Tutorial
Probability, Statistics and Concept Lattices

Richard EMILION (MAPMO Lab., Orléans University, France)

DATA ANALYSIS AND MODELING LAB

Palacky University, Olomouc, Czech Republic

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

Outline

- **Part I - Motivations**
- **Part II - Models**
- **Part III - Sampling**
- **Part IV - Pointwise convergence of empirical CLs**
- **Part V - Experiments, Regression**

Part I - MOTIVATIONS

Models
Sampling

• Context $\mathcal{C} = (I, J, \mathcal{D})$ (Binary matrix case), $\mathcal{L}$ its concept lattice.

• Examples of complex and time consuming tasks : listing $\mathcal{L}$, the frequent itemsets, the associative rules

• Probabilistic and Statistical methods can be used at least for :

1. *Modelling*
2. *Sampling, Bootstrapping*

I.1. Motivations

• Context $\mathcal{C} = (I, J, \mathcal{D})$ (Binary matrix case), $\mathcal{L}$ its concept lattice.

• Examples of complex and time consuming tasks : listing $\mathcal{L}$, the frequent itemsets, the associative rules

• Probabilistic and Statistical methods can be used at least for :

1. *Modelling*

2. *Sampling, Bootstrapping*

I.1. Motivations

• Context $\mathcal{C} = (I, J, \mathcal{D})$ (Binary matrix case), $\mathcal{L}$ its concept lattice.

• Examples of complex and time consuming tasks : listing $\mathcal{L}$, the frequent itemsets, the associative rules

• Probabilistic and Statistical methods can be used at least for :

1. *Modelling*
2. *Sampling, Bootstrapping*

I.1. Motivations

• Context $\mathcal{C} = (I, J, \mathcal{D})$ (Binary matrix case), $\mathcal{L}$ its concept lattice.

• Examples of complex and time consuming tasks : listing $\mathcal{L}$, the frequent itemsets, the associative rules

• Probabilistic and Statistical methods can be used at least for :

1. *Modelling*
2. *Sampling, Bootstrapping*

I.2 Modelling

- Model: Mathematical representation of a real context

- *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment:
customer purchases

meteorological measurements

patient diseases ...

- *Observed* context is considered as an *outcome* of the model.
- *Estimating* the parameters of the model from the observations
- Performing *Tests*
- Proposing *Confidence Intervals*
- *Model selection*

- Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling

- Model: Mathematical representation of a real context

- *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment:
customer purchases

meteorological measurements

patient diseases ...

- *Observed* context is considered as an *outcome* of the model.
- *Estimating* the parameters of the model from the observations
- Performing *Tests*
- Proposing *Confidence Intervals*
- *Model selection*

- Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment: customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment:
customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment: customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling 5 / 43

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment: customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.

• *Estimating* the parameters of the model from the observations

• Performing *Tests*

• Proposing *Confidence Intervals*

• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*

Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment:
customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment: customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.2 Modelling

• Model: Mathematical representation of a real context

• *Modelling* a real context (and $\mathcal{L}$, if possible) submitted to a random environment:
customer purchases

meteorological measurements

patient diseases ...

• *Observed* context is considered as an *outcome* of the model.
• *Estimating* the parameters of the model from the observations
• Performing *Tests*
• Proposing *Confidence Intervals*
• *Model selection*

• Some Interest of models: Framework for exact computations (concerning, e.g., $\mathcal{L}$) and *prediction*
Framework for defining the right concepts and not only the empirical concepts

I.3 Sampling 6 / 43

• Considering a given $\mathcal{C}$ or $\mathcal{L}$ as a population and *Sampling*, *Bootstrapping* from $\mathcal{C}$ or from $\mathcal{L}$

Application : *Concept Counting* (estimating $|\mathcal{L}|$), and quickly check the feasibility of an potentially exponential time listing of all concepts

I.3 Sampling

• Considering a given $\mathcal{C}$ or $\mathcal{L}$ as a population and *Sampling*, *Bootstrapping* from $\mathcal{C}$ or from $\mathcal{L}$
Application : *Concept Counting* (estimating $|\mathcal{L}|$), and quickly check the feasibility of an potentially exponential time listing of all concepts

II - MODELS OF RANDOM BINARY CONTEXTS

Bernoulli Model
Hierarchical Bernoulli Models
Indian Buffet
Latent Block Model
Survival Analysis with frailtyness

II.1 Bernoulli(p) Model: Simulation

Illustration In R software :

$p = 0.4$: probability that an entry be equal to 1

$m = 10$ rows (customers, objects), $I = 1, \ldots, m$

$n = 5$ columns (items, attributes) $J = 1, \ldots, n$

$\mathcal{D} = m \times n$ random binary matrix

$\{1, 4\}$ may be closed or not closed, depending on the outcome D.

Illustration In R software :

$p = 0.4$: probability that an entry be equal to 1

$m = 10$ rows (customers, objects), $I = 1, \ldots, m$

$n = 5$ columns (items, attributes) $J = 1, \ldots, n$

$\mathcal{D} = m \times n$ random binary matrix

$\{1, 4\}$ may be closed or not closed, depending on the outcome D.

p_j probability that any entry of column j be equal to 1

The entries of the matrix $\mathcal{D}$ are *independent* r.v.s.

O a subset of objects, A a subset of attributes (itemset)

Probability that the rectangle $O \times A$ be a concept ?

The rectangle $O \times A$ is a concept (maximal rectangle of ones) iff

1. $O \times A$ is filled of ones

and

2. each row of the rectangle $(I - O) \times A$ contains at least one zero

and

3. each column of the rectangle $O \times (J - A)$ contains at least one zero

II.2 Random Concepts

p_j probability that any entry of column j be equal to 1

The entries of the matrix $\mathcal{D}$ are *independent* r.v.s.

O a subset of objects, A a subset of attributes (itemset)

Probability that the rectangle $O \times A$ be a concept ?
The rectangle $O \times A$ is a concept (maximal rectangle of ones) iff

1. $O \times A$ is filled of ones

and

2. each row of the rectangle $(I - O) \times A$ contains at least one zero

and

3. each column of the rectangle $O \times (J - A)$ contains at least one zero

p_j probability that any entry of column j be equal to 1

The entries of the matrix $\mathcal{D}$ are *independent* r.v.s.

O a subset of objects, A a subset of attributes (itemset)

Probability that the rectangle $O \times A$ be a concept ?
The rectangle $O \times A$ is a concept (maximal rectangle of ones) iff

1. $O \times A$ is filled of ones

and
2. each row of the rectangle $(I - O) \times A$ contains at least one zero

and
3. each column of the rectangle $O \times (J - A)$ contains at least one zero

p_j probability that any entry of column j be equal to 1

The entries of the matrix $\mathcal{D}$ are *independent* r.v.s.

O a subset of objects, A a subset of attributes (itemset)

Probability that the rectangle $O \times A$ be a concept ?
The rectangle $O \times A$ is a concept (maximal rectangle of ones) iff

1. $O \times A$ is filled of ones

and
2. each row of the rectangle $(I - O) \times A$ contains at least one zero

and
3. each column of the rectangle $O \times (J - A)$ contains at least one zero

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m - |O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at
Proposition 1
$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m - |O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m-|O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at

Proposition 1

$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m - |O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at
Proposition 1
$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m - |O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)

Let $p_A := \Pi_{j \in A} p_j$.

1.

One row of $O \times A$ is filled with ones with probability (w.p.): p_A

$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.

One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$

each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m-|O|}$

3.

Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$

each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at

Proposition 1

$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

II.3 Computation in the Bernoulli model case

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m-|O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at
Proposition 1
$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

II.3 Computation in the Bernoulli model case

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m - |O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at
Proposition 1
$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|} \Pi_{j \notin A}(1 - p_j^{|O|})$

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)
Let $p_A := \Pi_{j \in A} p_j$.

1.
One row of $O \times A$ is filled with ones with probability (w.p.): p_A
$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.
One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$
each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m-|O|}$

3.
Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$
each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at
Proposition 1
$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

R. Emilion, G. Lévy, Disc. Appl. Math. J, 157, 2945-2957 (2009)

Let $p_A := \Pi_{j \in A} p_j$.

1.

One row of $O \times A$ is filled with ones with probability (w.p.): p_A

$O \times A$ is filled of ones w.p. $p_A^{|O|}$

2.

One row of $(I - O) \times A)$ contains at least one zero w.p. $1 - p_A$

each row of $(I - O) \times A)$ contains at least one zero w.p. $(1 - p_A)^{m-|O|}$

3.

Column j of $O \times (J - A)$ contains at least one zero w.p.: $1 - p_j^{|O|}$

each column of $O \times (J - A)$ contains at least one zero w.p.: $\Pi_{j \notin A}(1 - p_j^{|O|})$

Due to independency we arrive at

Proposition 1

$O \times A$ is a maximal rectangle w.p. $p_A^{|O|}(1 - p_A)^{m-|O|}\Pi_{j \notin A}(1 - p_j^{|O|})$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

• Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O

• As Prob(A is k-closed) $= \sum_{O \in \mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1-p_A)^{m-k} \Pi_{j \notin A} (1-p_j^k)$

• If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and
Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|} (1-p^{|A|})^{m-k} (1-p^k)^{n-|A|}$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

- Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O
- As Prob(A is k-closed) $= \sum_{O \in \mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1-p_A)^{m-k} \Pi_{j \notin A}(1-p_j^k)$

- If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and

Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|}$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

• Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O

• As Prob(A is k-closed) $= \sum_{O \in \mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1-p_A)^{m-k} \Pi_{j \notin A}(1-p_j^k)$

• If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and

Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|}$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

• Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O

• As Prob(A is k-closed) $= \sum_{O \in \mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1 - p_A)^{m-k} \Pi_{j \notin A} (1 - p_j^k)$

• If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and

Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|} (1 - p^{|A|})^{m-k} (1 - p^k)^{n-|A|}$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

• Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O

• As Prob(A is k-closed) $= \sum_{O\in\mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1-p_A)^{m-k} \Pi_{j\notin A}(1-p_j^k)$

• If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and
Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|}$

II.4 Probability of A be closed, in the Bernoulli model case 11 / 43

• Given A, the preceding proposition shows that the probability only depends on the size $|O|$ of O

• As Prob(A is k-closed) $= \sum_{O\in\mathcal{P}(I)}$Prob($O \times A$ is a concept)

and there are $\binom{m}{k}$ subsets O such that $|O| = k$ we arrive at the

Proposition 2
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p_A^k (1-p_A)^{m-k} \Pi_{j\notin A}(1-p_j^k)$

• If $p_j = p$ does not depend on j, we have $p_A = p^{|A|}$ and
Proposition 3
Prob(A is k-closed) $= \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|}$

II.5 Expectation of $|\mathcal{L}|$ in the Bernoulli model case

- Since the number of concepts is equal to the number of k-closed itemsets, we have

$$|L| = \sum_{A \in \mathcal{P}(J)} 1_{A \text{ is } k\text{-closed}}$$

- Taking expectation we get

$$\begin{aligned}\mathbb{E}(|L|) &= \sum_{A \in \mathcal{P}(J)} prob(A \text{ is } k\text{-closed}) \\ &= \sum_{A \in \mathcal{P}(J)} \sum_{k=0}^{m} \binom{m}{k} p^{k|A|} (1-p^{|A|})^{m-k} (1-p^k)^{n-|A|}\end{aligned}$$

and grouping the subsets A with same cardinality we get

Theorem 1

$$\mathbb{E}(|L|) = \sum_{l=0}^{n} \binom{n}{l} \sum_{k=0}^{m} \binom{m}{k} p^{kl} (1-p^l)^{m-k} (1-p^k)^{n-l}$$

II.5 Expectation of $|\mathcal{L}|$ in the Bernoulli model case

- Since the number of concepts is equal to the number of k-closed itemsets, we have

$$|L| = \sum_{A \in \mathcal{P}(J)} 1_{A \text{ is } k\text{-closed}}$$

- Taking expectation we get

$$\begin{aligned} \mathbb{E}(|L|) &= \sum_{A \in \mathcal{P}(J)} prob(A \text{ is } k\text{-closed}) \\ &= \sum_{A \in \mathcal{P}(J)} \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|} \end{aligned}$$

and grouping the subsets A with same cardinality we get

Theorem 1

$$\mathbb{E}(|L|) = \sum_{l=0}^{n} \binom{n}{l} \sum_{k=0}^{m} \binom{m}{k} p^{kl}(1-p^l)^{m-k}(1-p^k)^{n-l}$$

II.5 Expectation of $|\mathcal{L}|$ in the Bernoulli model case

- Since the number of concepts is equal to the number of k-closed itemsets, we have

$$|L| = \sum_{A \in \mathcal{P}(J)} 1_{A \text{ is } k\text{-closed}}$$

- Taking expectation we get

$$\begin{aligned} \mathbb{E}(|L|) &= \sum_{A \in \mathcal{P}(J)} prob(A \text{ is } k\text{-closed}) \\ &= \sum_{A \in \mathcal{P}(J)} \sum_{k=0}^{m} \binom{m}{k} p^{k|A|}(1-p^{|A|})^{m-k}(1-p^k)^{n-|A|} \end{aligned}$$

and grouping the subsets A with same cardinality we get

Theorem 1

$$\mathbb{E}(|L|) = \sum_{l=0}^{n} \binom{n}{l} \sum_{k=0}^{m} \binom{m}{k} p^{kl}(1-p^l)^{m-k}(1-p^k)^{n-l}$$

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.7 μ, σ exact values in the Bernoulli model case

m	n	p	μ	σ	95% CI for L
14	10	0.3	32.48	6.47	$[3, 62]$
15	15	0.9	489.47	373.74	$[1, 2161]$
20	15	0.25	62.78	11.09	$[13, 113]$
20	20	0.65	1945.49	469.16	$[1, 4044]$
25	15	0.85	3758.31	1625.93	$[1, 11030]$
30	12	0.85	1598.66	538.70	$[1, 4008]$

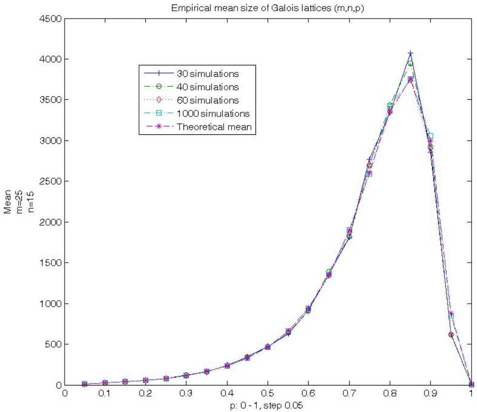

Figure: Estimated and Exact Mean size of Bernoulli Concept Lattices

II.9 Experiments for σ in the Bernoulli model case

m	n	p	σ	S_{300}	95% CI	S_{1000}
14	10	0.3	6.47	5.94	5.03 - 7.94	6.40
15	15	0.9	373.74	321.43	284.39 - 386.57	370.6
20	15	0.25	11.09	11.14	8.92 - 12.96	11.04
20	20	0.65	469.16	469.65	433.60 - 497.42	468.25
25	15	0.85	1625.93	1688.60	1493.90 - 1743.60	1626.20
30	12	0.85	538.70	549.30	503.96 - 566.11	535.79

II.10 Hierarchical Bernoulli context

R.E., *Selected contributions in Data Analysis and Classification*, 247-259, Springer, 2007

Context: $m \times r$ random binary matrix $\mathcal{C}$
U a latent class variable $\in \{1, \ldots, K\}$ over the individuals

$$\begin{cases} q = (q_1, \ldots, q_K) & \sim & Dirichlet(\gamma_1, \ldots, \gamma_K) \\ U \in \{1, \ldots, K\} : P(U = u|q) & = & q_u \\ \mathcal{C}|_{U=u,q} & \sim & \otimes_{j=1}^{r} B(p_{u,j}) \\ \mathcal{C}|_q & \sim & \sum_{u=1}^{K} q_u \otimes_{j=1}^{r} B(p_{u,j}) \end{cases}$$

Y. W. Teh, D. Gorur, Z. Ghahramani

$Beta - Bernoulli$ Context: $m \times r$ random binary matrix $\mathcal{C}$

$$\begin{cases} p_1, \ldots, p_r & \overset{i.i.d.}{\sim} & Beta(\frac{\alpha}{r}, 1) \\ \mathcal{C}_{ij} | p_1, \ldots, p_r & \overset{ind}{\sim} & Bernoulli(p_j) \end{cases}$$

Limit:

Step 1: Customer 1 chooses $K^{(1)}$ different items, where $K^{(1)} \sim Poisson(\alpha)$

Step 2: Customer 2 arrives and chooses to enjoy each of the items already chosen with probability $1/2$. In addition, he chooses $K^{(2)}$ new items , where $K^{(2)} \sim Poisson(\alpha/2)$

Steps 3 through N: The ith customer arrives and chooses to enjoy each of the items already chosen with probability m_{ki}/i, where m_{ki} is the number of customers who have chosen the kth item before the ith customer. In addition, the ith customer chooses $K^{(i)} \sim Poisson(\alpha/i)$ new items.

II.12 Latent Block model

G. Govaert, M. Nadif, Co-clustering.
Context: $m \times r$ random binary matrix $\mathcal{C}$
$\mathcal{Z}$ set of partitions of I into g subsets
$\mathcal{W}$ set of partitions of J into h subsets

$$f(\mathcal{C};\theta) = \sum_{(z,w)\in\mathcal{Z}\times\mathcal{W}} p(z;\theta)p(w;\theta) \prod_{i,j,k,l} Bernoulli(c_{i,j};\alpha_{k,l})^{z_{i,k}w_{j,l}}$$

II.13 Recurrent events with frailty models

A disease (crisis), or a failure, appearing several times.
Context: $m \times n$ random binary matrix $\mathcal{C}$
$1 : m$ set of patients
$1 : n$ Observation times (deterministic right censoring)
or locations
If the disease starts at time j for patient i then $\mathcal{C}_{i,j} = 1$ else $\mathcal{C}_{i,j} = 0$.
X_i a random variable representing frailty of patient i
The interarrival times (between two diseases) *given* X_i are i.i.d.
Simple case: $X_i \overset{i.i.d.}{\sim} \gamma$
Non Parametric Bayesian case $X_i|P \sim P, P \sim Dirichlet(c\gamma)$
In the case of locations: spatial dependance.
A. Adekpedjou, R. Emilion, S. Niang (in progress)

III - Sampling

- Sampling in a large set

- Markov Chains in $\mathcal{L}$

- Sampling and Counting concepts

- Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

- At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

- When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

- Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

- More general problem : $Prob(X = l) \ \propto \ v(l)$ a function of l which no need to sum up to 1.

- Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

- At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

- When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

- Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

- More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is unknown

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \ldots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

III.2 Markov Chains

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Markov triying to generalizing for non i.i.d. r.v.s found his famous definition:
- $X_0, \dots, X_n, \dots : \Omega \to \mathcal{L}$ (the state space)

$$P(X_{n+1} = x_{n+1} | X_n, \dots, X_0) = P(X_{n+1} = x_{n+1} | X_n)$$

- The chain 'forgets' its past.
- Transitions:

$$P(X_{n+1} = x_{n+1} | X_n = x_{n+1}) = p(x_{n+1}, x_n)$$

- $P(X_0 = x_0)$ initial distribution.
- Simulation in R software.

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

III.3 Markov Chains and sampling

- Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

- Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n\rightarrow+\infty}P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

III.3 Markov Chains and sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

III.4 Markov Chains in a graph 25 / 43

- The chain moves from on node to its neighbourhood nodes
- Define the neighbourhood nodes of a node
- Define the transitions

III.4 Markov Chains in a graph

- The chain moves from on node to its neighbourhood nodes
- Define the neighbourhood nodes of a node
- Define the transitions

III.4 Markov Chains in a graph

- The chain moves from on node to its neighbourhood nodes
- Define the neighbourhood nodes of a node
- Define the transitions

Mario Boley et. al. SIAM DM 2010

Context $\mathcal{C} = (A, O, \mathcal{D})$
$O[\,] : \mathcal{P}(A) \to \mathcal{P}(O)$ extent mapping
$A[\,] : \mathcal{P}(O) \to \mathcal{P}(A)$ intent mapping
$\Phi = A \; o \; O$ and $\Psi = O \; o \; A$ the closure mappings
Concepts $C = (I, E), I \in \mathcal{P}(A), E \in \mathcal{P}(O)$
I' is a $\Phi-$neighbourhood of I if there exists an $a \in A$ such that $\Phi(I \cup a) = I'$
E' is a $\Psi-$neighbourhood of E if there exists an $o \in O$ such that $\Psi(E \cup E) = E'$

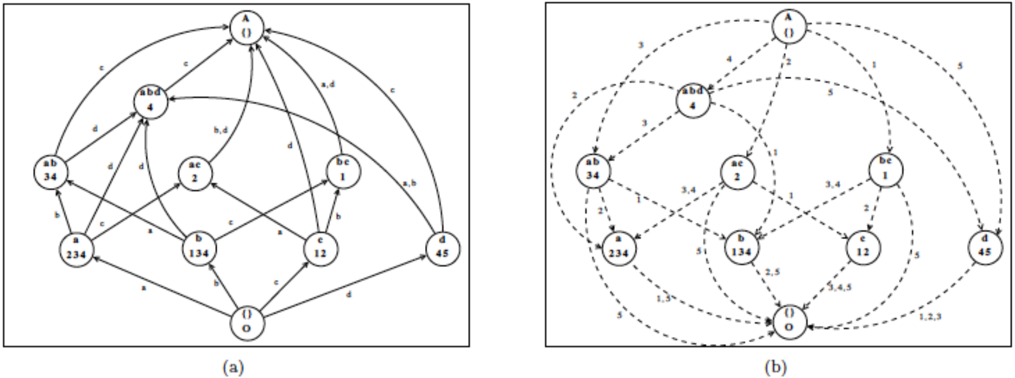

Figure: Both graphs are used

$$q(C,C') = \begin{cases} |G_\phi(I,I')| / (2|A|), & \text{if } C \prec C' \\ |G_\psi(E,E')| / (2|O|), & \text{if } C \succ C' \\ |I| / (2|A|) + |E| / (2|O|), & \text{if } C = C' \end{cases}$$

Figure: Transitions between two concepts

$$p(C,C') = \begin{cases} q(C,C')\min\{\alpha\frac{\pi(C')}{\pi(C)},1\}, & \text{if } q(C,C')>0 \\ 0, & \text{otherwise} \end{cases}$$

where $\alpha = q(C',C)/q(C,C')$. This is the Metropolis-

Figure: MH- Transitions between two concepts

Algorithm 1 Metropolis-Hastings Concept Sampling

Input : context $(A, O, \mathcal{D})$, number of iterations s, oracle of map $f : \mathcal{C} \to \mathbb{R}_+$
Output : concept $\langle I, E \rangle$

1. **init** $\langle I, E \rangle \sim u(\{\top, \bot\})$ and $i \leftarrow 0$
2. $i \leftarrow i + 1$
3. **draw** $d \sim u(\{\text{up}, \text{down}\})$
4. **if** $d = \text{up}$ **then**
5.   **draw** $a \sim u(A)$
6. $\langle I', E' \rangle \leftarrow \langle \phi(I \cup \{a\}), O[\phi(I \cup \{a\})] \rangle$
7. $\alpha \leftarrow (|G_\psi(E', E)|\,|A|) \,/\, (|G_\phi(I, I')|\,|O|)$
8. **else**
9.   **draw** $o \sim u(O)$
10. $\langle I', E' \rangle \leftarrow \langle A[\psi(E \cup \{o\})], \psi(E \cup \{o\}) \rangle$
11. $\alpha \leftarrow (|G_\phi(I', I)|\,|O|) \,/\, (|G_\psi(E, E')|\,|A|)$
12. **draw** $x \sim u([0, 1])$
13. **if** $x < \alpha f(I') \,/\, f(I)$ **then** $\langle I, E \rangle \leftarrow \langle I', E' \rangle$
14. **if** $i = s$ **then return** $\langle I, E \rangle$ **else goto** 2

Figure: MH- Algorithm

III.9 Application of sampling: concept counting in $\mathcal{L}$ 31 / 43

Mario Boley et. al. SIAM DM 2010

Context $\mathcal{C} = (A, O_n = 1 : n, \mathcal{D})$
$O_n[\,] : \mathcal{P}(A) \to \mathcal{P}(O_n)$ extent mapping
$A[\,] : \mathcal{P}(O_n) \to \mathcal{P}(A)$ intent maping
$\Phi_n = A \; o \; O_n$ and $\Psi_n = O_n \; o \; A$ the closure mappings
For any $I \in \mathcal{P}(A)$, we have $I \subseteq \Phi_{n+1}(I) \subseteq \Phi_n(I)$ and thus $\mathcal{L}_n \subseteq \mathcal{L}_{n+1}$

Mario Boley et. al. SIAM DM 2010

Context $\mathcal{C} = (A, O_n = 1 : n, \mathcal{D})$
$O_n[\,] : \mathcal{P}(A) \to \mathcal{P}(O_n)$ extent mapping
$A[\,] : \mathcal{P}(O_n) \to \mathcal{P}(A)$ intent maping
$\Phi_n = A \; o \; O_n$ and $\Psi_n = O_n \; o \; A$ the closure mappings
For any $I \in \mathcal{P}(A)$, we have $I \subseteq \Phi_{n+1}(I) \subseteq \Phi_n(I)$ and thus $\mathcal{L}_n \subseteq \mathcal{L}_{n+1}$

Mario Boley et. al. SIAM DM 2010

$\mathcal{L}_0 = \{A\}$,$|\mathcal{L}_0| = 1$
Sample r concepts in $\mathcal{L}_{n+1}$
c of them belong to $\mathcal{L}_n$ (does not contain $n+1$)
Estimate $\frac{|\mathcal{L}_n|}{|\mathcal{L}_{n+1}|}$ by $\frac{c}{r}$
If $O = 1 : m$ then use that

$$|\mathcal{L}| = |\mathcal{L}_m| = \frac{|\mathcal{L}_m|}{|\mathcal{L}_{m-1}|} \cdots \frac{|\mathcal{L}_1|}{|\mathcal{L}_0|}$$

to estimate $|\mathcal{L}|$

IV - Pointwise convergence of empirical RCLs

i.i.d. case

Markov chain case

IV.1 Random empirical Intents/Extents

(Ω, P) a probability space, $\mathcal{F}$ a countable semilattice
Examples: $\mathcal{F} = \mathcal{P}(A)$ A finite set, binary tree, set of subsets of $\mathbb{R}$ that are countable or their complementary is countable.
$X : \Omega \longrightarrow \mathcal{F}$ a random variable
Support of $X = Supp_X$:
any subset S of F such that $P(X \in S) = 1$
Observations: $X_1(\omega), \ldots, X_n(\omega), \ldots,$
For any $d \in \mathcal{F}$

$$g_n(d) = g_{X_1(\omega),\ldots,X_n(\omega)}(d) = \{i = 1, \ldots, n : d \leq X_i(\omega)\} \tag{1}$$

$$k_n(d) = f_n(g_n(d)) = \bigwedge_{u = X_1(\omega),\ldots,X_n(\omega), d \leq u} u \tag{2}$$

(Ω, P) a probability space, $\mathcal{F}$ a countable semilattice
Examples: $\mathcal{F} = \mathcal{P}(A)$ A finite set, binary tree, set of subsets of $\mathbb{R}$ that are countable or their complementary is countable.
$X : \Omega \longrightarrow \mathcal{F}$ a random variable
Support of $X = Supp_X$:
any subset S of F such that $P(X \in S) = 1$
Observations: $X_1(\omega), \ldots, X_n(\omega), \ldots,$
For any $d \in \mathcal{F}$

$$g_n(d) = g_{X_1(\omega),\ldots,X_n(\omega)}(d) = \{i = 1, \ldots, n : d \leq X_i(\omega)\} \quad (1)$$

$$k_n(d) = f_n(g_n(d)) = \bigwedge_{u = X_1(\omega),\ldots,X_n(\omega), d \leq u} u \quad (2)$$

IV.2 Empirical Intents Pointwise convergence

Theorem

(R.E., Springer 2007)

If $\mathcal{F}$ is a countable σ-semilattice

$X_1(\omega), \ldots, X_n(\omega), \ldots$, *i.i.d. sample of X*

Then

$g_n(d) \uparrow g_\infty(d) = \{i = 1, ..., n, ... : d \leq X_i(\omega)\}$

$k_n(d) \downarrow k_\infty(d) = \bigwedge_{u \in Supp_X, u \leq d} u$: *deterministic limit*

- (g_∞, k_∞) is a GC, $k_\infty(\mathcal{F})$ deterministic lattice generated by $Supp_X$.
- Induces the CL of a discrete r.v.
- $k_\infty(d)$: deterministic ideal concept (intent)
- Does not depend on the observations. Is the limit of empirical intents
- Streaming. Learning.

IV.3 Pointwise convergence, Markov chain

Theorem

(Emilion 2011)

If $\mathcal{F}$ is a countable σ-semilattice
$X_1(\omega), \ldots, X_n(\omega), \ldots$, *recurrent Markov chain with inv. meas.* μ
Then

$$g_n(d) \uparrow g_\infty(d) = \{i = 1, ..., n, ... : d \leq X_i(\omega)\}$$

$k_n(d) \downarrow k_\infty(d) = \bigwedge_{u \in Supp_\mu, u \leq d} u$: *deterministic limit*

(g_∞, k_∞) is a GC which induces a CL: the CL of a discrete Markov Chain

Since $Supp_X \subseteq \mathcal{F}$ is countable, $\{X_1(\omega), \ldots, X_n(\omega), \ldots\} = Supp_X$ for a.a. ω
Indeed $X_i(\omega) \in Supp_X$ as $P(X_i \in Supp_X) = P(X_i \in Supp_X) = 1$
Conversely if $d \in Supp_X$, by the Large Number Law, $X_i = d$ for an infinity of i.

$$k_n(d) \downarrow k_\infty(d) = \bigwedge_{u \in \{X_1(\omega), \ldots, X_n(\omega), \ldots\}, d \leq u} u$$

$$= \bigwedge_{u \in Supp_X, d \leq u} u : \text{deterministic limit}$$

LNL also holds for a recurrent Markov Chain which has an invariant measure.

V - Experiments

Bernoulli case
Mushroom case, Regression

V.1 Bernoulli context, CL size distribution

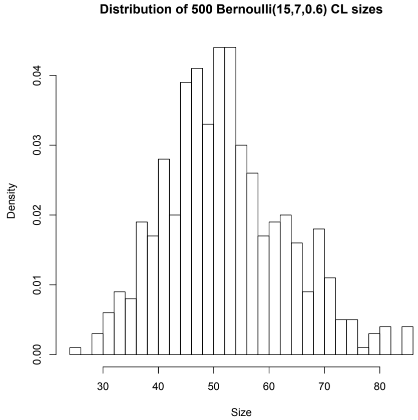

Figure: Distribution of Bernoulli CL Size

V.2 Bernoulli context, CL depth distribution

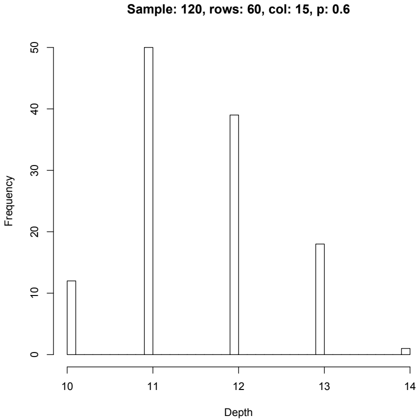

Figure: Distribution of Bernoulli CL Depth

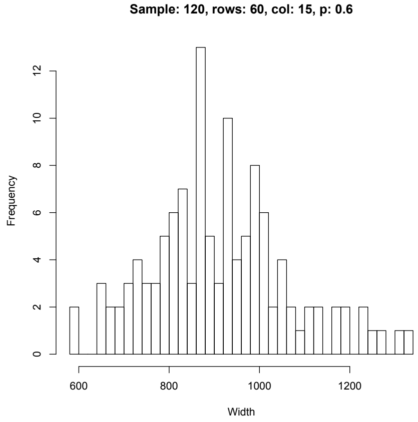

Figure: Distribution of Bernoulli CL Width

V.4 Bernoulli context, CL Size Depth Width PCA

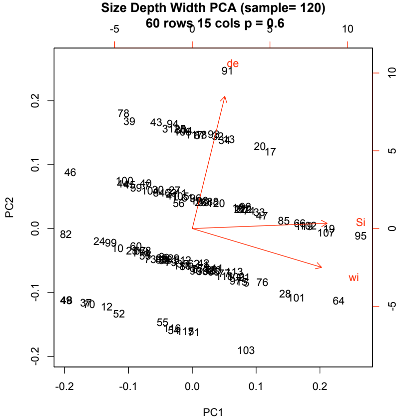

Figure: PCA on Bernoulli CLs

V.5 Mushroom context, Regression : Number of concepts w.r.t. number of ones 43 / 43

Figure: Linear regression, concepts of the r first rows, $r = 1000,1100, ..., 8124$