Uncertain Data Management
Relational Probabilistic Database Models

Antoine Amarilli[1], Silviu Maniu[2]

[1]Télécom ParisTech

[2]LRI

December 7th, 2015

Thanks to Pierre Senellart for proofreading these slides.

Table of contents

Uncertain instances

Remember from last class:

- Fix a finite set of possible tuples of same arity
- A possible world: a subset of the possible tuples
- An uncertain relation: set of possible worlds

Uncertain instances

Remember from last class:

- Fix a finite set of possible tuples of same arity
- A possible world: a subset of the possible tuples
- An uncertain relation: set of possible worlds

U_1

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

U_2

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

Probabilistic instances

- Support $\mathcal{U}$: uncertain relation

Probabilistic instances

- Support $\mathcal{U}$: uncertain relation
- Probability distribution π on $\mathcal{U}$:

Probabilistic instances

- Support $\mathcal{U}$: uncertain relation
- Probability distribution π on $\mathcal{U}$:
  - Function from $\mathcal{U}$ to reals in $[0, 1]$
  - It must sum up to 1: $\sum_{I \in \mathcal{U}} \pi(I) = 1$

Probabilistic instances

- Support $\mathcal{U}$: uncertain relation
- Probability distribution π on $\mathcal{U}$:
  - Function from $\mathcal{U}$ to reals in $[0, 1]$
  - It must sum up to 1: $\sum_{I \in \mathcal{U}} \pi(I) = 1$

U_1

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

U_2

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

Probabilistic instances

- Support $\mathcal{U}$: uncertain relation
- Probability distribution π on $\mathcal{U}$:
  - Function from $\mathcal{U}$ to reals in $[0, 1]$
  - It must sum up to 1: $\sum_{I \in \mathcal{U}} \pi(I) = 1$

U_1

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

$\pi(U_1) = 0.8$

U_2

date	teacher	room
04	Silviu	C017
04	Antoine	C017
04	Antoine	C47
11	Silviu	C017
11	Silviu	C47
11	Antoine	C017

$\pi(U_2) = 0.2$

What about NULLs?

Remember that last time we saw:

- Codd-tables and v-tables and c-tables, with NULLs
- Boolean c-tables, with NULLs only in conditions
  - $\rightarrow$ Boolean variables

What about NULLs?

Remember that last time we saw:

- Codd-tables and v-tables and c-tables, with NULLs
- Boolean c-tables, with NULLs only in conditions
  - → Boolean variables

→ We focus for probabilities on models like Boolean c-tables

→ Easier to define probabilities on a finite space!

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

U_1

04	S.	C017
11	S.	C47

U_2

11	A.	C017

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

U_1

04	S.	C017
11	S.	C47

$\bigcup$

U_2

11	A.	C017

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

U_1

04	S.	C017
11	S.	C47

U_2

11	A.	C017

$\bigcup$

V_1

V_2

11	A.	C017

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

U_1

04	S.	C017
11	S.	C47

U_2

11	A.	C017

$\bigcup$

V_1

(empty)

V_2

11	A.	C017

$=$

Relational algebra on uncertain instances

Remember from last class:

- Extend relational algebra operators to uncertain instances
- The possible worlds of the result should be...
  - take all possible worlds in the supports of the inputs
  - apply the operation and get the possible outputs

U_1

04	S.	C017
11	S.	C47

U_2

11	A.	C017

$\bigcup$

V_1

V_2

11	A.	C017

$=$

04	S.	C017
11	S.	C47

04	S.	C017
11	S.	C47
11	A.	C017

11	A.	C017

Relational algebra on probabilistic instances

- Let's adapt relational algebra to probabilistic instances
- The possible worlds of the result should be...

Relational algebra on probabilistic instances

- Let's adapt relational algebra to probabilistic instances
- The possible worlds of the result should be...
  - take all possible worlds of the inputs
  - apply the operation and get a possible output

Relational algebra on probabilistic instances

- Let's adapt relational algebra to probabilistic instances
- The possible worlds of the result should be...
  - take all possible worlds of the inputs
  - apply the operation and get a possible output
- The probability of each possible world should be...
  - consider all input possible worlds that give it
  - sum up their probabilities

Example of relational algebra on probabilistic instances

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

U_2

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\bigcup$

U_2

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

V_1

$\bigcup$

U_2

11	A.	C017

V_2

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

Example of relational algebra on probabilistic instances

U_1		
04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2		
11	A.	C017

$\pi(U_2) = 0.2$

$\cup$

V_1

$\pi(V_1) = 0.9$

V_2		
11	A.	C017

$\pi(V_2) = 0.1$

$=$

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\cup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

(empty)

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

W_2

04	S.	C017
11	S.	C47
11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\cup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

W_2

04	S.	C017
11	S.	C47
11	A.	C017

W_3

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

$\pi(W_1) = 0.8 \cdot 0.9$

W_2

04	S.	C017
11	S.	C47
11	A.	C017

W_3

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

$\pi(W_1) = 0.8 \cdot 0.9$

W_2

04	S.	C017
11	S.	C47
11	A.	C017

$\pi(W_1) = 0.8 \cdot 0.1$

W_3

11	A.	C017

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

$\pi(W_1) = 0.8 \cdot 0.9$

W_2

04	S.	C017
11	S.	C47
11	A.	C017

$\pi(W_1) = 0.8 \cdot 0.1$

W_3

11	A.	C017

$\pi(W_1) = 0.2 \cdot 0.9$

Example of relational algebra on probabilistic instances

U_1

04	S.	C017
11	S.	C47

$\pi(U_1) = 0.8$

U_2

11	A.	C017

$\pi(U_2) = 0.2$

$\bigcup$

V_1

$\pi(V_1) = 0.9$

V_2

11	A.	C017

$\pi(V_2) = 0.1$

$=$

W_1

04	S.	C017
11	S.	C47

$\pi(W_1) = 0.8 \cdot 0.9$

W_2

04	S.	C017
11	S.	C47
11	A.	C017

$\pi(W_1) = 0.8 \cdot 0.1$

W_3

11	A.	C017

$$\pi(W_1) = 0.2 \cdot 0.9 + 0.2 \cdot 0.1$$

Representation system

- Remember that if we have N possible tuples

Representation system

- Remember that if we have N possible tuples
  $\rightarrow$ there are 2^N possible instances

Representation system

- Remember that if we have N possible tuples
  - $\rightarrow$ there are 2^N possible instances
  - $\rightarrow$ there are 2^{2^N} possible uncertain instances

Representation system

- Remember that if we have N possible tuples
  - $\rightarrow$ there are 2^N possible instances
  - $\rightarrow$ there are 2^{2^N} possible uncertain instances
  - $\rightarrow$ writing out an uncertain instance is exponential

Representation system

- Remember that if we have N possible tuples
  - $\rightarrow$ there are 2^N possible instances
  - $\rightarrow$ there are 2^{2^N} possible uncertain instances
  - $\rightarrow$ writing out an uncertain instance is exponential

$\rightarrow$ Last time we saw Boolean c-tables as a concise way to represent uncertain instances

Representation system

- Remember that if we have N possible tuples
  - → there are 2^N possible instances
  - → there are 2^{2^N} possible uncertain instances
  - → writing out an uncertain instance is exponential

→ Last time we saw Boolean c-tables as a concise way to represent uncertain instances

- For probabilistic instances:
  - → there are infinitely many possible instances
  - → writing out a probabilistic instance is still exponential

Representation system

- Remember that if we have N possible tuples
  - $\rightarrow$ there are 2^N possible instances
  - $\rightarrow$ there are 2^{2^N} possible uncertain instances
  - $\rightarrow$ writing out an uncertain instance is exponential

$\rightarrow$ Last time we saw Boolean c-tables as a concise way to represent uncertain instances

- For probabilistic instances:
  - $\rightarrow$ there are infinitely many possible instances
  - $\rightarrow$ writing out a probabilistic instance is still exponential

$\rightarrow$ How to represent probabilistic instances?

Table of contents

Tuple-independent databases

- The simplest model: tuple-independent databases
- Annotate each instance fact with a probability

Tuple-independent databases

- The simplest model: tuple-independent databases
- Annotate each instance fact with a probability

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

Tuple-independent databases

- The simplest model: tuple-independent databases
- Annotate each instance fact with a probability

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

What's the probability of this outcome?

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

What's the probability of this outcome?

$$0.8$$

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

What's the probability of this outcome?

$$0.8 \times$$

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

What's the probability of this outcome?

$$0.8 \times (1 - 0.2)$$

Semantics of TID

- Each tuple is kept or discarded with the probability
- Probabilistic choices are independent across tuples

U

date	teacher	room	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

What's the probability of this outcome?

$$0.8 \times (1 - 0.2) \times 1$$

Getting a probability distribution

The semantics of a TID instance is a probabilistic instance...

$\rightarrow$ the possible worlds are the subsets

Getting a probability distribution

The semantics of a TID instance is a probabilistic instance...

$\rightarrow$ the possible worlds are the subsets

$\rightarrow$ always keeping tuples with probability 1

Getting a probability distribution

The semantics of a TID instance is a probabilistic instance...

$\rightarrow$ the possible worlds are the subsets

$\rightarrow$ always keeping tuples with probability 1

Formally, for a TID instance I, the probability of J:

Getting a probability distribution

The semantics of a TID instance is a probabilistic instance...

$\rightarrow$ the possible worlds are the subsets

$\rightarrow$ always keeping tuples with probability 1

Formally, for a TID instance I, the probability of J:

- we must have $J \subseteq I$
- product of $p_{\mathbf{t}}$ for each tuple $\mathbf{t}$ kept in J
- product of $1 - p_{\mathbf{t}}$ for each tuple $\mathbf{t}$ not kept in J

Is it a probability distribution?

Do the probabilities always sum to 1?

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$
- 2 tuples: $p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2)$

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$
- 2 tuples: $p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2)$
  $\rightarrow \ p_1(p_2 + (1 - p_2)) + (1 - p_1)(p_2 + (1 - p_2))$

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$
- 2 tuples: $p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2)$

  $\rightarrow$ $p_1(p_2 + (1 - p_2)) + (1 - p_1)(p_2 + (1 - p_2))$

  $\rightarrow$ $(p_1 + (1 - p_1))(p_2 + (1 - p_2))$

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$
- 2 tuples: $p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2)$
  - $\rightarrow\ p_1(p_2 + (1 - p_2)) + (1 - p_1)(p_2 + (1 - p_2))$
  - $\rightarrow\ (p_1 + (1 - p_1))(p_2 + (1 - p_2))$
  - $\rightarrow\ p_2 + (1 - p_2)$

Is it a probability distribution?

Do the probabilities always sum to 1?

- 0 tuples: vacuous
- 1 tuple: $p + (1 - p) = 1$
- 2 tuples: $p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2)$
  $\rightarrow\ p_1(p_2 + (1 - p_2)) + (1 - p_1)(p_2 + (1 - p_2))$
  $\rightarrow\ (p_1 + (1 - p_1))(p_2 + (1 - p_2))$
  $\rightarrow\ p_2 + (1 - p_2)$
- More tuples: $p_1(\cdots) + (1 - p_1)(\cdots)$ and induction hypothesis

Strong representation system

Remember from last class:

Uncertain instance: set of possible worlds

Strong representation system

Remember from last class:

Uncertain instance: set of possible worlds

Uncertainty framework: concise way to represent uncertain instances

Strong representation system

Remember from last class:

Uncertain instance:	set of possible worlds
Uncertainty framework:	concise way to represent uncertain instances
Query language:	here, relational algebra

Strong representation system

Remember from last class:

Uncertain instance: set of possible worlds

Uncertainty framework: concise way to represent uncertain instances

Query language: here, relational algebra

Definition (Strong representation system)

For any query in the language,

Strong representation system

Remember from last class:

Uncertain instance:	set of possible worlds
Uncertainty framework:	concise way to represent uncertain instances
Query language:	here, relational algebra

Definition (Strong representation system)

For any query in the language,
on uncertain instances represented in the framework,

Strong representation system

Remember from last class:

Uncertain instance:	set of possible worlds
Uncertainty framework:	concise way to represent uncertain instances
Query language:	here, relational algebra

Definition (Strong representation system)

For any query in the language,
on uncertain instances represented in the framework,
the uncertain instance obtained by evaluating the query

Strong representation system

Remember from last class:

Uncertain instance:	set of possible worlds
Uncertainty framework:	concise way to represent uncertain instances
Query language:	here, relational algebra

Definition (Strong representation system)

For any query in the language,
on uncertain instances represented in the framework,
the uncertain instance obtained by evaluating the query
can also be represented in the framework.

Strong representation system

Remember from last class:

Uncertain instance: set of possible worlds

Uncertainty framework: concise way to represent uncertain instances

Query language: here, relational algebra

Definition (Strong representation system)

For any query in the language,
on uncertain instances represented in the framework,
the uncertain instance obtained by evaluating the query
can also be represented in the framework.

$\rightarrow$ Are TID instances a strong representation system?

Implementing select

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1

Implementing select

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1

$\sigma_{\text{teacher}=\text{“Silviu”}}(U)$

date	teacher	room

Implementing select

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1

$\sigma_{\text{teacher}=\text{“Silviu”}}(U)$

date	teacher	room	
04	Silviu	C47	0.8
11	Silviu	C47	1

Implementing select

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1

$\sigma_{\text{teacher}=\text{"Silviu"}}(U)$

date	teacher	room	
04	Silviu	C47	0.8
11	Silviu	C47	1

$\rightarrow$ Is this correct?

Implementing select

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1

$\sigma_{\text{teacher}=\text{``Silviu''}}(U)$

date	teacher	room	
04	Silviu	C47	0.8
11	Silviu	C47	1

$\rightarrow$ Is this correct?

$\rightarrow$ So far, so good.

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date
04
11
18

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date	
04	
11	
18	0.9

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date	
04	
11	1
18	0.9

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date	
04	$1-(1-0.2)\cdot(1-0.8)$
11	1
18	0.9

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date	
04	$1-(1-0.2)\cdot(1-0.8)$
11	1
18	0.9

$\rightarrow$ Is this correct?

Implementing project

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2
11	Silviu	C47	1
11	Antoine	C47	0.1
18	Silviu	C47	0.9

$\pi_{\text{date}}(U)$

date	
04	$1-(1-0.2)\cdot(1-0.8)$
11	1
18	0.9

$\rightarrow$ Is this correct?

$\rightarrow$ So far, so good

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$U \times$ Repair

date	teacher	room	cause

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$U \times$ Repair

date	teacher	room	cause
04	Silviu	C47	leopard
04	Antoine	C47	leopard

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	0.8×0.1
04	Antoine	C47	leopard	0.2×0.1

Implementing product

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	0.8×0.1
04	Antoine	C47	leopard	0.2×0.1

$\rightarrow$ Is this correct?

Implementing product ... OR NOT!

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	0.8×0.1
04	Antoine	C47	leopard	0.2×0.1

→ Is this correct?

→ It's **WRONG**!

Why is it wrong?

U

date	**teacher**	**room**	
04	Silviu	C47	1
04	Antoine	C47	1

Why is it wrong?

U

date	teacher	room	
04	Silviu	C47	1
04	Antoine	C47	1

Repair

room	cause	
C47	leopard	$1/2$

Why is it wrong?

U

date	teacher	room	
04	Silviu	C47	1
04	Antoine	C47	1

Repair

room	cause	
C47	leopard	$1/2$

$U \times$ Repair

date	teacher	room	cause
04	Silviu	C47	leopard
04	Antoine	C47	leopard

Why is it wrong?

U

date	teacher	room	
04	Silviu	C47	1
04	Antoine	C47	1

Repair

room	cause	
C47	leopard	$1/2$

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

Why is it wrong?

U

date	teacher	room	
04	Silviu	C47	1
04	Antoine	C47	1

Repair

room	cause	
C47	leopard	$1/2$

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

→ The two tuples are not independent!

→ The first is there iff the second is there.

Why does it matter?

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

Why does it matter?

$U \times \text{Repair}$

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room
C47

Why does it matter?

$U \times \text{Repair}$

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	$1-(1-1/2)\times(1-1/2)$

Why does it matter?

$U \times \text{Repair}$

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	$1-(1-1/2)\times(1-1/2)$

$\rightarrow$ Probability of $3/4$...

Why does it matter?

$U \times$ Repair

date	teacher	room	cause	
04	Silviu	C47	leopard	$1/2$
04	Antoine	C47	leopard	$1/2$

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	$1-(1-1/2)\times(1-1/2)$

- → Probability of 3/4...
- → But the leopard had probability 1/2!

TID are not a strong representation system

- Remember how Codd tables required named nulls?
- The result of a query on TID may not be a TID

$\rightarrow$ We will see that the correlations can be complex

TID are not a strong representation system

- Remember how Codd tables required named nulls?
- The result of a query on TID may not be a TID

→ We will see that the correlations can be complex

- How to evaluate queries on a TID then?

→ List all possible worlds and count the probabilities

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\textbf{room}}(U \times \text{Repair})$

room
C47

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	???

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	???

- Either there is no leopard and then no result...

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause
C47	leopard 0.1

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	???

- Either there is no leopard and then no result...
- Or there is a leopard and then...

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause
C47	leopard 0.1

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	???

- Either there is no leopard and then no result...
- Or there is a leopard and then...
  - Non-empty result:

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\mathbf{room}}(U \times \text{Repair})$

room	
C47	???

- Either there is no leopard and then no result...
- Or there is a leopard and then...
  - Non-empty result: $1 - (1 - 0.8)(1 - 0.2)$

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\textbf{room}}(U \times \text{Repair})$

room	
C47	???

- Either there is no leopard and then no result...
- Or there is a leopard and then...
  - Non-empty result: $1 - (1 - 0.8)(1 - 0.2) = 0.84$

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause
C47	leopard 0.1

$\pi_{\textbf{room}}(U \times \text{Repair})$

room
C47

- Either there is no leopard and then no result...
- Or there is a leopard and then...
  - Non-empty result: $1 - (1 - 0.8)(1 - 0.2) = 0.84$
- The query probability is:

Query evalation done right

U

date	teacher	room	
04	Silviu	C47	0.8
04	Antoine	C47	0.2

Repair

room	cause	
C47	leopard	0.1

$\pi_{\textbf{room}}(U \times \text{Repair})$

room	
C47	0.084

- Either there is no leopard and then no result...
- Or there is a leopard and then...
  - Non-empty result: $1 - (1 - 0.8)(1 - 0.2) = 0.84$
- The query probability is: 0.1×0.84

Expressiveness of TID

Can we represent all probabilistic instances with TID?

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu

$\pi(U_1) = 0.8$

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu
$\pi(U_1) = 0.8$

U_2

teacher
Antoine
$\pi(U_2) = 0.1$

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu

$\pi(U_1) = 0.8$

U_2

teacher
Antoine

$\pi(U_2) = 0.1$

U_3

teacher

$\pi(U_3) = 0.1$

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1
teacher
Silviu
$\pi(U_1) = 0.8$

U_2
teacher
Antoine
$\pi(U_2) = 0.1$

U_3
teacher
$\pi(U_3) = 0.1$

U
teacher
Antoine
Silviu

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu

$\pi(U_1) = 0.8$

U_2

teacher
Antoine

$\pi(U_2) = 0.1$

U_3

teacher

$\pi(U_3) = 0.1$

U

teacher	
Antoine	0.1
Silviu	

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu

$\pi(U_1) = 0.8$

U_2

teacher
Antoine

$\pi(U_2) = 0.1$

U_3

teacher

$\pi(U_3) = 0.1$

U

teacher	
Antoine	0.1
Silviu	0.8

Expressiveness of TID

Can we represent all probabilistic instances with TID?

"The class is taught by Antoine or Silviu or no one but not both"

U_1

teacher
Silviu

$\pi(U_1) = 0.8$

U_2

teacher
Antoine

$\pi(U_2) = 0.1$

U_3

teacher

$\pi(U_3) = 0.1$

U

teacher	
Antoine	0.1
Silviu	0.8

$\rightarrow$ We cannot forbid that both teach the class!

Table of contents

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (underlined)

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (<u>underlined</u>)

U

<u>mon</u>	<u>day</u>	teacher	room
Jan	04	Silviu	C017
Jan	04	Antoine	C017
Jan	11	Silviu	C47
Jan	11	Antoine	C017

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (underlined)

U

<u>mon</u>	<u>day</u>	teacher	room
Jan	04	Silviu	C017
Jan	04	Antoine	C017
Jan	11	Silviu	C47
Jan	11	Antoine	C017

- The blocks are the sets of tuples with the same key

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (underlined)

U

mon	day	teacher	room
Jan	04	Silviu	C017
Jan	04	Antoine	C017
Jan	11	Silviu	C47
Jan	11	Antoine	C017

- The blocks are the sets of tuples with the same key
- Each tuple has a probability

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (underlined)

U

<u>mon</u>	<u>day</u>	teacher	room	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- The blocks are the sets of tuples with the same key
- Each tuple has a probability

Block-independent disjoint instances

- A more expressive framework than TID
- Call some attributes the key (underlined)

U

mon	day	teacher	room	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- The blocks are the sets of tuples with the same key
- Each tuple has a probability
- Probabilities must sum to ≤ 1 in each block

BID semantics

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

BID semantics

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- For each block:

BID semantics

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- For each block:
  - Pick one tuple according to probabilities

BID semantics

U

<u>mon</u>	<u>day</u>	teacher	room	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- For each block:
  - Pick one tuple according to probabilities
  - Possibly no tuple if probabilities are < 1

BID semantics

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

- For each block:
  - Pick one tuple according to probabilities
  - Possibly no tuple if probabilities are < 1

$\rightarrow$ Do choices independentently in each block

BID semantics

U

<u>mon</u>	<u>day</u>	teacher	room	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

U

<u>mon</u>	<u>day</u>	teacher	room

- For each block:
  - Pick one tuple according to probabilities
  - Possibly no tuple if probabilities are < 1

$\rightarrow$ Do choices independentently in each block

BID semantics

U

mon	day	teacher	room	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

U

mon	day	teacher	room
Jan	04	Silviu	C017
Jan	04	Antoine	C017

- For each block:
  - Pick one tuple according to probabilities
  - Possibly no tuple if probabilities are < 1

$\rightarrow$ Do choices independentently in each block

BID semantics

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**	
Jan	04	Silviu	C017	0.9
Jan	04	Antoine	C017	0.1
Jan	11	Silviu	C47	0.8
Jan	11	Antoine	C017	0.1

U

<u>**mon**</u>	<u>**day**</u>	**teacher**	**room**
Jan	04	Silviu	C017
Jan	04	Antoine	C017
Jan	11	Silviu	C47
Jan	11	Antoine	C017

- For each block:
  - Pick one tuple according to probabilities
  - Possibly no tuple if probabilities are < 1

→ Do choices independentently in each block

BID captures TID

- Each TID can be expressed as a BID...

BID captures TID

- Each TID can be expressed as a BID...
  - → Take all attributes as key
  - → Each block contains a single tuple

BID captures TID

- Each TID can be expressed as a BID...
  - → Take all attributes as key
  - → Each block contains a single tuple

U

date	**teacher**	**room**	
04	Silviu	C017	0.8
04	Antoine	C017	0.2
11	Silviu	C017	1

Expressiveness of BID

Can we represent all probabilistic instances with BID?

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

U_2

teacher
Antoine
Fabian

$\pi(U_2) = 0.1$

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

U_2

teacher
Antoine
Fabian

$\pi(U_2) = 0.1$

U_3

teacher
Antoine
Silviu

$\pi(U_3) = 0.1$

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

U_2

teacher
Antoine
Fabian

$\pi(U_2) = 0.1$

U_3

teacher
Antoine
Silviu

$\pi(U_3) = 0.1$

$\rightarrow$ If **teacher** is a key **teacher**, then TID

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

U_2

teacher
Antoine
Fabian

$\pi(U_2) = 0.1$

U_3

teacher
Antoine
Silviu

$\pi(U_3) = 0.1$

$\rightarrow$ If **teacher** is a key **teacher**, then TID

$\rightarrow$ If **teacher** is not a key, then only one tuple

Expressiveness of BID

Can we represent all probabilistic instances with BID?

"The class is taught by exactly two among Antoine, Silviu, Fabian."

U_1

teacher
Silviu
Fabian

$\pi(U_1) = 0.8$

U_2

teacher
Antoine
Fabian

$\pi(U_2) = 0.1$

U_3

teacher
Antoine
Silviu

$\pi(U_3) = 0.1$

→ If **teacher** is a key **teacher**, then TID

→ If **teacher** is not a key, then only one tuple

→ We cannot represent this probabilistic instance as a BID

Table of contents

Boolean c-tables

Remember Boolean c-tables:

- Set of Boolean variables $x_1, x_2, \ldots$
- Each tuple carries a condition
  - Variables and true and false
  - Boolean operators

Boolean c-tables

Remember Boolean c-tables:

- Set of Boolean variables $x_1, x_2, \ldots$
- Each tuple carries a condition
  - Variables and true and false
  - Boolean operators

date	teacher	room	
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

x_1 Silviu is sick

x_2 Projector in C017 is working

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations
- The probability of valuation ν is:

pc-tables

A (Boolean) pc-table is a Boolean c-table
plus a probability p_i for each x_i
indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations
- The probability of valuation ν is:
  - Product of the p_i for the x_i with $\nu(x_i) = 1$

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations
- The probability of valuation ν is:
  - Product of the p_i for the x_i with $\nu(x_i) = 1$
  - Product of the $1 - p_i$ for the x_i with $\nu(x_i) = 0$

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations
- The probability of valuation ν is:
  - Product of the p_i for the x_i with $\nu(x_i) = 1$
  - Product of the $1 - p_i$ for the x_i with $\nu(x_i) = 0$
  - Sounds familiar?

pc-tables

A (Boolean) pc-table is a Boolean c-table plus a probability p_i for each x_i indicating the independent probability that x_i is true.

Formally:

- A Boolean valuation ν of the x_i maps each to 0 or 1
  - Possible world of the Boolean c-instance under ν
  - The possible worlds are the worlds over all valuations
- The probability of valuation ν is:
  - Product of the p_i for the x_i with $\nu(x_i) = 1$
  - Product of the $1 - p_i$ for the x_i with $\nu(x_i) = 0$
  - Sounds familiar?

    $\rightarrow$ Yeah, it's like TID instances!

pc-table example

date	teacher	room	
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

pc-table example

date	teacher	room	
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

x_1 Silviu is sick

x_2 Projector in C017 is working

pc-table example

date	teacher	room	
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \land \neg x_1$
11	Antoine	C017	$x_2 \land x_1$
11	Silviu	C47	$\neg x_2 \land \neg x_1$
11	Antoine	C47	$\neg x_2 \land x_1$

x_1 Silviu is sick

$\rightarrow$ Probability 0.1

x_2 Projector in C017 is working

$\rightarrow$ Probability 0.2

pc-table semantics example

date	**teacher**	**room**	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \land \neg x_1$
11	Antoine	C017	$x_2 \land x_1$
11	Silviu	C47	$\neg x_2 \land \neg x_1$
11	Antoine	C47	$\neg x_2 \land x_1$

pc-table semantics example

date	**teacher**	**room**	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \land \neg x_1$
11	Antoine	C017	$x_2 \land x_1$
11	Silviu	C47	$\neg x_2 \land \neg x_1$
11	Antoine	C47	$\neg x_2 \land x_1$

- Take ν mapping x_1 to 0 and x_2 to 1

pc-table semantics example

date	teacher	room	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν:

pc-table semantics example

date	**teacher**	**room**	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν: $(1 - 0.1) \times 0.2 = 0.18$

pc-table semantics example

date	teacher	room	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \land \neg x_1$
11	Antoine	C017	$x_2 \land x_1$
11	Silviu	C47	$\neg x_2 \land \neg x_1$
11	Antoine	C47	$\neg x_2 \land x_1$

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν: $(1 - 0.1) \times 0.2 = 0.18$
- Evaluate the conditions

pc-table semantics example

date	**teacher**	**room**	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \land \neg x_1$
11	Antoine	C017	$x_2 \land x_1$
11	Silviu	C47	$\neg x_2 \land \neg x_1$
11	Antoine	C47	$\neg x_2 \land x_1$

date	**teacher**	**room**
04	Silviu	C42
04	Antoine	C42
11	Silviu	C017
11	Antoine	C017
11	Silviu	C47
11	Antoine	C47

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν: $(1 - 0.1) \times 0.2 = 0.18$
- Evaluate the conditions

pc-table semantics example

date	teacher	room	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

date	teacher	room
04	Silviu	C42
04	Antoine	C42
11	Silviu	C017
11	Antoine	C017
11	Silviu	C47
11	Antoine	C47

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν: $(1 - 0.1) \times 0.2 = 0.18$
- Evaluate the conditions

$\rightarrow$ Probability of possible world: sum over the valuations

pc-table semantics example

date	teacher	room	$x_1 : 0.1,\ x_2 : 0.2$
04	Silviu	C42	$\neg x_1$
04	Antoine	C42	x_1
11	Silviu	C017	$x_2 \wedge \neg x_1$
11	Antoine	C017	$x_2 \wedge x_1$
11	Silviu	C47	$\neg x_2 \wedge \neg x_1$
11	Antoine	C47	$\neg x_2 \wedge x_1$

date	teacher	room
04	Silviu	C42
04	Antoine	C42
11	Silviu	C017
11	Antoine	C017
11	Silviu	C47
11	Antoine	C47

- Take ν mapping x_1 to 0 and x_2 to 1
- Probability of ν: $(1-0.1)\times 0.2 = 0.18$
- Evaluate the conditions

→ Probability of possible world: sum over the valuations

→ Here: only this valuation, 0.18

pc-tables capture TID

Give each tuple its own variable:

U

date	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C017

pc-tables capture TID

Give each tuple its own variable:

U

date	teacher	room	
04	Silviu	C017	x_1
04	Antoine	C017	x_2
11	Silviu	C017	x_3

pc-tables capture TID

Give each tuple its own variable:

date	teacher	room	
04	Silviu	C017	x_1
04	Antoine	C017	x_2
11	Silviu	C017	x_3

U

$\rightarrow$ Give each variable the probability of the tuple

pc-tables capture mutually exclusive

- Remember non-Boolean c-tables

pc-tables capture mutually exclusive

- Remember non-Boolean c-tables

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 1$
Jan	04	Antoine	C017	$x = 2$
Jan	04	Fabian	C017	$x = 3$

pc-tables capture mutually exclusive

- Remember non-Boolean c-tables

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 1$
Jan	04	Antoine	C017	$x = 2$
Jan	04	Fabian	C017	$x = 3$

- Give a probability to each value of x, summing up to 1

pc-tables capture mutually exclusive

- Remember non-Boolean c-tables

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 1$
Jan	04	Antoine	C017	$x = 2$
Jan	04	Fabian	C017	$x = 3$

- Give a probability to each value of x, summing up to 1

  $\rightarrow$ Example: x has probability:

  - 0.8 to be 1
  - 0.1 to be 2
  - 0.1 to be 3

pc-tables capture mutually exclusive

- Remember non-Boolean c-tables

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 1$
Jan	04	Antoine	C017	$x = 2$
Jan	04	Fabian	C017	$x = 3$

- Give a probability to each value of x, summing up to 1
  - $\rightarrow$ Example: x has probability:
    - 0.8 to be 1
    - 0.1 to be 2
    - 0.1 to be 3
- Remember our rewriting from non-Boolean to Boolean...

Reminder: rewriting non-Boolean to Boolean

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 00$
Jan	04	Antoine	C017	$x = 01$
Jan	04	Fabian	C017	$x = 10$

Reminder: rewriting non-Boolean to Boolean

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 00$
Jan	04	Antoine	C017	$x = 01$
Jan	04	Fabian	C017	$x = 10$

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$\neg x_1 \wedge \neg x_2$
Jan	04	Antoine	C017	$\neg x_1 \wedge x_2$
Jan	04	Fabian	C017	$x_1 \wedge \neg x_2$

Reminder: rewriting non-Boolean to Boolean

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 00$
Jan	04	Antoine	C017	$x = 01$
Jan	04	Fabian	C017	$x = 10$

U

mon	day	teacher	room	
Jan	04	Silviu	C017	$\neg x_1 \wedge \neg x_2$
Jan	04	Antoine	C017	$\neg x_1 \wedge x_2$
Jan	04	Fabian	C017	$x_1 \wedge \neg x_2$

$\rightarrow$ How to choose the probabilities?

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0
- See the rewriting as a decision tree:
  Either the first bit is 0 or it is 1:
  - if the first bit is 0, then either the second is 0 or it is 1
  - if the first bit is 1, then either the second is 0 or it is 1

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0
- See the rewriting as a decision tree:
  Either the first bit is 0 or it is 1:
  - if the first bit is 0, then either the second is 0 or it is 1
  - if the first bit is 1, then either the second is 0 or it is 1
- Use variable x_1 for the first choice, proba 0.1
  - If $x_1 = 0$ use variable x_2 for the second choice, proba...

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0
- See the rewriting as a decision tree:
  Either the first bit is 0 or it is 1:
  - if the first bit is 0, then either the second is 0 or it is 1
  - if the first bit is 1, then either the second is 0 or it is 1
- Use variable x_1 for the first choice, proba 0.1
  - If $x_1 = 0$ use variable x_2 for the second choice, proba... $1/9$

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0
- See the rewriting as a decision tree:
  Either the first bit is 0 or it is 1:
  - if the first bit is 0, then either the second is 0 or it is 1
  - if the first bit is 1, then either the second is 0 or it is 1
- Use variable x_1 for the first choice, proba 0.1
  - If $x_1 = 0$ use variable x_2 for the second choice, proba... $1/9$
  - If $x_1 = 1$ use variable x_2' for the second choice, proba...

Choosing the probabilities

- We start with the probabilities:
  - $x = 00$ has probability 0.8
  - $x = 01$ has probability 0.1
  - $x = 10$ has probability 0.1
  - $x = 11$ has probability 0
- See the rewriting as a decision tree:
  Either the first bit is 0 or it is 1:
  - if the first bit is 0, then either the second is 0 or it is 1
  - if the first bit is 1, then either the second is 0 or it is 1
- Use variable x_1 for the first choice, proba 0.1
  - If $x_1 = 0$ use variable x_2 for the second choice, proba... $1/9$
  - If $x_1 = 1$ use variable x_2' for the second choice, proba... 0

Converting mutually exclusive to pc-tables

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 00$
Jan	04	Antoine	C017	$x = 01$
Jan	04	Fabian	C017	$x = 10$

- Probabilities: x has proba 0.8 to be 1, 0.1 to be 2, 0.1 to be 3

→ Rewriting:

Converting mutually exclusive to pc-tables

mon	day	teacher	room	
Jan	04	Silviu	C017	$x = 00$
Jan	04	Antoine	C017	$x = 01$
Jan	04	Fabian	C017	$x = 10$

- Probabilities: x has proba 0.8 to be 1, 0.1 to be 2, 0.1 to be 3

→ Rewriting:

mon	day	teacher	room	
Jan	04	Silviu	C017	$\neg x_1 \wedge \neg x_2$
Jan	04	Antoine	C017	$\neg x_1 \wedge x_2$
Jan	04	Fabian	C017	$x_1 \wedge x_2'$

→ x_1 has proba $1/9$, x_2 has proba $1/2$, x_2' has proba 0

Capturing BID with pc-tables

- This process generalizes: create decision trees

Capturing BID with pc-tables

- This process generalizes: create decision trees
- We can capture BID by doing this in each block

day	teacher	room	
04	Silviu	C017	0.9
04	Antoine	C017	0.1
11	Silviu	C47	0.8
11	Antoine	C017	0.1

Capturing BID with pc-tables

- This process generalizes: create decision trees
- We can capture BID by doing this in each block

day	teacher	room	
04	Silviu	C017	0.9
04	Antoine	C017	0.1
11	Silviu	C47	0.8
11	Antoine	C017	0.1

day	teacher	room
04	Silviu	C017
04	Antoine	C017
11	Silviu	C47
11	Antoine	C017

Capturing BID with pc-tables

- This process generalizes: create decision trees
- We can capture BID by doing this in each block

day	teacher	room	
04	Silviu	C017	0.9
04	Antoine	C017	0.1
11	Silviu	C47	0.8
11	Antoine	C017	0.1

day	teacher	room	
04	Silviu	C017	$\neg x_1$
04	Antoine	C017	x_1
11	Silviu	C47	$\neg y_1 \land \neg y_2$
11	Antoine	C017	$\neg y_1 \land y_2$

Capturing BID with pc-tables

- This process generalizes: create decision trees
- We can capture BID by doing this in each block

day	teacher	room	
04	Silviu	C017	0.9
04	Antoine	C017	0.1
11	Silviu	C47	0.8
11	Antoine	C017	0.1

day	teacher	room	
04	Silviu	C017	$\neg x_1$
04	Antoine	C017	x_1
11	Silviu	C47	$\neg y_1 \land \neg y_2$
11	Antoine	C017	$\neg y_1 \land y_2$

x_1 has probability 0.1

y_1 has probability 0.1

y_2 has probability $1/9$

Strong representation system

- Remember from last class:
  Boolean c-tables are a strong representation system
  - ... because c-tables are

Strong representation system

- Remember from last class:
  Boolean c-tables are a strong representation system
  - ... because c-tables are
- Further, each valuation of the output
  is the output for the same valuation of the inputs
  → assuming that variables in the input relations are different
  → this preserves probabilities

Strong representation system

- Remember from last class:
  Boolean c-tables are a strong representation system
  - ... because c-tables are
- Further, each valuation of the output
  is the output for the same valuation of the inputs
  - $\rightarrow$ assuming that variables in the input relations are different
  - $\rightarrow$ this preserves probabilities

$\rightarrow$ pc-tables are a strong representation system

Capturing all probabilistic instances

- Remember:
  - Support $\mathcal{U}$: uncertain relation
    - Here, set of subsets of a finite set of tuples
  - Probability distribution π on $\mathcal{U}$

Capturing all probabilistic instances

- Remember:
  - Support $\mathcal{U}$: uncertain relation
    - Here, set of subsets of a finite set of tuples
  - Probability distribution π on $\mathcal{U}$

$\rightarrow$ Can any probabilistic instance be represented by a pc-table?

Capturing uncertain instances with Boolean c-tables (1)

Remember from last time:

- Number the possible worlds in binary
- For each tuple, write the possible worlds where it appears

Capturing uncertain instances with Boolean c-tables (1)

Remember from last time:

- Number the possible worlds in binary
- For each tuple, write the possible worlds where it appears

00

v	w
a	d
b	e
c	f

01

v	w
a	d
b	e
c	f

10

v	w
a	d
b	e
c	f

11

v	w
a	d
b	e
c	f

Capturing uncertain instances with Boolean c-tables (1)

Remember from last time:

- Number the possible worlds in binary
- For each tuple, write the possible worlds where it appears

00		01		10		11	
v	**w**	**v**	**w**	**v**	**w**	**v**	**w**
a	d	a	d	a	d	a	d
b	e	b	e	b	e	b	e
c	f	c	f	c	f	c	f

v	**w**	
a	d	$x = 00 \vee x = 01 \vee x = 10 \vee x = 11$
b	e	$x = 01$
c	f	$x = 01 \vee x = 10 \vee x = 11$

Capturing uncertain instances with Boolean c-tables (1)

Remember from last time:

- Number the possible worlds in binary
- For each tuple, write the possible worlds where it appears

00

v	**w**
a	d
b	e
c	f

01

v	**w**
a	d
b	e
c	f

10

v	**w**
a	d
b	e
c	f

11

v	**w**
a	d
b	e
c	f

v	**w**	
a	d	$x = 00 \lor x = 01 \lor x = 10 \lor x = 11$
b	e	$x = 01$
c	f	$x = 01 \lor x = 10 \lor x = 11$

$\rightarrow$ We can also do this with pc-tables

Capturing uncertain instances with Boolean c-tables (2)

Remember: the second step was to reduce to binary:

v	**w**	
a	d	$x = 00 \lor x = 01 \lor x = 10 \lor x = 11$
b	e	$x = 01$
c	f	$x = 01 \lor x = 10 \lor x = 11$

Capturing uncertain instances with Boolean c-tables (2)

Remember: the second step was to reduce to binary:

v	w	
a	d	$x = 00 \lor x = 01 \lor x = 10 \lor x = 11$
b	e	$x = 01$
c	f	$x = 01 \lor x = 10 \lor x = 11$

v	w	
a	d	$\neg x_1 \land \neg x_2 \lor \neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$
b	e	$\neg x_1 \land x_2$
c	f	$\neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$

Capturing uncertain instances with Boolean c-tables (2)

Remember: the second step was to reduce to binary:

v	w	
a	d	$x = 00 \lor x = 01 \lor x = 10 \lor x = 11$
b	e	$x = 01$
c	f	$x = 01 \lor x = 10 \lor x = 11$

v	w	
a	d	$\neg x_1 \land \neg x_2 \lor \neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$
b	e	$\neg x_1 \land x_2$
c	f	$\neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$

- For pc-instances, how to choose the probabilities?

Capturing uncertain instances with Boolean c-tables (2)

Remember: the second step was to reduce to binary:

v	w	
a	d	$x = 00 \lor x = 01 \lor x = 10 \lor x = 11$
b	e	$x = 01$
c	f	$x = 01 \lor x = 10 \lor x = 11$

v	w	
a	d	$\neg x_1 \land \neg x_2 \lor \neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$
b	e	$\neg x_1 \land x_2$
c	f	$\neg x_1 \land x_2 \lor x_1 \land \neg x_2 \lor x_1 \land x_2$

- For pc-instances, how to choose the probabilities?

$\rightarrow$ We have seen this: this is encoding a mutually exclusive choice

Table of contents

Summary

We have seen relational formalisms for probabilistic instances:

- TID, a simple model with independent probabilities on tuples
- BID, adding blocks with mutually exclusive choices
- pc-tables, i.e., Boolean c-tables with probabilities on variables

$\rightarrow$ pc-tables can capture any probabilistic instance

Summary

We have seen relational formalisms for probabilistic instances:

- TID, a simple model with independent probabilities on tuples
- BID, adding blocks with mutually exclusive choices
- pc-tables, i.e., Boolean c-tables with probabilities on variables

$\rightarrow$ pc-tables can capture any probabilistic instance

- In the next class: how to evaluate queries efficiently
- Let's see a few advanced topics

Conditioning

- With probabilities, conditioning is a common operation
  - *What is the probability that it rains given that the grass is wet?*

Conditioning

- With probabilities, conditioning is a common operation
  - *What is the probability that it rains given that the grass is wet?*
- Conditioning a pc-table:
  - $\rightarrow$ Condition the distribution on possible worlds
  - $\rightarrow$ Easy to condition by $x_1 = 1$
  - $\rightarrow$ Hard to condition on "this tuple is there"

Conditioning

- With probabilities, conditioning is a common operation
  - *What is the probability that it rains given that the grass is wet?*
- Conditioning a pc-table:
  - → Condition the distribution on possible worlds
  - → Easy to condition by $x_1 = 1$
  - → Hard to condition on "this tuple is there"
- Idea: pc-table with global condition

Conditioning

- With probabilities, conditioning is a common operation
  - *What is the probability that it rains given that the grass is wet?*
- Conditioning a pc-table:
  - → Condition the distribution on possible worlds
  - → Easy to condition by $x_1 = 1$
  - → Hard to condition on "this tuple is there"
- Idea: pc-table with global condition

$x_1 \vee x_2 \vee x_3$

teacher	
Antoine	x_1
Fabian	x_2
Silviu	x_3

Conditioning

- With probabilities, conditioning is a common operation
  - *What is the probability that it rains given that the grass is wet?*
- Conditioning a pc-table:

  → Condition the distribution on possible worlds

  → Easy to condition by $x_1 = 1$

  → Hard to condition on "this tuple is there"
- Idea: pc-table with global condition

$x_1 \vee x_2 \vee x_3$

teacher	
Antoine	x_1
Fabian	x_2
Silviu	x_3

→ Semantics: ignore valuations that violate the global condition

→ Easier to add things to the global condition

Other models

- We will see non-relational models of uncertainty

→ There are also other relational models

Other models

- We will see non-relational models of uncertainty

$\rightarrow$ There are also other relational models

- Possibilistic databases:
  - Do not consider the probability that a fact is true but the degree of surprise caused by a fact
  - The possibility of a world is its highest degree of surprise
  - Also, fuzzy databases: facts can be any intermediate value between true (1) and false (0)

Other models

- We will see non-relational models of uncertainty

→ There are also other relational models

- Possibilistic databases:
  - Do not consider the probability that a fact is true but the degree of surprise caused by a fact
  - The possibility of a world is its highest degree of surprise
  - Also, fuzzy databases: facts can be any intermediate value between true (1) and false (0)

- Continuous distributions: impose conditions like "this value follows a normal distribution"
  - Usually intractable to reason with
  - MCDBs: Monte Carlo DataBases: use sampling

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance
  - $\rightarrow$ Choose the world and join with the table of the worlds

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance
  - $\rightarrow$ Choose the world and join with the table of the worlds
- Any probabilistic relation can be obtained with a relational algebra query on a TID instance

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance
  - $\rightarrow$ Choose the world and join with the table of the worlds
- Any probabilistic relation can be obtained with a relational algebra query on a TID instance
  - $\rightarrow$ Code mutually exclusive with TID + RA ("largest value")

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance
  - $\rightarrow$ Choose the world and join with the table of the worlds
- Any probabilistic relation can be obtained with a relational algebra query on a TID instance
  - $\rightarrow$ Code mutually exclusive with TID + RA ("largest value")
- TID instances plus UCQ do not suffice

Views and normalization

Idea: obtain complex probabilistic instances as the result of evaluating a query (or view) on simple instances

- Any probabilistic relation can be obtained with a CQ on a (single-block) BID instance
  - $\rightarrow$ Choose the world and join with the table of the worlds
- Any probabilistic relation can be obtained with a relational algebra query on a TID instance
  - $\rightarrow$ Code mutually exclusive with TID + RA ("largest value")
- TID instances plus UCQ do not suffice
  - $\rightarrow$ Always a maximal world for inclusion

References I

Abiteboul, S., Hull, R., and Vianu, V. (1995).
Foundations of Databases.
Addison-Wesley.
http://webdam.inria.fr/Alice/pdfs/all.pdf.

Barbará, D., Garcia-Molina, H., and Porter, D. (1992).
The management of probabilistic data.
IEEE Transactions on Knowledge and Data Engineering, 4(5).
http://www.iai.uni-bonn.de/III/lehre/AG/IntelligenteDatenbanken/Seminar/SS05/Literatur/%5BBGP92%5DProbData_IEEE_TKDE.pdf.

References II

Dalvi, N. N. and Suciu, D. (2007).
Efficient query evaluation on probabilistic databases.
VLDB Journal.
http://www.vldb.org/conf/2004/RS22P1.PDF.

Green, T. J. and Tannen, V. (2006).
Models for incomplete and probabilistic information.
IEEE Data Eng. Bull.
http://sites.computer.org/debull/A06mar/green.ps.

Huang, J., Antova, L., Koch, C., and Olteanu, D. (2009).
MayBMS: a probabilistic database management system.
In *SIGMOD.*
https://www.cs.ox.ac.uk/dan.olteanu/papers/hako-sigmod09.pdf.

References III

Lakshmanan, L. V. S., Leone, N., Ross, R. B., and Subrahmanian, V. S. (1997).
ProbView: A flexible probabilistic database system.
ACM Transactions on Database Systems.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.293&rep=rep1&type=pdf.

Ré, C. and Suciu, D. (2007).
Materialized views in probabilistic databases: for information exchange and query optimization.
In *VLDB*.
http://www.cs.stanford.edu/people/chrismre/papers/prob_materialized_views_TR.pdf.

References IV

Suciu, D., Olteanu, D., Ré, C., and Koch, C. (2011).
Probabilistic Databases.
Morgan & Claypool.
Unavailable online.