Approximation in Value-Based Potentials

Retamero O. Cano A. Gómez-Olmedo M.

University of Granada

New Bridges between Mathematics and Data Science, 2021

Table of contents

Prerequisites

Probabilistic graphical models

Probabilistic graphical models (PGMs) are graph-based representations able to encode the information of problems under uncertainty in a compact and efficient way.

PGMs are defined by two parts:

- A **qualitative component** — graph —, where nodes represent the variables of the domain and edges the dependencies among them.
- A **quantitative component**, consisting of a set of functions quantifying such dependencies.

Prerequisites

Probabilistic graphical models

More formally, a *PGM* is defined by its three elements $\langle \mathbf{X}, P, \mathcal{G} \rangle$ where:

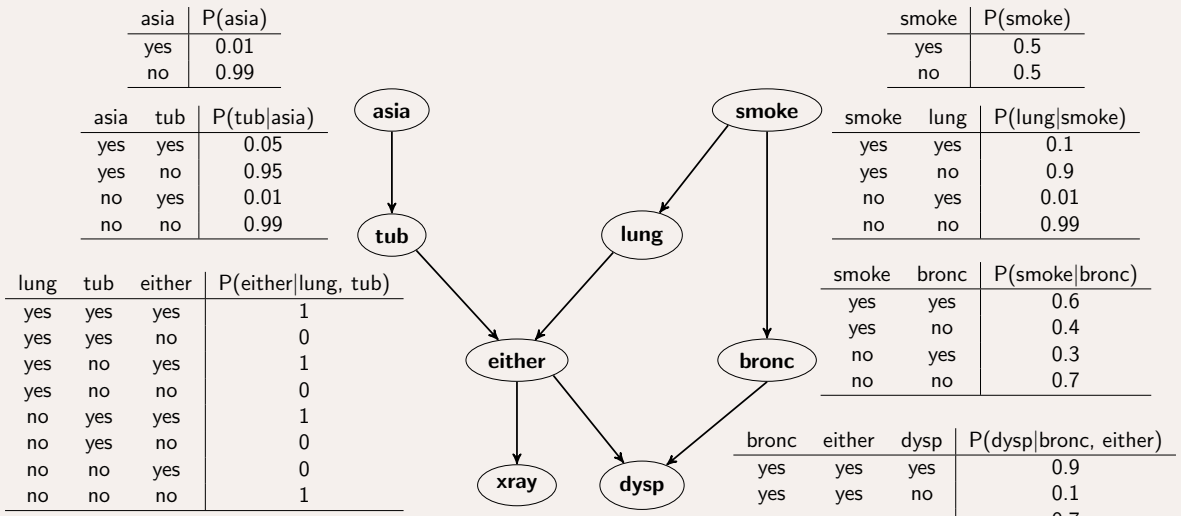
- $\mathbf{X} = \{X_1, X_2, ..., X_N\}$ is the set of variables involved in the problem,
- with the joint probability distribution $P(\mathbf{X})$,
- and $\mathcal{G}$ is a graph which represents the dependency (and independency) between the variables.

Prerequisites

Probabilistic graphical models

Example

An example of a *PGM*, more specifically of a Bayesian network, is the well known *asia* network:



asia	P(asia)
yes	0.01
no	0.99

smoke	P(smoke)
yes	0.5
no	0.5

asia	tub	P(tub|asia)
yes	yes	0.05
yes	no	0.95
no	yes	0.01
no	no	0.99

smoke	lung	P(lung|smoke)
yes	yes	0.1
yes	no	0.9
no	yes	0.01
no	no	0.99

lung	tub	either	P(either|lung, tub)
yes	yes	yes	1
yes	yes	no	0
yes	no	yes	1
yes	no	no	0
no	yes	yes	1
no	yes	no	0
no	no	yes	0
no	no	no	1

smoke	bronc	P(smoke|bronc)
yes	yes	0.6
yes	no	0.4
no	yes	0.3
no	no	0.7

either	xray	P(xray|either)
yes	yes	0.05
yes	no	0.95
no	yes	0.01
no	no	0.99

bronc	either	dysp	P(dysp|bronc, either)
yes	yes	yes	0.9
yes	yes	no	0.1
yes	no	yes	0.7
yes	no	no	0.3
no	yes	yes	0.8
no	yes	no	0.2
no	no	yes	0.1
no	no	no	0.9

Figure: Asia network.

Prerequisites

Representation potentials

In *PGMs* over discrete domains, such as Bayesian networks (*BN*) or influence diagrams, the potentials are traditionally represented with **tables or unidimensional arrays (1DA)**.

Prerequisites

Representation potentials

Example

Let us consider the variables X_1, X_2, and X_3, with 2, 3 and 2 states respectively. Then $\phi(X_1, X_2, X_3)$ is a potential defined on such variables, representing the conditional distribution $P(X_3|X_1, X_2)$, and whose values are presented in the following figure:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

Prerequisites

Representation potentials

Example

The *1DA* representation for our considered example has 12 entries and looks as follows:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

0	1	2	3	4	5	6	7	8	9	10	11
0.1	0.9	0.5	0.5	0.0	1.0	0.8	0.2	0.2	0.8	0.9	0.1

Figure: $\phi(X_1, X_2, X_3)$ as a *1DA*.

Prerequisites

Representation potentials

Another widely used representation would be **probability trees (PTs)**. A *PT* is a directed and labeled tree in which each internal node represents a variable and each leaf node is a non-negative real number. From each internal node there will be as many arcs as states the variable labeling the node has.

Prerequisites

Representation potentials

Example

The same potential given in the previous example is presented as a PT with a size of 21 nodes (12 leaves and 9 internal nodes).

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

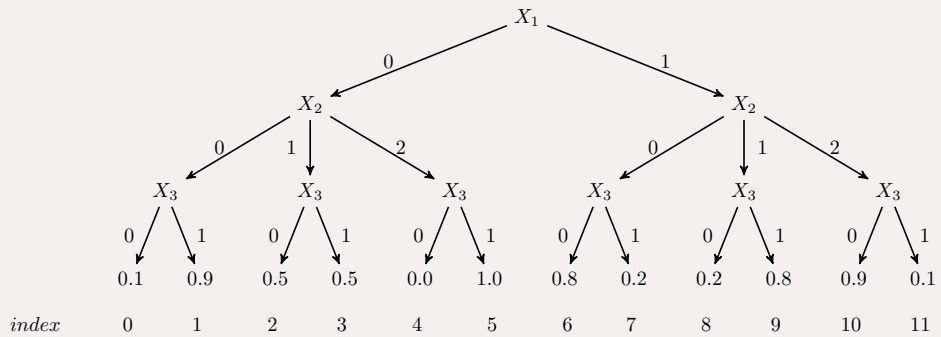


Figure: $\phi(X_1, X_2, X_3)$ as PT.

Prerequisites

Representation potentials

PTs can take advantage of contex-specific independencies, joining equal values into a single one. This operation is called **pruning**.

Prerequisites

Representing potentials

Example

The potential in previous examples presents a *context-specific independence* that allows reducing its size: The value for $X_1 = 0, X_2 = 1$ is 0.5, regardless of the value of X_3. Once the *pruning* is complete, the result is a *PPT* (*pruned PT*) of size 19.

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

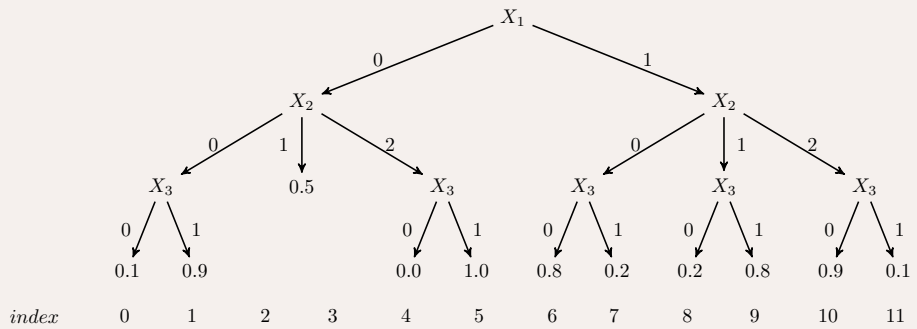


Figure: $\phi(X_1, X_2, X_3)$ as *PPT*.

Value-Based Potentials

Introduction to VBPs

The four alternative structures aptly named **Value-Based Potentials (VBPs)** from being based on the values themselves and not on the position they appear, can be divided in two categories:

- Structures driven by values, using dictionaries in which the keys will be values. Called **value-driven with grains (VDG)** and **value-driven with indices (VDI)**.
- Structures driven by indices, whose keys are indices. Called **index-driven with indices (IDP)** and **index-driven with map (IDM)**.

Value-Based Potentials

Introduction to VBPs

Note that the four alternatives present a default value which will be returned if the search fails.

This default value is set to 0 in our case, but could be any other, considering the most space-saving option might be the most repeated one.

Value-Based Potentials

Introduction to VBPs

Example

Let us present our studied example as a simple set of probabilities, considering the values taken by all the possible configurations of the variables. The correspondent indices can be found between brackets.

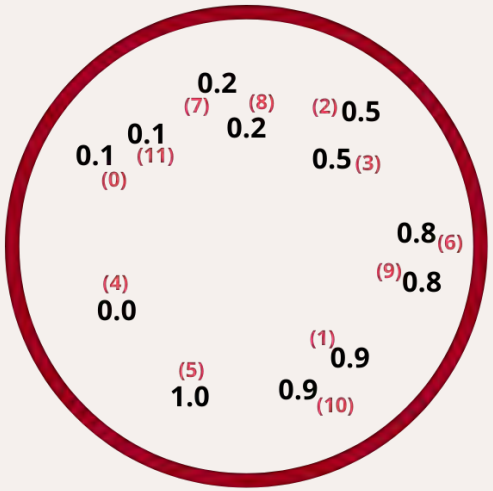


Figure: Visual idea of the set of the values in ϕ.

Value-Based Potentials

Introduction to VBPs

Example

The main idea behind *VBPs* is to take advantage of the repetition of values in the potential. We can easily observe some of the values are equal, consecuently we could gather them into sets and save memory space:

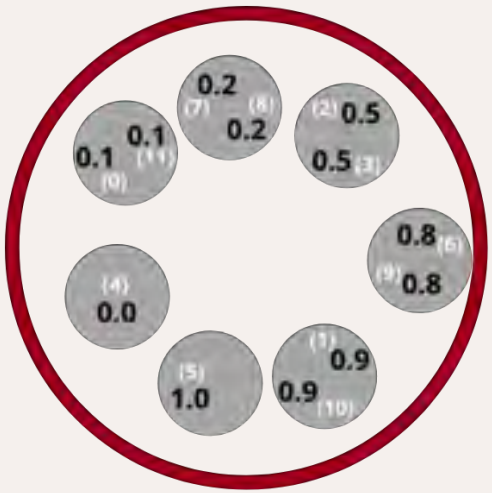


Figure: Visual idea of the equal values joined in the set of the values in ϕ.

Value-Based Potentials

VDG: value-driven with grains

Example

The potential $\phi(X_1, X_2, X_3)$ used in our example will be represented as *VDG*, as follows:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

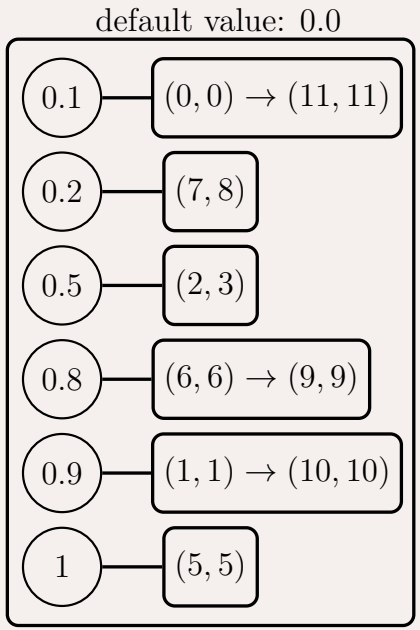


Figure: $\phi(X_1, X_2, X_3)$ as VDG.

Value-Based Potentials

VDI: value-driven with indices

Example

The potential $\phi(X_1, X_2, X_3)$ used in our example will be represented as *VDI*, as follows:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

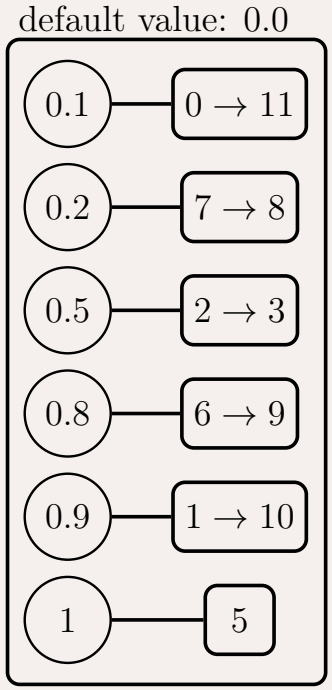


Figure: $\phi(X_1, X_2, X_3)$ as VDI.

Value-Based Potentials

IDP: index-driven with pairs

Example

The potential $\phi(X_1, X_2, X_3)$ used in our example will be represented as *IDP*, as follows:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

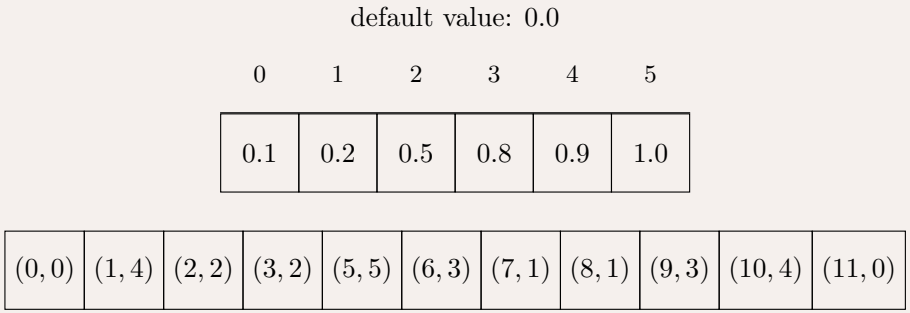


Figure: $\phi(X_1, X_2, X_3)$ as IDP.

Value-Based Potentials

IDM: index-driven with map

Example

The potential $\phi(X_1, X_2, X_3)$ used in our example will be represented as *IDM*, as follows:

index	x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	0	0.1
1	0	0	1	0.9
2	0	1	0	0.5
3	0	1	1	0.5
4	0	2	0	0.0
5	0	2	1	1
6	1	0	0	0.8
7	1	0	1	0.2
8	1	1	0	0.2
9	1	1	1	0.8
10	1	2	0	0.9
11	1	2	1	0.1

Figure: Table representation of the potential $\phi(X_1, X_2, X_3)$ as a mapping that assigns a numerical value to each configuration.

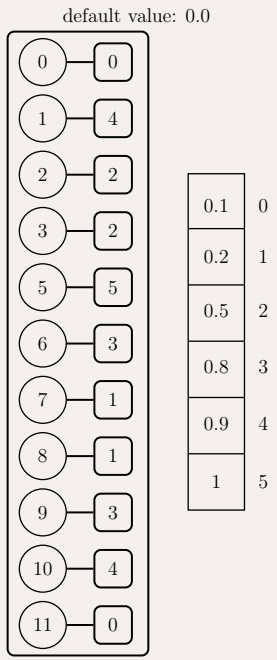


Figure: $\phi(X_1, X_2, X_3)$ as IDM.

Approximation of VBPs

Motivation

Potentials present a big limitation, their size grows exponentially as the number of variables increases. Therefore, in the case of complex models, doing inference or even dealing with them might be computationally infeasible. *PPTs* already could present a good improvement when values appear consecutively, but what if that is not the case?

Approximation of VBPs

Motivation

VBPs were tested with several *BNs* from the *bnlearn* package and *UAI*'s inference competitions, proving they imply saving memory space. For this purpose and as well as it has been done before in the case of *PTs*, we go a step further and present the approximation of such *VBPs*.

Note, the chosen *VBP* alternative will be **VDI**, as it performed the best of the four in most of the studied *BNs*.

Approximation of VBPs

Algorithm

1. Create the new *VDI* candidate structures to become the chosen approach. Each one will contain one combination of two contiguous dictionary entries. The key of the combined cell will be the average weight of both keys and the associated indices, the union of the two lists of indices.
2. Calculate the Kullback-Leibler divergences between the original potential ϕ and each one of the possibilities.
3. The chosen alternative will be the one whose Kullback-Leibler divergence has the smallest value, accordingly the loss of information will be minimum.
4. Repeat the process until the selected stopping condition.

Value-Based Potentials

Algorithm

Example

In the visual idea, we can observe the 7 different possible combinations of groups of values:

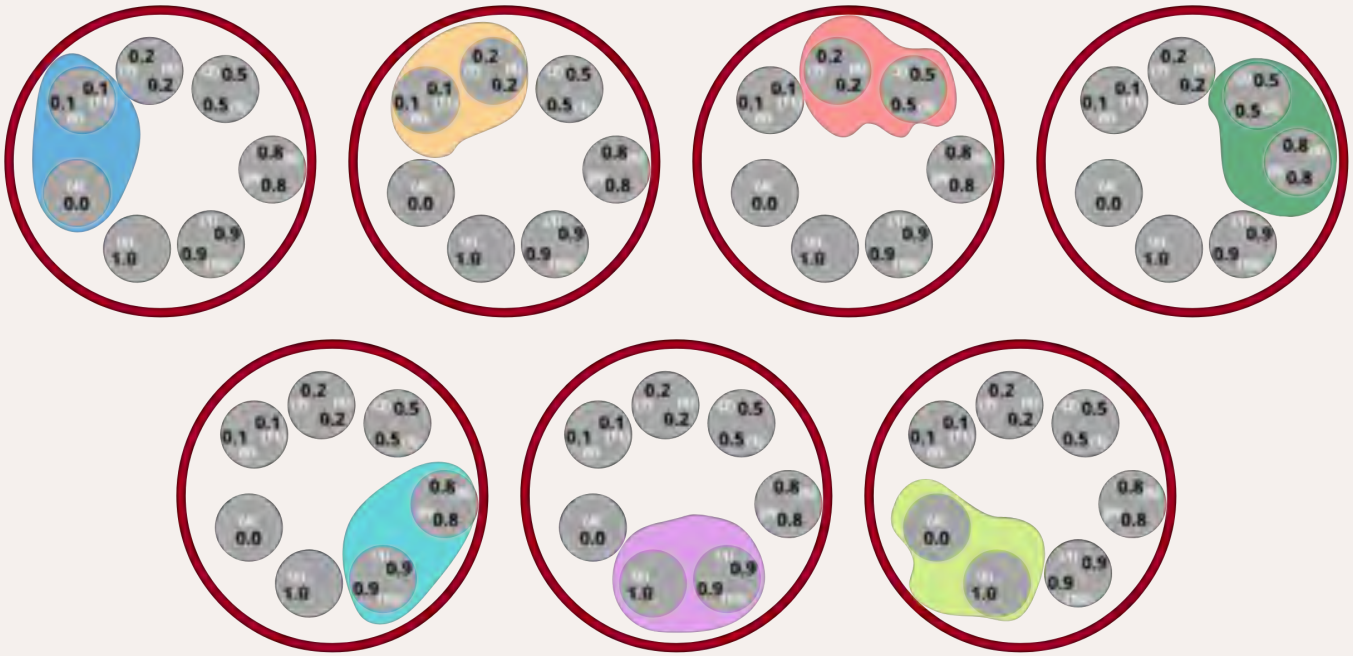


Figure: Visual idea of the possible combination of equal values joined in ϕ.

Approximation of VBPs

Fundamentals

To measure the distance between the original potential ϕ and the approximated *VDI*, V, we will use the Kullback-Leibler divergence between their normalized *potentials* ($\overline{\phi}$ and $\overline{V}$), defined as follows:

$$D\left(\phi, V\right) = \sum_{x_i \in \Omega_{X_i}} \overline{\phi}\left(x_i\right) \; log \; \frac{\overline{\phi}\left(x_i\right)}{\overline{V}\left(x_i\right)}$$

Approximation of VBPs

Fundamentals

Proposition

The information loss obtained by joining the sets of configurations A_l and A_r into the set A in the data structure V, can be calculated as follows:

$$I\,(V,\ A,\ A_l,\ A_r) = \frac{1}{sum\ (\phi)} \left[log\ \frac{|A|}{sum\ (\phi^{R(A)})}\ sum\ (\phi^{R(A)}) - log\ \frac{|A_l|}{sum\ (\phi^{R(A_l)})}\ sum\ (\phi^{R(A_l)}) - log\ \frac{|A_r|}{sum\ (\phi^{R(A_r)})}\ sum\ (\phi^{R(A_r)}) \right]$$

where $sum(\phi)$ denotes the addition of all the values of the potential ϕ.

Approximation of VBPs

Example

Example

Following with our studied example, we consider the first iteration of the process. As **Step 1** of the algorithm, the possible approximations are:

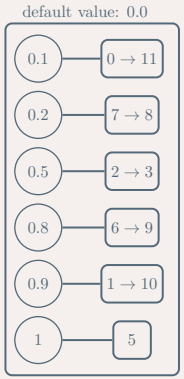


Figure: Original ϕ as VDI.

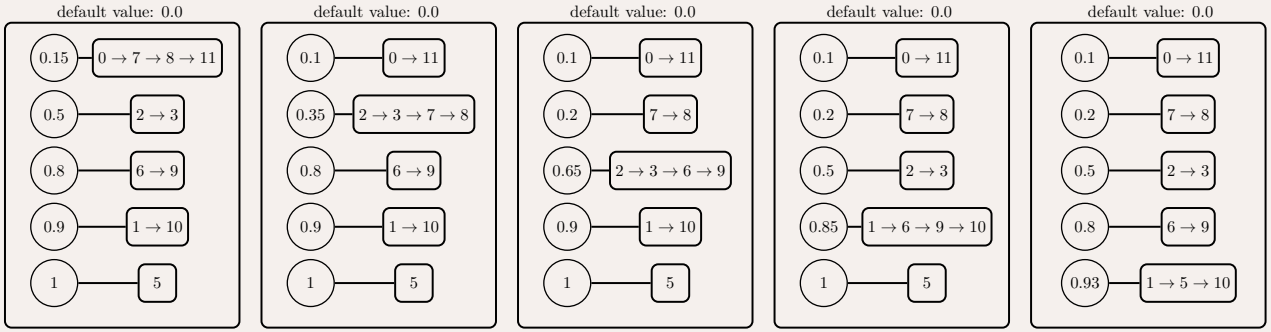


Figure: Possible approximated structures on the first step of the algorithm

Approximation of VBPs

Example

Example

Step 2: Their Kullback-Leibler divergences to the original potential ϕ are:

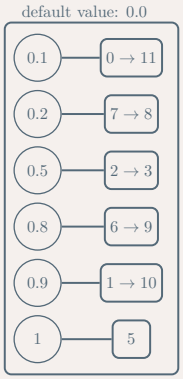
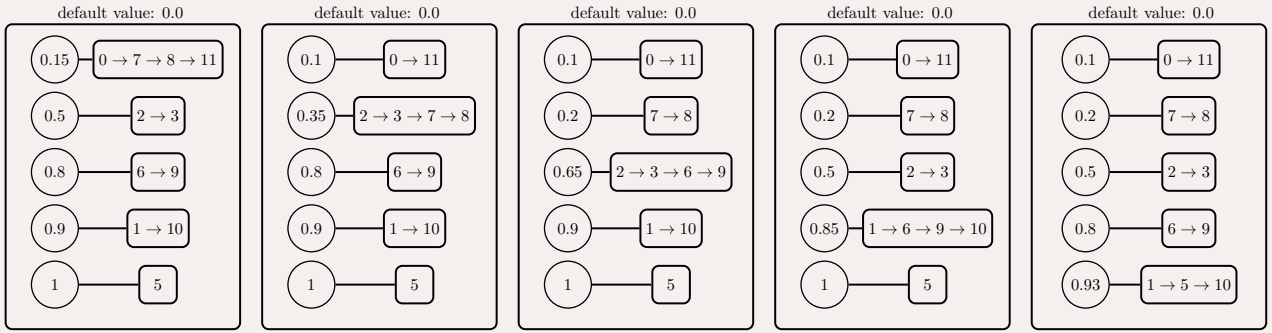


Figure: Original ϕ as VDI.



Correspondent Kullback-Leibler divergences:

0,002459	0,009614	0,005056	0,000426	0,000255

Figure: Possible approximated structures on the first step of the algorithm and their correspondent distances to the original potential ϕ

Approximation of VBPs

Example

Example

Step 3: Therefore, we can select the option producing the least loss of information, i. e. the minimum Kullback-Leibler divergence:

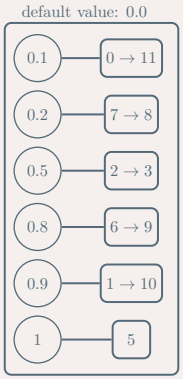
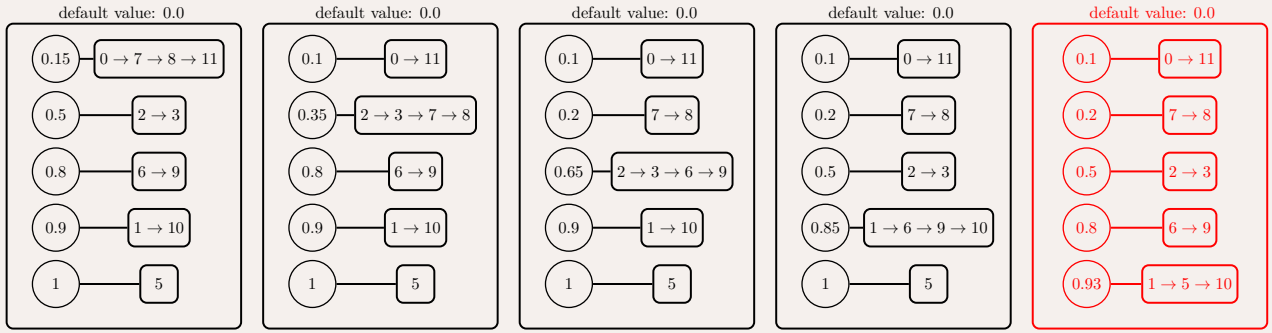


Figure: Original ϕ as VDI.



Correspondent Kullback-Leibler divergences:

0,002459 0,009614 0,005056 0,000426 0,000255

Figure: Chosen approximated structure on the first iteration

Approximation of VBPs

Example

Example

If the chosen stopping criterion is not satisfied the process is repeated. The table on the left represents the Kullback-Leibler divergence and the resultant approximated *VDI* structure is shown on the right:

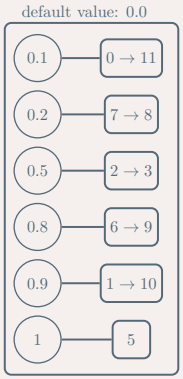


Figure: Original ϕ as VDI.

KEY COMBINATION	KULLBACK – LEIBLER DIVERGENCE
0,1 & 0,2	0,0027151309
0,2 & 0,5	0,0098700472
0,5 & 0,8	0,0053121331
0,8 & 0,93	0,0011427898

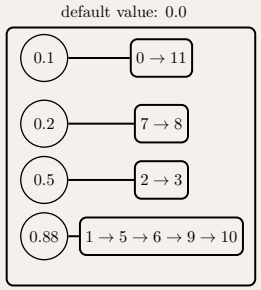


Figure: Distance of the potential ϕ to the possible approximations calculated on the second iteration and chosen option

Approximation of VBPs

Example

Example

And a third iteration will produce the following results. The table on the left represents the Kullback-Leibler divergence and the resultant and final approximated *VDI* structure is shown on the right:

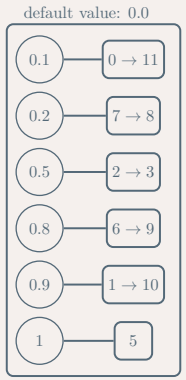


Figure: Original ϕ as VDI.

KEY COMBINATION	KULLBACK – LEIBLER DIVERGENCE
0,1 & 0,2	0,0038579207
0,2 & 0,5	0,0096144568
0,5 & 0,88	0,0105494791

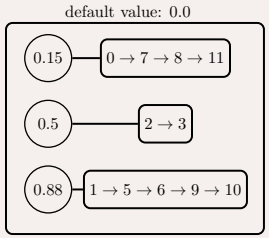


Figure: Distance of the potential ϕ to the possible approximations calculated on the third iteration and chosen option

Thank you for attending!