fair(coin)}\}$ \hspace{1cm} $\pi^*_P(I_2) = 0.45$
$I_3 = \{\text{heads(coin), toss(coin), biased(coin)}\}$ \hspace{1cm} $\pi^*_P(I_3) = 0.06$
$I_4 = \{\text{tails(coin), toss(coin), biased(coin)}\}$ \hspace{1cm} $\pi^*_P(I_4) = 0.04$

The conditional probability of $\text{heads(coin)}$ given $\text{toss(coin) \land biased(coin)}$ is

$$\pi^*_P(\text{heads(coin)}|\text{toss(coin) \land biased(coin)}) = \frac{\pi^*_P(\text{heads(coin) \land toss(coin) \land biased(coin)})}{\pi^*_P(\text{toss(coin) \land biased(coin)})} = \frac{0.06}{(0.06 + 0.04)} = 0.6$$
Properties of LPADs

Definition 1: Ground clauses \( H_1 \leftarrow B_1 \) and \( H_2 \leftarrow B_2 \) have mutually exclusive bodies over a set of interpretations \( J \) iff, \( \forall I \in J, B_1 \) and \( B_2 \) are not both true in \( I \).

Theorem 1: Consider the grounding \( P' \) of a locally stratified LPAD \( P \) and a clause \( C \in P' \) of the form 
\[
C = h_1 : p_1 \lor h_2 : p_2 \lor \ldots h_m : p_m \leftarrow B.
\]
Suppose you are given the function \( \pi^*_P \) and suppose that all the clauses of \( P' \) that share an atom in the head with \( C \) have mutually exclusive bodies with \( C \) over the set of interpretations \( J = \{ I | \pi^*_P(I) > 0 \} \). The probabilities \( p_i \) can be computed with the following formula:

\[
p_i = \frac{\sum_{I \in \mathcal{I}_P, I \models B, h_i} \pi^*_P(I)}{\sum_{I \in \mathcal{I}_P, I \models B} \pi^*_P(I)}
\]
Example

Given the following function $\pi^*_P$

$I_1 = \{\text{heads}(\text{coin}), \text{toss}(\text{coin}), \text{fair}(\text{coin})\}$ \hspace{1cm} $\pi^*_P(I_1) = 0.45$

$I_2 = \{\text{tails}(\text{coin}), \text{toss}(\text{coin}), \text{fair}(\text{coin})\}$ \hspace{1cm} $\pi^*_P(I_2) = 0.45$

$I_3 = \{\text{heads}(\text{coin}), \text{toss}(\text{coin}), \text{biased}(\text{coin})\}$ \hspace{1cm} $\pi^*_P(I_3) = 0.06$

$I_4 = \{\text{tails}(\text{coin}), \text{toss}(\text{coin}), \text{biased}(\text{coin})\}$ \hspace{1cm} $\pi^*_P(I_4) = 0.04$

The probabilities of clause $\text{heads}(\text{coin}) : p_1 \lor \text{tails}(\text{coin}) : p_2 \leftarrow \text{toss}(\text{coin}), \text{biased}(\text{coin})$. can be computed in the following way:

\[
p_1 = \frac{0.06}{0.06 + 0.04} = 0.6
\]

\[
p_2 = \frac{0.04}{0.06 + 0.04} = 0.4
\]
Observations

Consider the class of LPADs \( P \) such that all the couples of clauses of its grounding \( P' \) that share an atom in the head with have mutually exclusive bodies over the set of interpretations \( J = \{ I | \pi_P^*(I) > 0 \} \).

This class has the important property that, given a clause \( C \in P' \) of the form
\[
C = h_1 : p_1 \lor h_2 : p_2 \lor \ldots h_m : p_m \leftarrow B.
\]
\[p_i = \pi_P^*(h_i|B), \sum_{i=1}^{m} \pi_P^*(h_i) = 1, \text{ and } h_i \text{ and } h_j \text{ are mutually exclusive, i.e. not both true in any interpretation of } J\]

Intuitive reading, simpler writing

For generic LPADs this is not true
Example

\[ P = \]

\[
\begin{align*}
\text{heads} : 0.5 \lor \text{tails} : 0.5 & \leftarrow \text{fair}. \\
\text{heads} : 0.6 \lor \text{tails} : 0.4 & \leftarrow \text{square}. \\
\text{fair} : 0.1 \lor \text{biased} : 0.9. \\
\text{square} : 0.3 \lor \text{round} : 0.7.
\end{align*}
\]

\[
\begin{align*}
\pi_P^*(\text{heads}|\text{fair}) &= 0.59 \\
\pi_P^*(\text{tails}|\text{fair}) &= 0.56 \\
\pi_P^*(\text{heads}|\text{square}) &= 0.62 \\
\pi_P^*(\text{tails}|\text{square}) &= 0.43 \\
\pi_P^*(\{\text{heads, tails, fair, square}\}) &= 0.015
\end{align*}
\]
Example of an LPAD

Mendel’s inheritance mechanism for peas

\[\text{cg}(X, 1, A) : 0.5 \lor \text{cg}(X, 1, B) : 0.5 \quad \leftarrow \quad \text{father}(Y, X),\]
\[\quad \text{cg}(Y, 1, A), \text{cg}(Y, 2, B).\]
\[\text{cg}(X, 2, A) : 0.5 \lor \text{cg}(X, 2, B) : 0.5 \quad \leftarrow \quad \text{mother}(Y, X),\]
\[\quad \text{cg}(Y, 1, A), \text{cg}(Y, 2, B).\]
\[\text{color}(X, p) \quad \leftarrow \quad \text{cg}(X, A, p).\]
\[\text{color}(X, w) \quad \leftarrow \quad \text{cg}(X, 1, w), \text{cg}(X, 2, w).\]
Probability Theory

- $A$ and $B$ propositions
- $P(A)$ probability of $A$ or degree of belief in $A$
- $P(A|B)$ probability of $A$ given $B$ or degree of belief of $A$ given that I know $B$
- $P(A, B) ::= P(A \land B)$

Properties:
- $0 \leq P(A) \leq 1$
- $P($sure proposition$) = 1$
- $P(A \lor B) = P(A) + P(B)$ if $A$ and $B$ are mutually exclusive
- $P(A) = P(A, B) + P(A, \neg B)$
- $P(A|B) = P(A, B)/P(B)$ (Bayes theorem)
We consider propositions of the form “Attribute=value”

E.g. attribute $A$ has values $a_1, a_2, \ldots, a_n$, generic value of $A$: $a$

E.g. attribute $B$ has values $b_1, b_2, \ldots, b_m$, generic value of $B$: $b$

$P(a) ::= P(A = a)$

Properties:

$P(a_1 \lor \ldots \lor a_n) = P(a_1) + \ldots + P(a_n) = 1$

$P(a) = \sum_{i=1}^{m} P(a, b_i) = \sum_{b} P(a, b)$

$P(a|b) = P(a, b)/P(b)$ (Bayes theorem)
Joint Distribution Function

- Attributes=discrete random variables
- $U$ set of all the variables of the universe of discourse
- Let $U = \{ X_1, \ldots, X_n \}$
- Sets of variables: $X, Y, Z \subseteq U$, values $x, y, z$
- If $X = \{ X_i, X_j \}$, then $x = \{ x_i, x_j \}$
- Joint distribution function: $P(x_1, \ldots, x_n)$

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{\sum_{x_i: X_i \notin X, Y} P(x_1, \ldots, x_n)}{\sum_{x_i: X_i \notin Y} P(x_1, \ldots, x_n)}$$
\[ P(x_1, x_2, \ldots, x_n) = \]
\[ = P(x_1|x_2, \ldots, x_n)P(x_2, \ldots, x_n) = \]
\[ = P(x_1|x_2, \ldots, x_n)P(x_2|x_3, \ldots, x_n)P(x_3, \ldots, x_n) = \]
\[ = \ldots \]
\[ = P(x_1|x_2, \ldots, x_n)P(x_2|x_3, \ldots, x_n) \ldots P(x_{n-1}|x_n)P(x_n) \]
Independence

\[ X \text{ is conditionally independent from } Y \text{ given } Z \text{ iff } \]

\[ P(y, z) > 0 \rightarrow P(x|y, z) = P(x|z) \]

for all the possible configurations \( x, y, z \), i.e., iff

\[ P(x|y, z)P(y, z) = P(x|z)P(y, z) \]

for all the possible configurations \( x, y, z \).
Given node $A$, let $\Pi_A$ be the set of its parents

Given a directed acyclic graph $G$ and a probability distribution $P$, $G$ is a Bayesian network for $P$ iff every node $A$ is independent from its non-descendants given its parents $\Pi_A$, and no proper subset of $\Pi_A$ satisfies this condition.
If $G$ is a Bayesian network then there exist an ordering of the variables such that no variable is preceded by any of its ancestors.

If $G$ is a Bayesian network then there exist an ordering of the variables such that no variable is preceded by any of its ancestors.

$(I \mid p) \prod (l \mid q) \prod (\epsilon \mid p) \prod (\epsilon \mid l) \prod (l \mid u) \prod (l \mid u \mid \epsilon) \prod (l \mid u \mid l) \prod (l \mid u \mid \epsilon) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q) \prod (l \mid u \mid q)$

Bayesian Networks
Bayesian Networks

- In general

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \pi_{X_i}) \]

- I have to specify only \( P(x_i | \pi_{X_i}) \) rather than \( P(x_1, \ldots, x_n) \).

- If binary variables, \( k \) maximum parents then \( n \times 2^k \) parameters instead of \( 2^n \).
Bayesian Networks

Conditional probability tables (CPT):

<table>
<thead>
<tr>
<th>Neighbourhood</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>Earthquake</td>
</tr>
</tbody>
</table>

| Alarm | Report |

| N=good  | 0.3 |
| N=average | 0.4 |
| N=bad   | 0.3 |

<table>
<thead>
<tr>
<th>N =&gt;</th>
<th>g</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>B=true</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>B=false</td>
<td>0.9</td>
<td>0.8</td>
<td>0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B,E =&gt;</th>
<th>t,t</th>
<th>t,f</th>
<th>f,t</th>
<th>f,f</th>
</tr>
</thead>
<tbody>
<tr>
<td>A=true</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>A=false</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Reasoning with Bayesian Networks

- Compute $P(\text{event}|\text{evidence})$, i.e. $P(x|y)$ for all the values of $X$
- The problem is NP-hard
- Algorithms:
  - Variable elimination: $P(x|y) = P(x, y)/P(y)$ and then summation, careful choice of the ordering of variables to be summed up [Zhang & Poole 96, Dechter 98]
  - Belief propagation: Junction Trees [Lauritzen & Spiegelhalter 88], Conditioning [Pearl 86]
  - Stochastic simulation (approximate methods): [Pearl 97, MacKey 98]
Reasoning with Bayesian Networks

- Evidential reasoning or abduction or diagnosis: $P(\text{Burglary}|\text{Alarm})$
- Causal reasoning or prediction: $P(\text{Burglary}|\text{Neighbourhood})$
- Mixed reasoning: $P(\text{Burglary}|\text{Earthquake})$
Bayesian Networks vs LPADs

\[
\begin{align*}
\text{neigh}(X, g) & : 0.3 \lor \text{neigh}(X, a) : 0.4 \lor \text{neigh}(X, b) : 0.3 \\
\text{burg}(X, t) & : 0.1 \lor \text{burg}(X, f) : 0.9 \leftarrow \text{neigh}(X, \text{good}) \\
\text{burg}(X, t) & : 0.2 \lor \text{burg}(X, f) : 0.8 \leftarrow \text{neigh}(X, \text{average}) \\
\text{burg}(X, t) & : 0.4 \lor \text{burg}(X, f) : 0.6 \leftarrow \text{neigh}(X, \text{bad}) \\
\text{alarm}(X, t) & \leftarrow \text{burg}(X, t), \text{earthq}(X, t) \\
\text{alarm}(X, t) & : 0.8 \lor \text{alarm}(X, f) : 0.2 \leftarrow \text{burg}(X, t), \text{earthq}(X, f) \\
\text{alarm}(X, t) & : 0.8 \lor \text{alarm}(X, f) : 0.2 \leftarrow \text{burg}(X, f), \text{earthq}(X, t) \\
\text{alarm}(X, t) & : 0.1 \lor \text{alarm}(X, f) : 0.9 \leftarrow \text{burg}(X, f), \text{earthq}(X, f)
\end{align*}
\]
Bayesian Logic Programs

- [Kersting & De Raedt 00]
- Every ground atom $a$ with a predicate $p$ represents a random variable, which can take on a value from a domain $d_p = \{v_1^p, \ldots, v_{m_p}^p\}$

- A Bayesian Logic Program (BLP) is a set of clauses of the form
  
  $$h|b_1, \ldots, b_n$$

  where $h, b_1, \ldots, b_n$ are atoms

- Each clause must be range-restricted

- A CPT is associated with each clause

- Assumption: for each ground atom $a$, only one ground clause with $a$ in the head
Semantics of BLP

- Given by relating the BLP to a Bayesian network
- Each element of the Least Herbrand Model is a node in the network
- There is an edge from $h'$ to $b'$ iff the program contains one clause
  \[ h | \ldots b \ldots \]
  for which a substitution $\theta$ exists such that $h\theta = h'$ and $b\theta = b'$
- The CPT associated with $h'$ is the CPT of the only ground clause that has $h'$ in the head
- For the Bayesian network to be well defined, the program must be acyclic
Example of a BLP

Mendel’s laws for peas

\[
\begin{align*}
  cg(X, 1) & \mid father(Y, X), cg(Y, 1), cg(Y, 2) \\
  cg(X, 2) & \mid mother(Y, X), cg(Y, 1), cg(Y, 2) \\
  color(X) & \mid cg(X, 1), cg(X, 2)
\end{align*}
\]
**LPAD vs BLP**

- BLPs can be transformed into LPADs preserving the semantics
- Function free, acyclic LPADs can be transformed into BLPs preserving the semantics

The probability distribution defined by an LPAD representing a Bayesian network is the same as that defined by the network.
An Independent Choice Logic (ICL) program consists of two parts:

- A set $C$ of declarations of the following form

  $$\text{random}([a_1 : p_1, \ldots, a_n : p_n])$$

  with $a_i$ atoms, $p_i \in [0, 1]$ such that $\sum p_i = 1$

- A normal acyclic logic program $F$

Moreover,

- No atom $a_i$ appearing in $C$ may unify with another atom $a_j$ which also appears in $C$

- No head of a clause of $F$ may unify with an atom $a_i$ appearing in $C$
Independent Choice Logic

Example:

\[\text{goes\_to}(X, \text{football}) \leftarrow \text{man}(X), \text{man\_to\_f}(X)\]
\[\text{goes\_to}(X, \text{pub}) \leftarrow \text{man}(X), \text{man\_to\_p}(X)\]
\[\text{goes\_to}(X, \text{ballet}) \leftarrow \text{woman}(X), \text{woman\_to\_b}(X)\]
\[\text{goes\_to}(X, \text{pub}) \leftarrow \text{woman}(X), \text{woman\_to\_f}(X)\]

\[\text{random}([\text{man}(a) : 0.7, \text{woman}(a) : 0.3])\]
\[\text{random}([\text{man\_to\_f}(X) : 0.5, \text{man\_to\_p}(X) : 0.5])\]
\[\text{random}([\text{woman\_to\_b}(X) : 0.5, \text{woman\_to\_p}(X) : 0.5])\]
Independent Choice Logic

- An ICL program $P$ defines a probability distribution $\pi_P^{ICL}$ over interpretations and formulas.
- A meta interpreter exists that, given a program $P$ and a goal $G$, returns its probability.
- A transformation $\alpha$ exists such that every acyclic LPAD $P$ is transformed into an ICL $\alpha(P)$ such that the semantics is preserved.
- A meta interpreter for LPADs exists.
Metainterpreter

- Can handle non-ground formulas:
  
  ```
  :— explain(p(X),P).
  Yes, X=a, P=0.465;
  Yes, X=b, P=0.34
  ```

- Clauses must be disjoint: clauses with the same head must have mutually exclusive bodies

- Arbitrary formulas can be formed: & and, ; or, ~ not

- Prob(Interpretation)

- \( P(a|b) \):
  
  ```
  :— explain((a&b),P1),explain(b,P2),P is P1/P2.
  ```
A transformation $\beta$ exists such that every ICL $P$ is transformed into an LPAD $\beta(P)$ such that the semantics is preserved.
LPAD to ICL

\[ R = (h_1 : p_1) \lor (h_2 : p_2) \lor \ldots \lor (h_n : p_n) \leftarrow b_1, b_2, \ldots b_m \]

\[ h_1 \leftarrow b_1, b_2, \ldots b_m, choice_R(1) \]
\[ \ldots \]
\[ h_n \leftarrow b_1, b_2, \ldots b_m, choice_R(n) \]

\[ random([choice_R(1) : p_1, \ldots, choice_R(n) : p_n]) \]
Example of LPAD to ICL

\[
\begin{align*}
\text{heads}(\text{Coin}) & \leftarrow \text{toss}(\text{Coin}), \neg \text{biased}(\text{Coin}), \text{non}_{-}\text{bi}_{-}\text{heads}(\text{Coin}) \\
\text{tails}(\text{Coin}) & \leftarrow \text{toss}(\text{Coin}), \neg \text{biased}(\text{Coin}), \text{non}_{-}\text{bi}_{-}\text{tails}(\text{Coin}) \\
\text{heads}(\text{Coin}) & \leftarrow \text{toss}(\text{Coin}), \text{biased}(\text{Coin}), \text{biased}_{-}\text{heads}(\text{Coin}) \\
\text{tails}(\text{Coin}) & \leftarrow \text{toss}(\text{Coin}), \text{biased}(\text{Coin}), \text{biased}_{-}\text{tails}(\text{Coin}) \\
\text{toss}(\text{Coin}) & \\
\text{random}([\text{non}_{-}\text{bi}_{-}\text{heads}(\text{Coin}) : 0.5, \text{non}_{-}\text{bi}_{-}\text{tails}(\text{Coin}) : 0.5]) \\
\text{random}([\text{biased}_{-}\text{heads}(\text{Coin}) : 0.6, \text{biased}_{-}\text{tails}(\text{Coin}) : 0.4]) \\
\text{random}([\text{fair}(\text{Coin}) : 0.9, \text{biased}(\text{Coin}) : 0.1])
\end{align*}
\]
Example of ICL to LPAD

goes_to(X, football) ← man(X), man_to_f(X)
goes_to(X, pub) ← man(X), man_to_p(X)
goes_to(X, ballet) ← woman(X), woman_to_b(X)
goes_to(X, pub) ← woman(X), woman_to_f(X)

random([man(a) : 0.7, woman(a) : 0.3])
random([man_to_f(X) : 0.5, man_to_p(X) : 0.5])
random([woman_to_b(X) : 0.5, woman_to_p(X) : 0.5])


goes_to(X, football) : 0.5 ∨ goes_to(X, pub) : 0.5 ← man(X)
goes_to(X, ballet) : 0.5 ∨ goes_to(X, pub) : 0.5 ← woman(X)
man(a) : 0.7 ∨ woman(a) : 0.3
Research Issues

Is it possible to make the metainterpreter more efficient in computing $P(a|b)$?
Properties of LPADs

[Theorem 2: Consider an interpretation $I$ and a locally stratified LPAD $P$ with grounding $P'$ such that all the couples of clauses of $P'$ that share an atom in the head have mutually exclusive bodies with respect to the set of interpretations $\{I\}$. If $S(I) \neq \emptyset$ then all the selection $\sigma \in S(I)$ agree on the clauses of $P'$ with body true in $I$ and

$$\pi^*_P(I) = \prod_{R \in P', I \models body(R)} \sigma_{prob}(R)$$

where $\sigma$ is any element of $S(I)$. If $S(I) = \emptyset$ then $\pi^*_P(I) = 0$.}
Example

Consider the grounding $P'$ of program $P$ given before and the interpretation
$I = \{\text{toss(coin)}, \text{fair(coin)}, \text{heads(coin)}\}.

\begin{align*}
C_1 & = \text{heads(coin)} : 0.5 \lor \text{tails(coin)} : 0.5 \leftarrow \text{toss(coin)}, \neg \text{biased(coin)}. \\
C_2 & = \text{heads(coin)} : 0.6 \lor \text{tails(coin)} : 0.4 \leftarrow \text{toss(coin)}, \text{biased(coin)}. \\
C_3 & = \text{fair(coin)} : 0.9 \lor \text{biased(coin)} : 0.1. \\
C_4 & = \text{toss(coin)}. 
\end{align*}
In $P$, all the clauses that share an atom in the head have mutually exclusive bodies with respect to $\{I\}$.

Then all the selection $\sigma \in S(I)$ agree on the clauses with body true in $I$, namely $C_1$, $C_3$, and $C_4$.

In order to find the values of $\sigma(C_1)$, $\sigma(C_3)$ and $\sigma(C_4)$, we find a $P_\sigma$ such that $WFM(P_\sigma) = I = \{\text{toss}(\text{coin}), \text{fair}(\text{coin}), \text{heads}(\text{coin})\}$.

One such $P_\sigma$ is

\[
\begin{align*}
\text{heads}(\text{coin}) & \leftarrow \text{toss}(\text{coin}), \neg\text{biased}(\text{coin}). \\
\text{tails}(\text{coin}) & \leftarrow \text{toss}(\text{coin}), \text{biased}(\text{coin}). \\
\text{fair}(\text{coin}). \\
\text{toss}(\text{coin}).
\end{align*}
\]
Example

Therefore we have $\sigma(C_1) = (\text{heads}(\text{coin}) : 0.5)$, $\sigma(C_3) = (\text{fair}(\text{coin}) : 0.9)$ and $\sigma(C_4) = (\text{toss}(\text{coin}) : 1)$ and this is true $\forall \sigma \in S(I)$.

Then

$$\pi_P^*(I) = \sigma_{\text{prob}}(C_1) \cdot \sigma_{\text{prob}}(C_3) \cdot \sigma_{\text{prob}}(C_4) = 0.5 \cdot 0.9 \cdot 1 = 0.45$$

Before we had $\pi_P^*(I) = 0.5 \cdot 0.4 \cdot 0.9 \cdot 1 + 0.5 \cdot 0.6 \cdot 0.9 \cdot 1 = 0.5 \cdot (0.4 + 0.6) \cdot 0.9 \cdot 1 = 0.45$. 
Function Symbols Example: Throwing Dice

\[
on(0, 1) : 1/6 \lor on(0, 2) : 1/6 \lor \ldots \lor on(0, 6) : 1/6 \\
on(s(T), 1) : 1/6 \lor on(s(T), 2) : 1/6 \lor \ldots \lor on(s(T), 6) : 1/6 \\
\text{← on}(T, X), \neg on(T, 6)
\]

- We continue to throw a die until we obtain a 6
- \(\Psi_n = \{\psi | \psi \text{ is of the form } on(0, X_1) \land on(s(0), X_2) \land \ldots \land on(s(s(\ldots)), X_{n-1}) \land on(s(s(s(\ldots))), 6) \text{ with } X_i \neq 6\} \)
- \(\phi_n = \bigvee_{\psi \in \Psi_n} \psi, \pi^*_P(\phi_n) \text{ is the probability that we obtain a 6 at the } n-\text{th throw} \)
- \(\phi = \bigvee_{i=1}^{n} \phi_i, \pi^*_P(\phi) \text{ is the probability that we obtain a 6 in } n \text{ throws} \)
Proposal: let’s define $\pi_P^*(I)$ as
$$\pi_P^*(I) = \prod_{R \in P', I \models \text{body}(R)} \sigma_{\text{prob}}(R)$$

For a locally stratified LPADs without function symbols $P$ such that all the couples of clauses of $P'$ that share an atom in the head have mutually exclusive bodies with respect to the set of interpretations $\{I\}$ it coincides with the old definition.

For a program with function symbols such that the number of ground clauses with the body true in $I$ is finite $=>$ $\prod_{R \in P', I \models \text{body}(R)} \sigma_{\text{prob}}(R)$ is different from 0.

For what programs is this an admissible distribution?
Throwing Dice

- Time 0: throw of a die, if we get a 6 we stop, i.e. 
  \[ I_6 = \{on(0, 6)\} \text{ is such that } S(I_6) \neq \emptyset \]

\[
on(0, 6) \\
on(s(0), A) \leftarrow on(0, B), \neg on(0, 6) \\
\ldots
\]

- \( \pi^*_P(I_6) = 1/6 \)
- If at 0 we don’t get a 6, we throw the die again
Throwing Dice

- Time $s(0)$: if we get a 6 we stop, i.e.
  
  $I_{X6} = \{on(0, X), on(s(0), 6)\}$ is such that $S(I_{X6}) \neq \emptyset$ with $X = 1, \ldots, 5$

\[
\begin{align*}
on(0, X) \\
on(s(0), 6) & \leftarrow on(0, X), \neg on(0, 6) \\
on(s(s(0)), A) & \leftarrow on(s(0), B), \neg on(s(0), 6)
\end{align*}
\]

- \(\pi_P^*(I_{X6}) = 1/6^2\)
- There are 5 $I_{X6}$
- If at $s(0)$ we don’t get a 6, we throw the die again
Throwing Dice

- Time $s(s(0))$: if we get a 6 we stop, i.e.
  \[ I_{XY6} = \{on(0, X), on(s(0), Y), on(s(s(0)), 6)\} \]
  is such that
  \[ S(I_{XY6}) \neq \emptyset \text{ with } X, Y = 1, \ldots, 5 \]

  \[
  \begin{align*}
  on(0, X) \\
  on(s(0), Y) & \leftarrow on(0, X), \neg on(0, 6) \\
  on(s(s(0)), 6) & \leftarrow on(s(0), Y), \neg on(s(0), 6) \\
  on(s(s(s(0)))) , A) & \leftarrow on(s(s(0)), B), \neg on(s(s(0)), 6) \\
  \ldots
  \end{align*}
  \]

- $\pi_P^*(I_{XY6}) = 1/6^3$
- There are $5^2 I_{XY6}$
Throwing Dice

Overall:

\[
\frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} + \ldots = \\
= \frac{1}{5} \left( \frac{5}{6} + \left( \frac{5}{6} \right)^2 + \left( \frac{5}{6} \right)^3 + \ldots \right) = \\
= \frac{1}{5} \sum_{i=1}^{\infty} \left( \frac{5}{6} \right)^i = \\
= \frac{1}{5} \cdot \frac{\frac{5}{6}}{1 - \frac{5}{6}} = \frac{1}{5} \cdot \frac{5}{1} = \\
= 1
\]
Function Symbols

Under what conditions $\pi_P^*$ is an admissible distribution?

1. Every ground atom must depend on a finite number of ground atoms:

$$a(T) : p_a \lor b(T) : p_b \leftarrow a(s(T))$$

2. There exist a finite interpretation $I$ such that $S(I) \neq \emptyset$

$$a(0) : p_a \lor b(0) : p_b$$
$$a(s(T)) : p_a \lor b(s(T)) : p_b \leftarrow a(T)$$
$$a(s(T)) : p_a \lor b(s(T)) : p_b \leftarrow b(T)$$
Learning LPADs

- We consider a learning problem of the following form:

  **Given:**
  - a set $E$ of examples that are couples $(I, Pr(I))$ where $I$ is an interpretation and $Pr(I)$ is its associated probability
  - a space of possible LPAD $S$ (described by a language bias $LB$)

  **Find:**
  - an LPAD $P \in S$ such that
    $\forall (I, Pr(I)) \in E \quad \pi^*_P(I) = Pr(I)$

- Instead of a set of couples $(I, Pr(I))$, the input of the learning problem can be a multiset $E'$ of interpretations
**Preliminaries**

- **Definition** (adapted from [De Raedt, Dehaspe 96]): A ground clause $C$ is **non-trivially true** in an interpretation $I$ if $C$ is true in $I$ and $\text{body}(C)$ is true in $I$.

- **Definition**: The disjuncts in the head of a ground clause are **mutually exclusive with respect to a set of interpretations $J$** if there is no interpretation $I \in J$ such that two or more disjuncts are true in $I$.
Learning LPADs (LLPAD)

- LLPAD proceeds in three stages

- First stage: it searches for all the ground non-annotated disjunctive clauses allowed by the language bias that are true in all interpretations,
  - non trivially-true in at least one interpretation
  - whose disjuncts in the head are mutually exclusive,
  - every disjunct is true in at least one interpretation where the body is true

When it finds one such clause, it annotates the head disjuncts with a probability using theorem 1
Second stage: a constraint satisfaction problem is solved in order to find subsets of the annotated disjunctive clauses that form programs that assign to each interpretation the associated probability.

Third stage: the program is generalized to obtain non-ground clauses.

LLPAD learns a restricted class of LPADs: those such that every couple of clauses that share a literal in the head have mutually exclusive bodies over the set of interpretations $E$. 

Learning LPADs
First Stage

- Search for bodies that are true in at least one interpretation starting from the empty body.

- Every time a body is true in at least one interpretation, the algorithm searches, starting from the empty head, for all the heads that are true in the interpretations where the body is true ($EB$), whose disjuncts are mutually exclusive with respect to $EB$ and such that no disjunct is false on $EB$.

- Every clause that satisfy these requirements is annotated using theorem 1.
Example

$E$ is:

$I_1 = \{\text{heads}(\text{coin}), \text{toss}(\text{coin}), \text{fair}(\text{coin})\} \quad Pr(I_1) = 0.45$
$I_2 = \{\text{tails}(\text{coin}), \text{toss}(\text{coin}), \text{fair}(\text{coin})\} \quad Pr(I_2) = 0.45$
$I_3 = \{\text{heads}(\text{coin}), \text{toss}(\text{coin}), \text{biased}(\text{coin})\} \quad Pr(I_3) = 0.06$
$I_4 = \{\text{tails}(\text{coin}), \text{toss}(\text{coin}), \text{biased}(\text{coin})\} \quad Pr(I_4) = 0.04$
Example

toss:-true.

heads:0.51;tails:0.49:-true.
biased:0.1;fair:0.9:-true.
heads:0.51;tails:0.49:-toss.
biased:0.1;fair:0.9:-toss.
heads:0.6;tails:0.3999:-biased.
heads:0.5;tails:0.5:-fair.
biased:0.0816;fair:0.9183:-tails.
biased:0.1176;fair:0.8823:-heads.
heads:0.6;tails:0.3999:-toss,biased.
heads:0.5;tails:0.5:-toss,fair.
biased:0.0816;fair:0.9183:-toss,tails.
biased:0.1176;fair:0.8823:-toss,heads.
Second Stage

- Aim: partition the found disjunctive clauses in subsets that are solutions of the learning problem
- Assign a binary variable $x_i$ to each found clause $C_i$
Second Stage

- The couples of clauses that share a literal in the head must have mutually exclusive bodies over the set of interpretations $E$.

- Test, for each couple of clauses $(C_i, C_j)$, if they share a literal in the head and, if so, if the intersections of the two sets of interpretations where their body is true is non-empty.

- If so, we assert the constraint $x_i + x_j \leq 1$ for all such couples of clauses.
Second Stage

- For each interpretation we have the constraint:

\[
\prod_{C_i \in SC(I)} p_i^{x_i} = Pr(I)
\]

where \( SC(I) \) is the subset of the clauses returned by the algorithm whose body is true in \( I \) and \( p_i \) is the probability of the single head of \( C_i \) that is true in \( I \). This constraint is based on theorem 2.

- Definite clauses are not considered in the constraints because they would contribute only with a factor \( 1^{x_{ij}} \).
Second Stage

- Taking the logarithm of both members $\Rightarrow$ linear constraint:

$$\sum_{C_i \in SC(I)} x_i \log p_i = \log Pr(I)$$

- $\Rightarrow$ integer programming problem
Using the CLP(R) solver of Sicstus Prolog we obtain 16 solutions. The original program is among the 16 solutions:

biased: 0.1; fair: 0.9; true.
heads: 0.6; tails: 0.3999; -toss, biased.
heads: 0.5; tails: 0.5; -toss, fair.
toss; true.
Third Stage

Generalization of the ground clauses using lgg

cg(s,1,p):0.6;cg(s,1,w):0.4:-cg(f,1,p),cg(f,2,w).
cg(s,1,p):0.4;cg(s,1,w):0.6:-cg(f,1,w),cg(f,2,p).
cg(s,1,p):-cg(f,1,p),cg(f,2,p).
cg(s,1,w):-cg(f,1,w),cg(f,2,w).
cg(s,2,p):0.5;cg(s,2,w):0.5:-cg(m,1,p),cg(m,2,w).
cg(s,2,p):0.5;cg(s,2,w):0.5:-cg(m,1,w),cg(m,2,p).
cg(s,2,p):-cg(m,1,p),cg(m,2,p).
cg(s,2,w):-cg(m,1,w),cg(m,2,w).
cg(f,1,p):0.5;cg(f,1,w):0.5:-true.
cg(f,2,p):0.5;cg(f,2,w):0.5:-true.
cg(m,1,p):0.5;cg(m,1,w):0.5:-true.
cg(m,2,p):0.5;cg(m,2,w):0.5:-true.
Example, Third Stage

- Generalization of

\[
\begin{align*}
\text{cg}(s, 2, p) &: 0.5; \text{cg}(s, 2, w) &: 0.5: -\text{cg}(m, 1, p), \text{cg}(m, 2, w). \\
\text{cg}(s, 2, p) &: 0.5; \text{cg}(s, 2, w) &: 0.5: -\text{cg}(m, 1, w), \text{cg}(m, 2, p). \\
C &= \text{cg}(s, 2, A) &: 0.5; \text{cg}(s, 2, B) &: 0.5: -\text{cg}(m, 1, A), \text{cg}(m, 2, B), \\
&\quad \text{cg}(m, C, p), \text{cg}(m, D, w). \\
D &= \text{cg}(s, 2, w) &: 0.5; \text{cg}(s, 2, p) &: 0.5: -\text{cg}(m, 1, A), \text{cg}(m, 2, B), \\
&\quad \text{cg}(m, C, p), \text{cg}(m, D, w). \\
\end{align*}
\]

- Generalization of \(C\) with

\[
\begin{align*}
\text{cg}(s, 2, p) &: 1.0: -\text{cg}(m, 2, p), \text{cg}(m, 1, p). \\
E &= \text{cg}(s, 2, A) &: 0.5; \text{cg}(s, 2, B) &: 0.5: -\text{cg}(m, 1, A), \text{cg}(m, 2, B), \text{cg}(m, C, p). \\
\end{align*}
\]
Example, Third Stage

- Generalization of

\[
E = \text{cg}(s, 2, A): 0.5; \text{cg}(s, 2, B): 0.5; \neg \text{cg}(m, 1, A), \text{cg}(m, 2, B), \text{cg}(m, C, p).
\]

\[
\text{cg}(s, 2, w): 1.0; \neg \text{cg}(m, 2, w), \text{cg}(m, 1, w).
\]

\[
\text{cg}(s, 2, A): 0.5; \text{cg}(s, 2, B): 0.5; \neg \text{cg}(m, 1, A), \text{cg}(m, 2, B).
\]
# Mendel Experiment

<table>
<thead>
<tr>
<th>Members</th>
<th>Ex.</th>
<th>Phase 1</th>
<th>Rules</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>36</td>
<td>0.68</td>
<td>117</td>
<td>1.53</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>324</td>
<td>5.60</td>
<td>226</td>
<td>11.69</td>
<td>31.6</td>
</tr>
<tr>
<td>7</td>
<td>2916</td>
<td>144</td>
<td>338</td>
<td>856</td>
<td>5263</td>
</tr>
<tr>
<td>7 diff. bias</td>
<td>2916</td>
<td>712</td>
<td>4128</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Improvements

- LPADs can be represented as a Bayesian network
- => use an algorithm for learning BN for learning LPADs
http://www.ing.unife.it/software/LLPAD/