Institute for Risk and Uncertainty

WPMSIIP 2018, Oviedo, Spain

Approximate Inference methods for Advanced Bayesian networks

Presenter: Hector Diego Estrada Lugo

Second year PhD student

Dr E. Patelli, Dr M. de Angelis

Institute for Risk and Uncertainty

University of Liverpool, U.K.

CONACYT
Consejo Nacional de Ciencia y Tecnología

30/07/2018

Motivation

Bayesian nets methodology

- Different data sets implemented

Bayesian Update (Inference)

- Method 1: Naïve approximate inference
- Method 2: Approximate LP inference

Case study

Results

Conclusions

Motivation

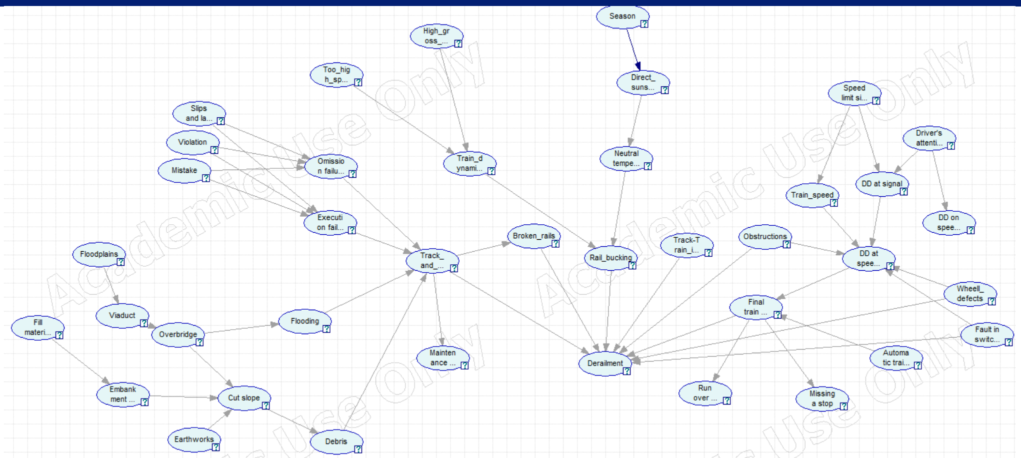
- Risk factors representation and uncertainty quantification is complicated in large infrastructure projects.

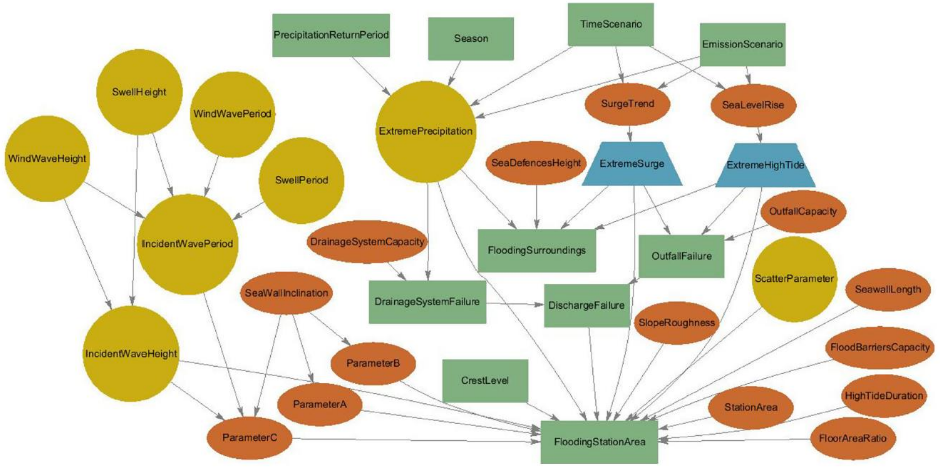
- Multidisciplinary nature needs a standard tool to facilitate risk communication.

- Risk management must take into consideration the uncertainty factors in the system.

Motivation

- Probabilistic graphical models (like Bayes nets), effective mathematical tool for uncertainty quantification and system modelling.
- Allows to capture variable dependencies of complex systems.
- Inference computation is a key method to update outcomes in Bayesian networks.
- Reliable method of inference computation in Credal networks is necessary.





Enhanced Bayesian Network[*].

[*]S. Tolo, E. Patelli, and M. Beer, “Robust vulnerability analysis of nuclear facilities subject to external hazards,” *Stoch. Environ. Res. Risk Assess.*, vol. 31, no. 10, pp. 2733-- 2756, 2017.

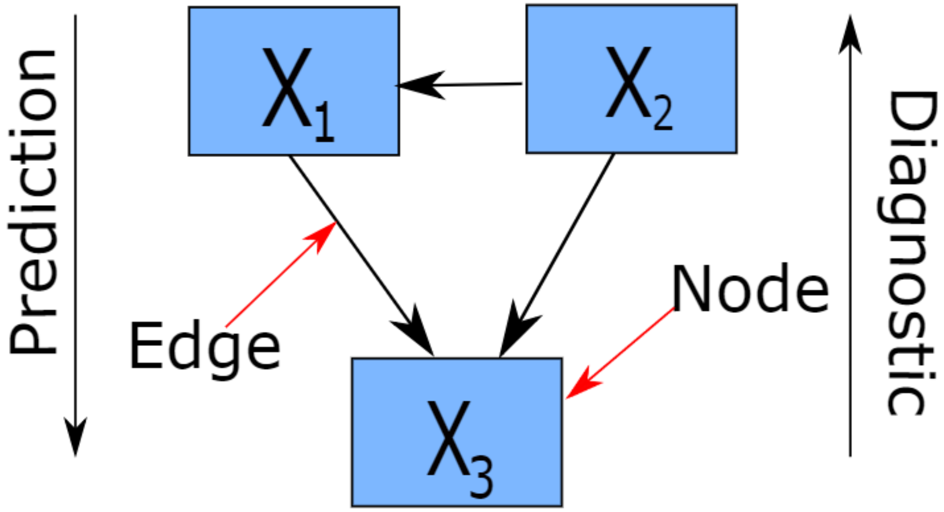
Bayesian Networks

A Bayesian network is a probabilistic graphical model to study and analyse the dependencies of components (random variables) that make up a system.

- The Joint Probability Distribution (JPD) describes entirely network's dependability,

$$P(x_i) = \prod_{i=1}^{n} P(x_i|\pi_i)$$

- By introducing evidence, infer updated outcomes.
- Intuitive and relatively easy to implement.



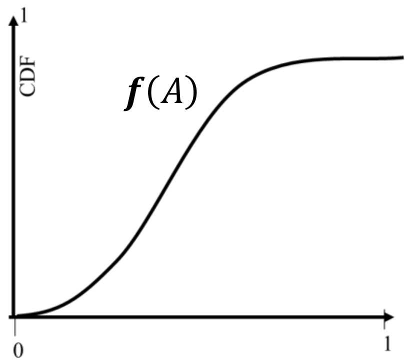
Enhanced Bayesian Networks

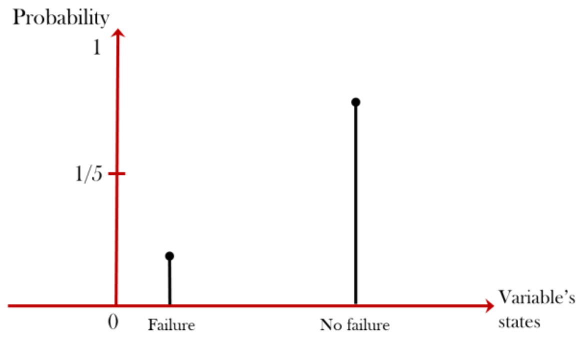
Bayesian Networks enhanced* with Structural Reliability Methods (SRM) permit to calculate the conditional probability values of discrete children that come from continuous-parent nodes.

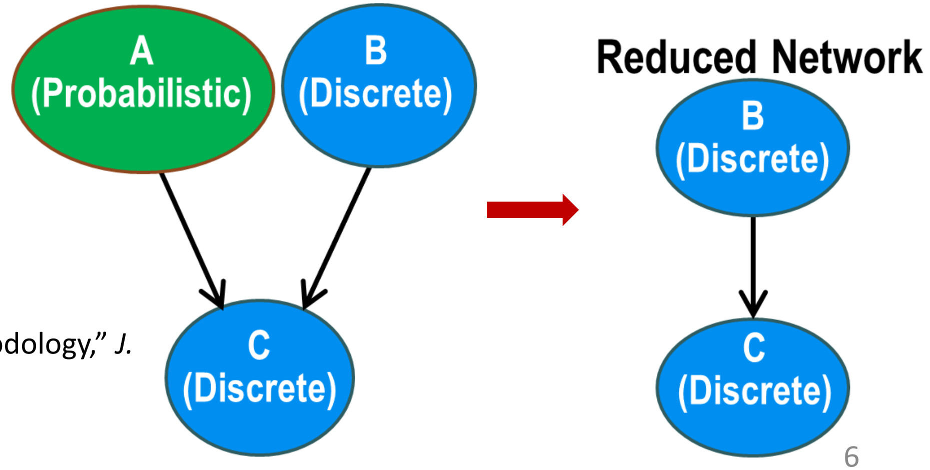
- Calculation of conditional probabilities consist in the approximation of the failure probability.

$$\boldsymbol{P}(C|B) = \int_{\Omega^{c}_{C,b}} \boldsymbol{f}(A)dA$$

f(A): Probability Density Function of continuous node A. $\Omega^{c}_{C,b}$ is the domain when C=c in the space of C given B=b.







[*] D. Straub and A. Der Kiureghian, "Bayesian Network Enhanced with Structural Reliability Methods: Methodology," *J. Eng. Mech.*, vol. 136, no. 10, pp. 1248--1258, Oct. 2010.

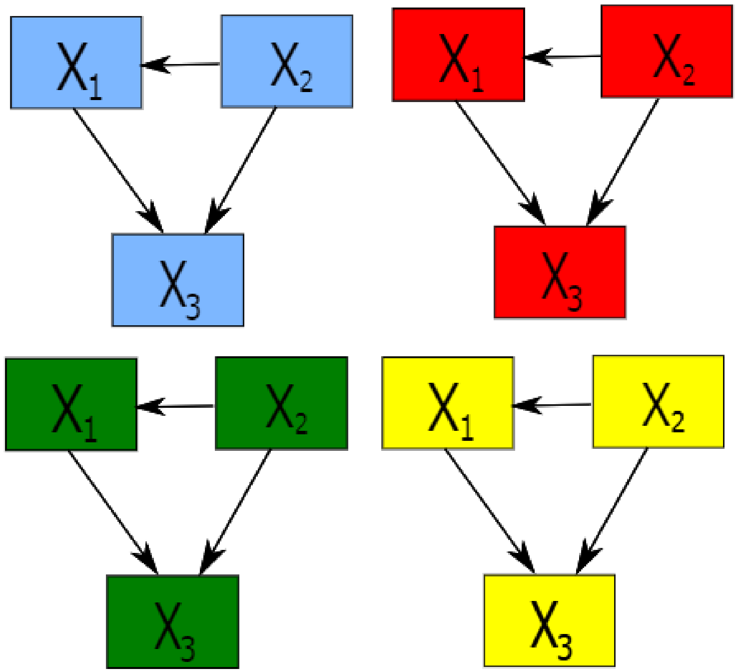
Imprecise data sets (discrete): Credal Networks

Generalization of BN to implement imprecise discrete variables in the form of intervals.

- Imprecision is represented through the so called credal sets $K(x_i)$.

$$K(x_i) := CH\left\{P(x_i) \middle| P(x_i) = \prod_{i=1}^{n} P(x_i|\pi_i)\right\}$$

- CNs inherent all the probabilistic and graphical characteristics of BNs.
- A CN is a **set of BNs**, each with different probability values.



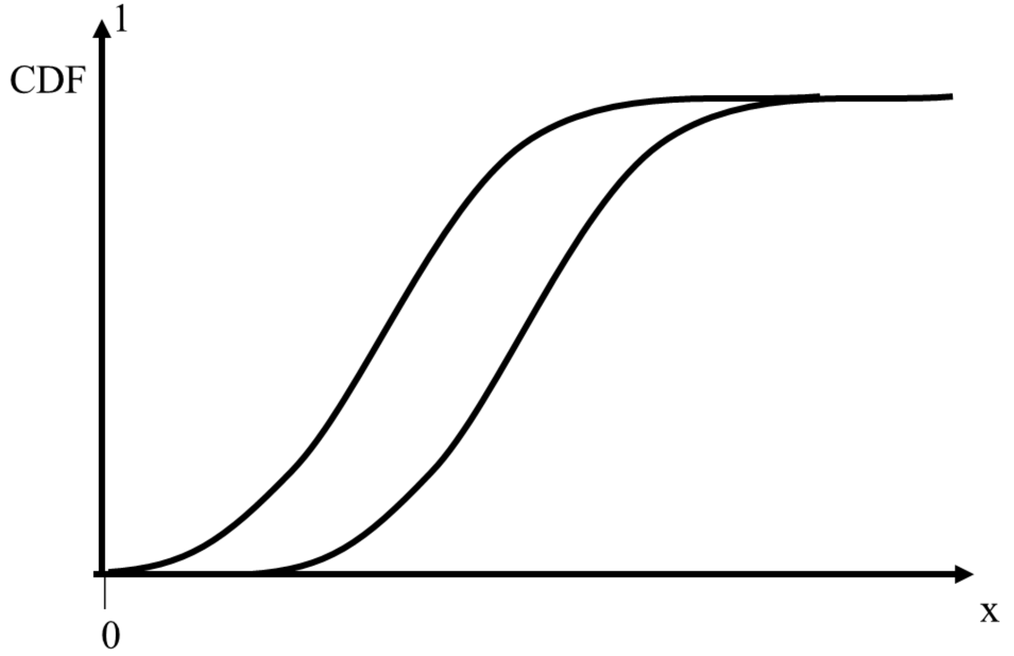
Different extreme points combinations make a set of BNs that makes up a CN.

Imprecise datasets (continuous): Probability boxes

A characterization of an uncertain continuous measure in the cumulative distribution space.

- When using SRM failure probability is now represented as:

$$\overline{P_f} = \max_{\theta} \int_{g(x)<0} p(x,\theta)dx$$



- In this way, the continuous probability distributions affected by **aleatoric** and **epistemic uncertainty** are taken into account.

Computational toolbox

OpenCossan

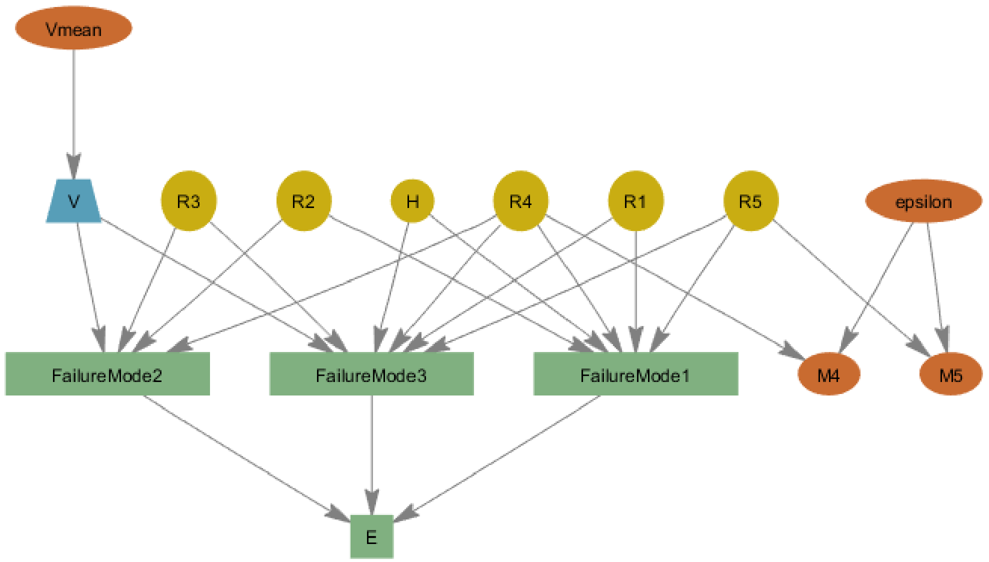
- It takes advantage of Object-Oriented programming in Matlab.
- Parallelization of high demanding tasks.
- Easy connectable with 3rd party toolboxes.
- Excellent platform for EBN.

UNIVERSITY OF LIVERPOOL | Institute for Risk and Uncertainty

www.cossan.co.uk

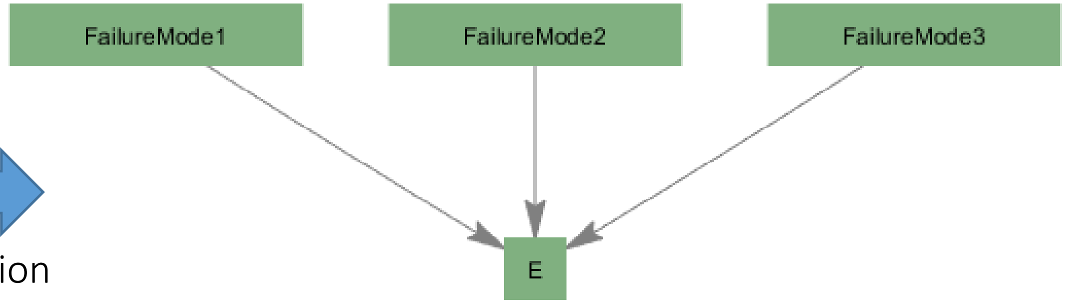
Enhanced BN to Credal nets

Enhanced Bayesian network[*]



Reduction process

Credal network[*]



Enhanced Bayesian network [*] (Advanced BN)

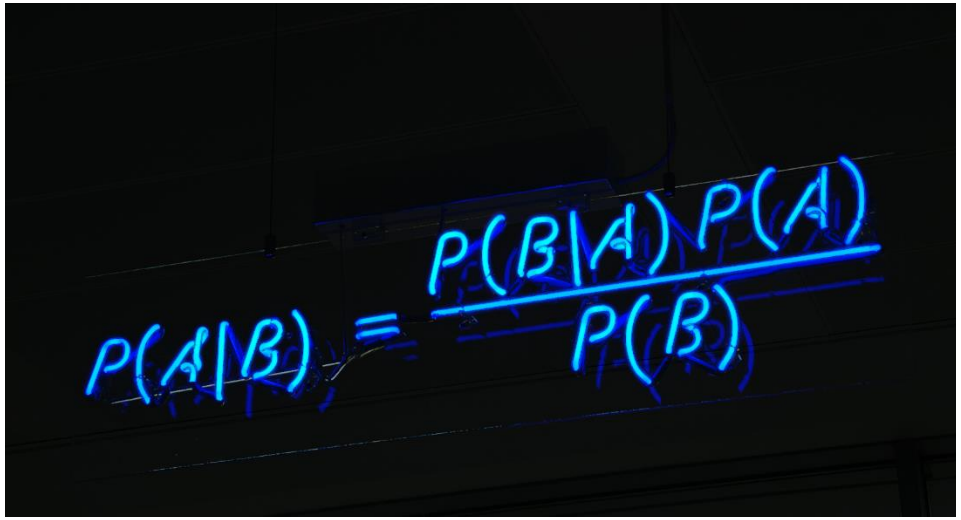
- Rectangle-Discrete
- Ellipse-Interval
- Circle-Continuous
- Trapezoid- P-box

- Rectangle-Interval

[*] Silvia Tolo, Tutorial Enhanced Bayesian networks. OpenCossan Tutorial.

Bayesian updating (Inference)

Computation of posterior distribution, P(A|B), of a query node (A) given (or not) evidence (B).



Bayes' Theorem

Bayesian updating (example)

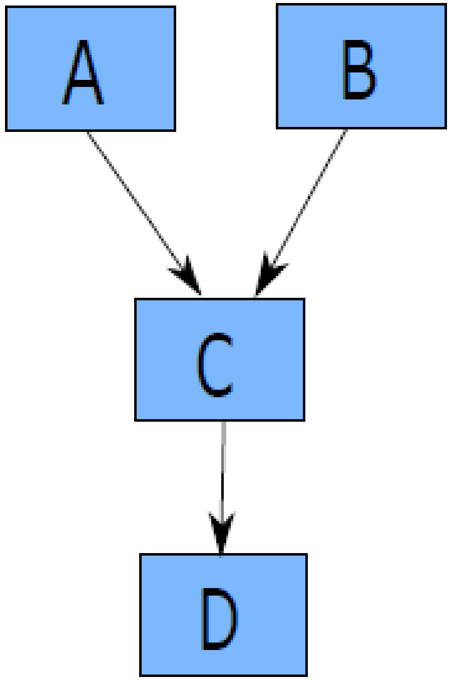
Computation of posterior distribution, P(A|B), of a query node (A) given (or not) evidence (B).

JPD of the network N with binary variables :

$$\mathrm{P(N) = P(A, B, C, D) = P(A)P(B)P(C|A, B)P(D|C)}$$

What if we can to compute $P(C_1|D_1)$?

$$\mathrm{P}(C_1|D_1) = \frac{\sum_{A,B} P(N)}{\sum_{A,B,C} P(D_1)}$$

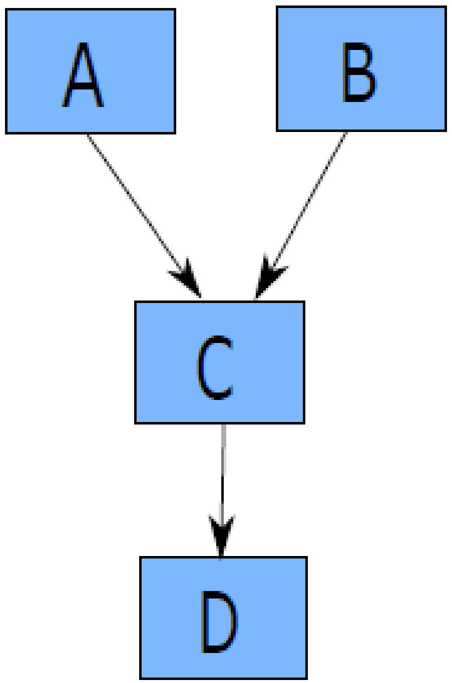


Traditional BN

Bayesian updating (example)

Where:

$$\sum_{A,B} \mathrm{P(N)} =$$
$$\mathrm{P}(\mathrm{D}{=}D_1 \mid C = C_1) \sum_A \mathrm{P(A)} \sum_B \mathrm{P(B)}\, \mathrm{P}(C = C_1 \mid \mathrm{A,B})$$
$$= \mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)\mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)$$
$$+ \mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)\mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)$$
$$+ \mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)\mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)$$
$$+ \mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)\mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)$$

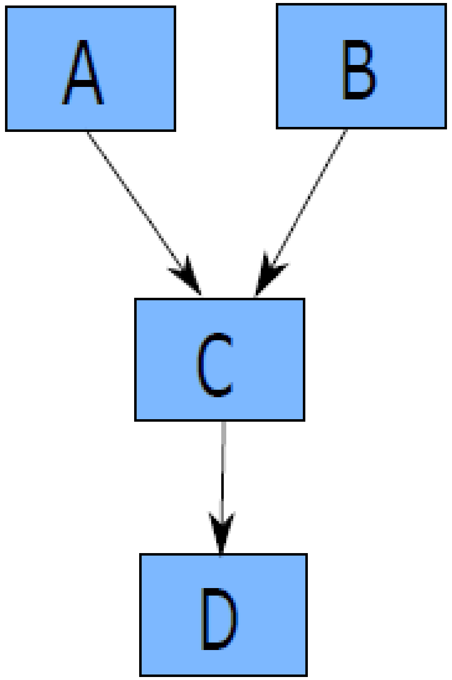


Traditional BN

Bayesian updating (example)

Where:

$$\sum_{A,B,C} \mathrm{P}(\mathrm{D}{=}D_1) =$$
$$\sum_{A} \mathrm{P}(\mathrm{A}) \sum_{B} \mathrm{P}(\mathrm{B})\, \mathrm{P}(\mathrm{C} \mid \mathrm{A},\mathrm{B}) \sum_{C} \mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C})$$
$$= \mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_2 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_2)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_1)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_2 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_2)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{C}{=}C_2 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_1)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_2)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_1 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_1)$$
$$+ \mathrm{P}(\mathrm{A}{=}A_2)\mathrm{P}(\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{C}{=}C_2 \mid \mathrm{A}{=}A_1,\mathrm{B}{=}B_2)\mathrm{P}(\mathrm{D}{=}D_1 \mid \mathrm{C}{=}C_2)$$

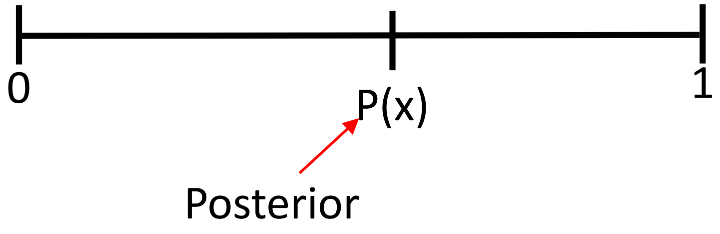


Traditional BN

Exact inference

Exact inference methods:

- Variable elimination (Marginalization).
- Junction tree algorithm (Clique tree).
- Recursive conditioning.
- And/Or search.

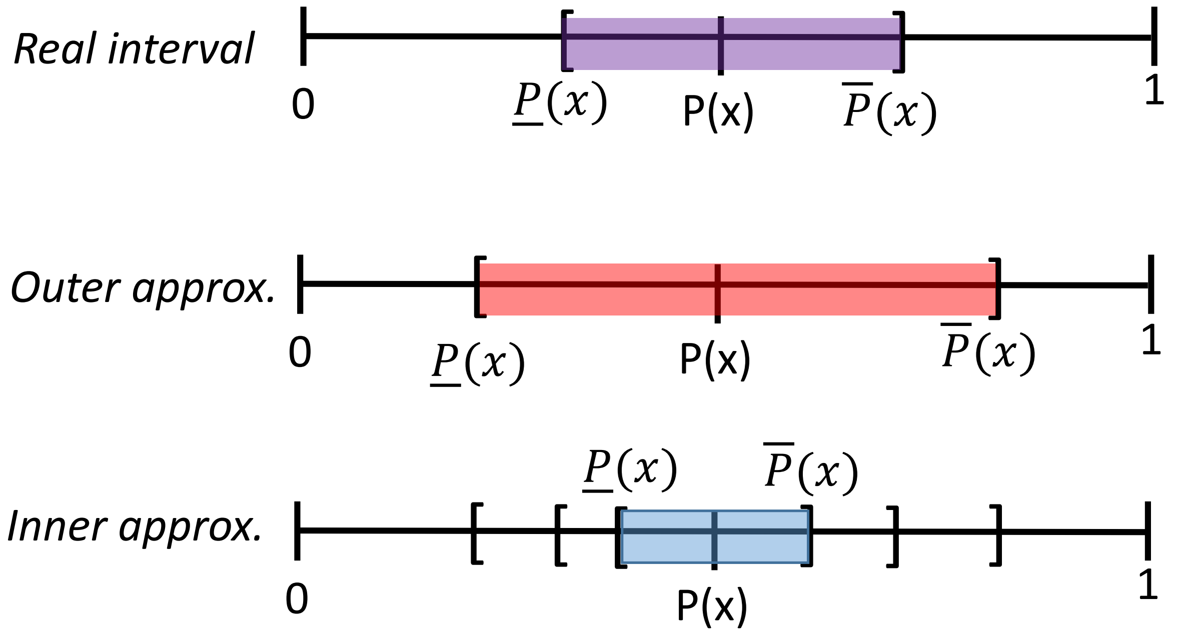


This method is applicable to traditional and relatively small BNs.

Inference with intervals

Approximate inference.

- Inner and outer approximation.
- Linear programming approximation.
- Importance sampling.
- Stochastic MCMC simulation.
- Mini-bucket elimination.
- Generalized belief propagation.
- Variational methods.



Inference with intervals

It is based on the joint credal set definition to calculate the bounds of the marginal probability as:

$$\underline{P}(x_0) := \min_{P(\mathbf{X}) \in K(\mathbf{X})} P(x_0) = \min_{\substack{P(X_i|\pi_i) \in K(X_i|\pi_i) \\ \pi_i \in \Omega_{\Pi_i}, i=0,\ldots,n}} \sum_{x_1, x_2, \ldots, x_n} \prod_{i=0}^{n} P(x_i|\pi_i)$$

$$\overline{P}(x_0) := \max_{P(\mathbf{X}) \in K(\mathbf{X})} P(x_0) = \max_{\substack{P(X_i|\pi_i) \in K(X_i|\pi_i) \\ \pi_i \in \Omega_{\Pi_i}, i=0,\ldots,n}} \sum_{x_1, x_2, \ldots, x_n} \prod_{i=0}^{n} P(x_i|\pi_i)$$

This represents a non-linear optimization problem with a multilinear objective function. (The head ache of CN inference).

Method 1: Naïve approach (Outer approximation)*

- Take the joint probability distribution function of upper bounds of all the variables in the net. Artificial JPDs are created (each containing a case of the query node).

$$P(\overline{\mathbf{F}}, \overline{\mathbf{S}}, \overline{\mathbf{A}}) = \begin{bmatrix} \begin{bmatrix} p(\overline{F_1}, \overline{S_1}, \overline{A_1}) & p(\overline{F_1}, \overline{S_2}, \overline{A_1}) \\ p(\overline{F_2}, \overline{S_1}, \overline{A_1}) & p(\overline{F_2}, \overline{S_2}, \overline{A_1}) \end{bmatrix} \\ \begin{bmatrix} p(\overline{F_1}, \overline{S_1}, \overline{A_2}) & p(\overline{F_1}, \overline{S_2}, \overline{A_2}) \\ p(\overline{F_2}, \overline{S_1}, \overline{A_2}) & p(\overline{F_2}, \overline{S_2}, \overline{A_2}) \end{bmatrix} \end{bmatrix}$$

Artificial Joint Probability Distribution

- Outer approximation obtained by computing exact inference in 2 artificial JPDs. 1 containing the all-lower and another the all-upper bounds.

[*]S. Tolo, E. Patelli, and M. Beer, "An Inference Method for Bayesian Networks with Probability Intervals," *ICVRAM ISUMA UNCERTAINTIES conference proceedings*, no. April, 2018.

Method 1: Naïve approach (inner approximation)

- Take the joint probability distribution function of upper bounds of all the variables in the net. Artificial JPDs are created (each containing a case of the query node).

$$P(\overline{\mathbf{F}}, \overline{\mathbf{S}}, \overline{\mathbf{A}}) = \begin{bmatrix} \begin{bmatrix} p(\overline{F_1}, \overline{S_1}, \overline{A_1}) & p(\overline{F_1}, \overline{S_2}, \overline{A_1}) \\ p(\overline{F_2}, \overline{S_1}, \overline{A_1}) & p(\overline{F_2}, \overline{S_2}, \overline{A_1}) \end{bmatrix} \\ \begin{bmatrix} p(\overline{F_1}, \overline{S_1}, \overline{A_2}) & p(\overline{F_1}, \overline{S_2}, \overline{A_2}) \\ p(\overline{F_2}, \overline{S_1}, \overline{A_2}) & p(\overline{F_2}, \overline{S_2}, \overline{A_2}) \end{bmatrix} \end{bmatrix}$$

Artificial Joint Probability Distribution

- Inner approximation is obtained by finding the artificial JPD that maximizes and minimizes the posterior probability of queried variable.

$$max \left[\frac{P(\overline{F_1}, \overline{\mathbf{S}}, \overline{\mathbf{A}})}{P(\overline{F_2}, \overline{\mathbf{S}}, \overline{\mathbf{A}})} \right] = max \begin{bmatrix} \frac{p(\overline{F_1}, \overline{S_1}, \overline{A_1})}{p(\overline{F_2}, \overline{S_1}, \overline{A_1})} & \frac{p(\overline{F_1}, \overline{S_2}, \overline{A_1})}{p(\overline{F_2}, \overline{S_2}, \overline{A_1})} \\ \frac{p(\overline{F_1}, \overline{S_1}, \overline{A_2})}{p(\overline{F_2}, \overline{S_1}, \overline{A_2})} & \frac{p(\overline{F_1}, \overline{S_2}, \overline{A_2})}{p(\overline{F_2}, \overline{S_2}, \overline{A_2})} \end{bmatrix}$$

$$min \left[\frac{P(\overline{F_1}, \overline{\mathbf{S}}, \overline{\mathbf{A}})}{P(\overline{F_2}, \overline{\mathbf{S}}, \overline{\mathbf{A}})} \right] = min \begin{bmatrix} \frac{p(\overline{F_1}, \overline{S_1}, \overline{A_1})}{p(\overline{F_2}, \overline{S_1}, \overline{A_1})} & \frac{p(\overline{F_1}, \overline{S_2}, \overline{A_1})}{p(\overline{F_2}, \overline{S_2}, \overline{A_1})} \\ \frac{p(\overline{F_1}, \overline{S_1}, \overline{A_2})}{p(\overline{F_2}, \overline{S_1}, \overline{A_2})} & \frac{p(\overline{F_1}, \overline{S_2}, \overline{A_2})}{p(\overline{F_2}, \overline{S_2}, \overline{A_2})} \end{bmatrix}$$

Method 2: Approximate inference

- Approximate inference with Linear programming. Optimization task.
- Reduce credal sets to singletons called Extreme Points $\tilde{P}(X_i|\pi_i) \in \mathrm{ext}[K(X_i|\pi_i)]$

different from the Free variable X_j.

So the constrained queried-variable (x_0) lower bound is:

$$\underline{P}'(x_0) := \min_{\substack{P(X_j|\pi_j)\in K(X_j|\pi_j) \\ \pi_j \in \Omega_{\Pi_j}}} \sum_{x_j,\pi_j} \underbrace{\left[\tilde{P}(x_0|x_j,\pi_j)\cdot\tilde{P}(\pi_j)\right]\cdot P(x_j|\pi_j)}$$

Linear combination of X_j local probabilities.

A. Antonucci, C. P. De Campos, D. Huber, and M. Zaffalon, "Approximate credal network updating by linear programming with applications to decision making," *Int. J. Approx. Reason.*, vol. 58, pp. 25–38, 2015.

Method 2: Approximate inference

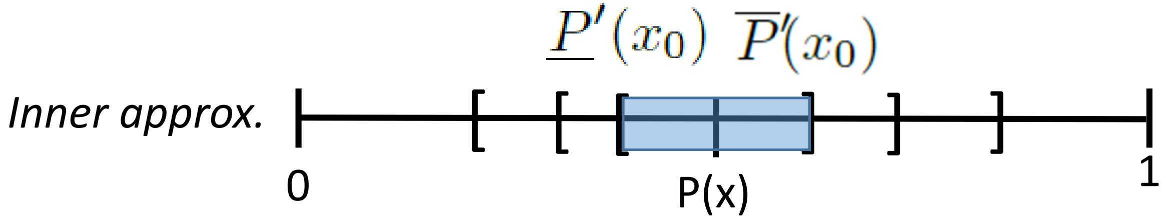
$$\underline{P}'(x_0) := \min_{\substack{P(X_j|\pi_j) \in K(X_j|\pi_j) \\ \pi_j \in \Omega_{\Pi_j}}} \sum_{x_j, \pi_j} \left[\tilde{P}(x_0|x_j, \pi_j) \cdot \tilde{P}(\pi_j) \right] \cdot P(x_j|\pi_j)$$

- Iterations over Xj are done to perform a local search.
- Once an approximation (extreme point) to the optimal solution is calculated. The Xj variable released and a new Xj is used as the free variable.
- The programme stops iterating when no further improved approximation is found.

A. Antonucci, C. P. De Campos, D. Huber, and M. Zaffalon, "Approximate credal network updating by linear programming with applications to decision making," *Int. J. Approx. Reason.*, vol. 58, pp. 25–38, 2015.

Method 2: Approximate inference

- $\underline{P}'(x_0)$ upper approximation of lower probability bound $\underline{P}'(x_0)$ of the CN.
- $\overline{P}'(x_0)$ is lower approximation of the upper bound $\overline{P}'(x_0)$ of the CN.

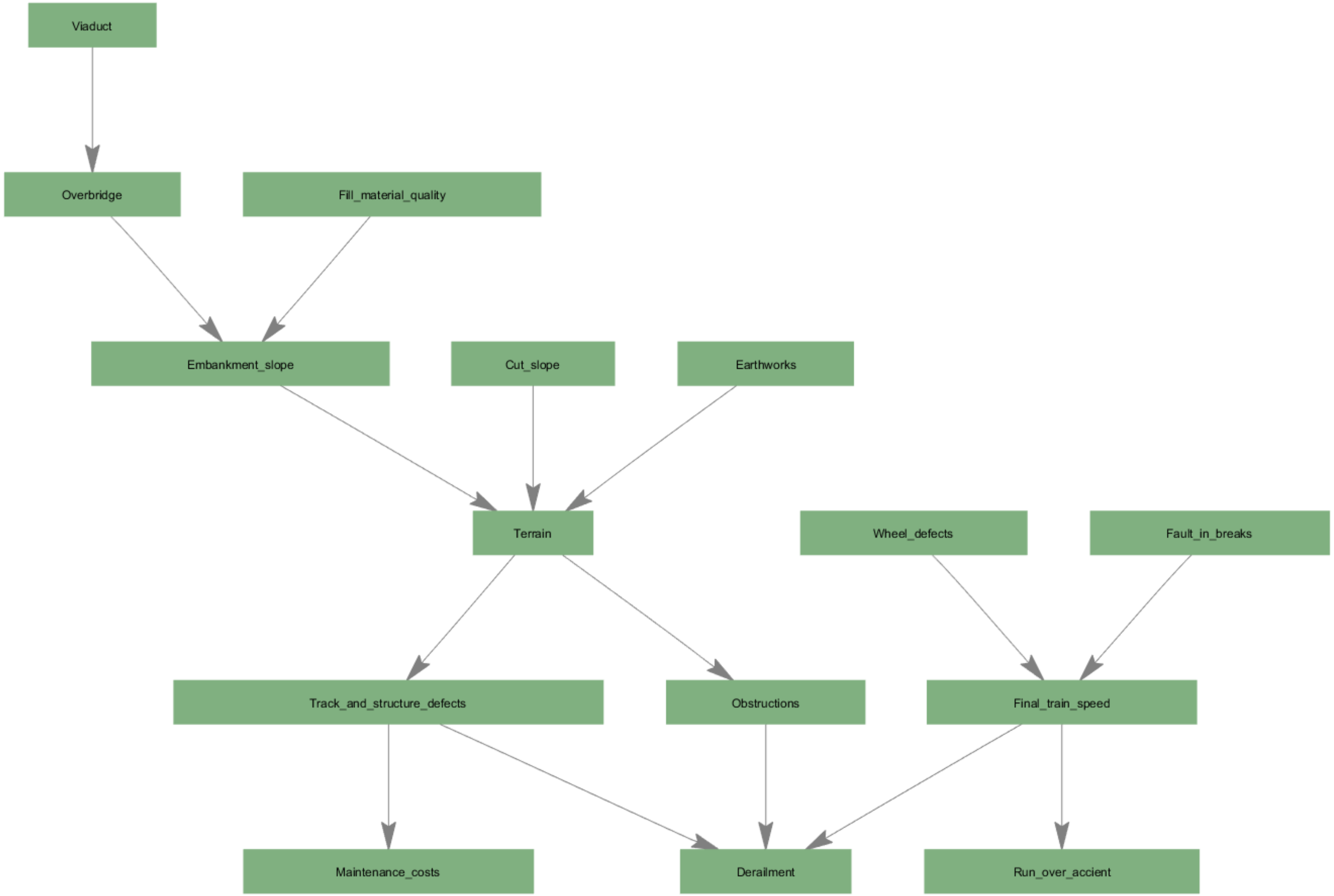


A. Antonucci, C. P. De Campos, D. Huber, and M. Zaffalon, "Approximate credal network updating by linear programming with applications to decision making," *Int. J. Approx. Reason.*, vol. 58, pp. 25–38, 2015.

Case of study: Railway system

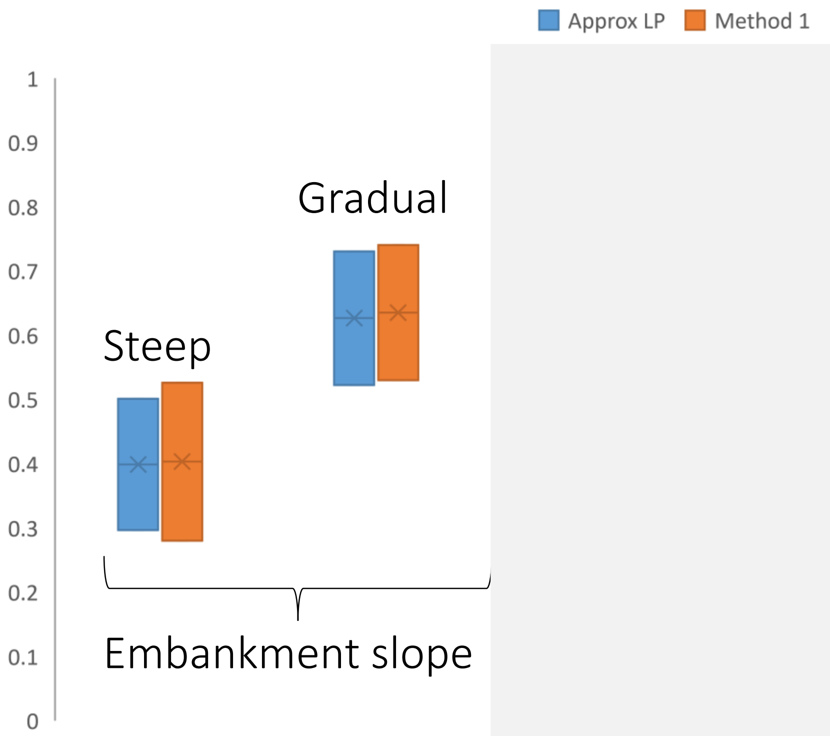
Derailment probability, taking into account:

- Obstructions in the railway due to:
    - Earthworks
    - Terrain
- Train speed.
- Damage in the tracks.



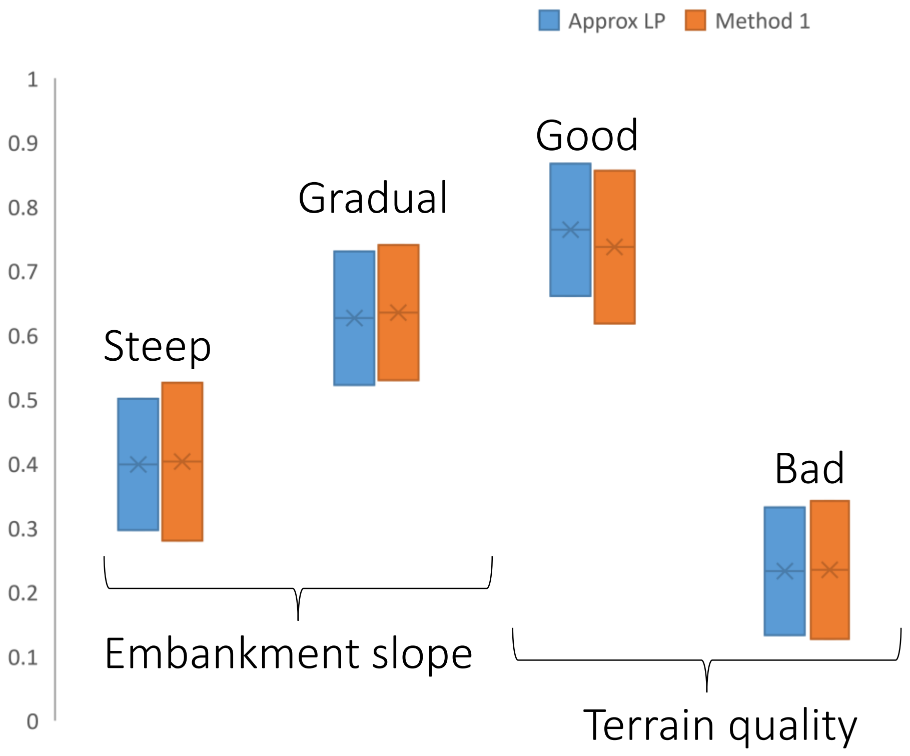
Results

- Embankment slope over which the rail tracks are placed.
- Terrain quality depending on:
  - Earthworks
  - Cut slopes
  - Embankment slope steepness
- Derailment, due to factors:
  - Final train speed
  - Track obstructions
  - Track defects



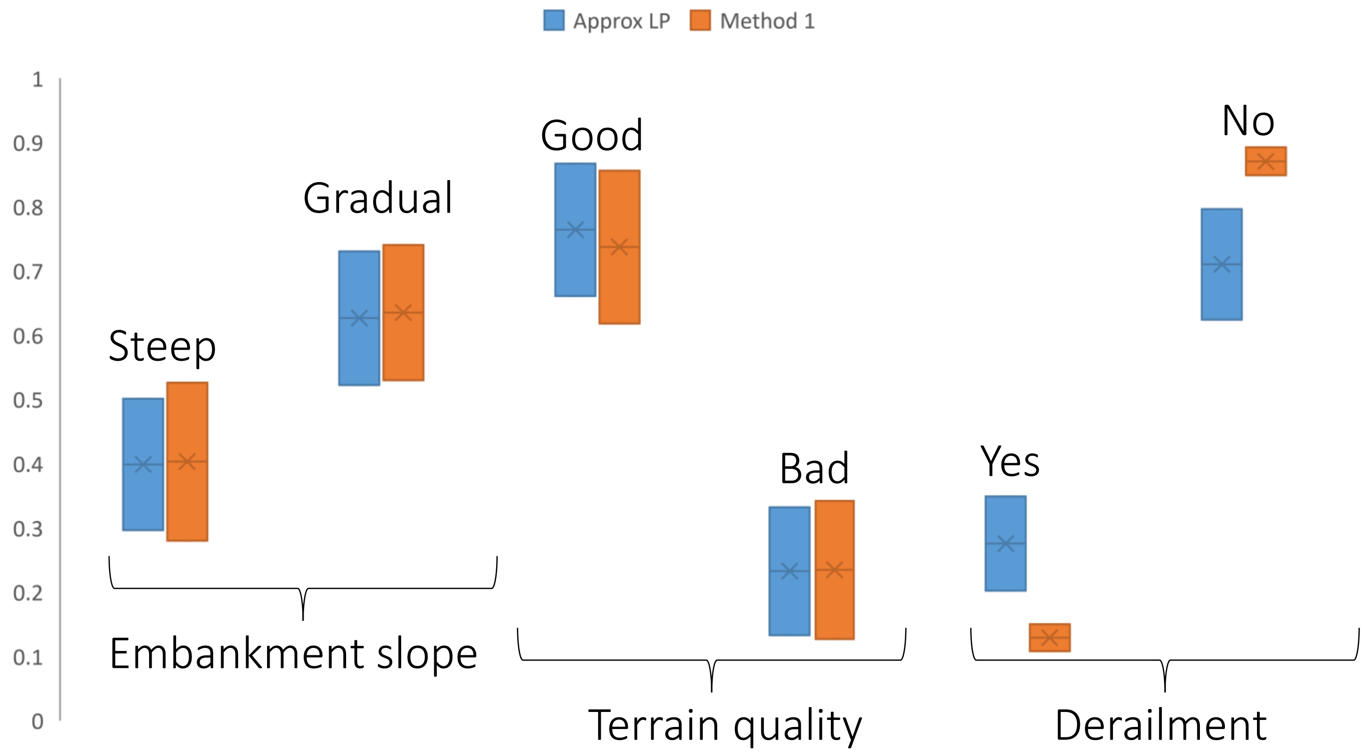
Results

- Embankment slope over which the rail tracks are placed.
- Terrain quality depending on:
  - Earthworks
  - Cut slopes
  - Embankment slope steepness
- Derailment, due to factors:
  - Final train speed
  - Track obstructions
  - Track defects



Results

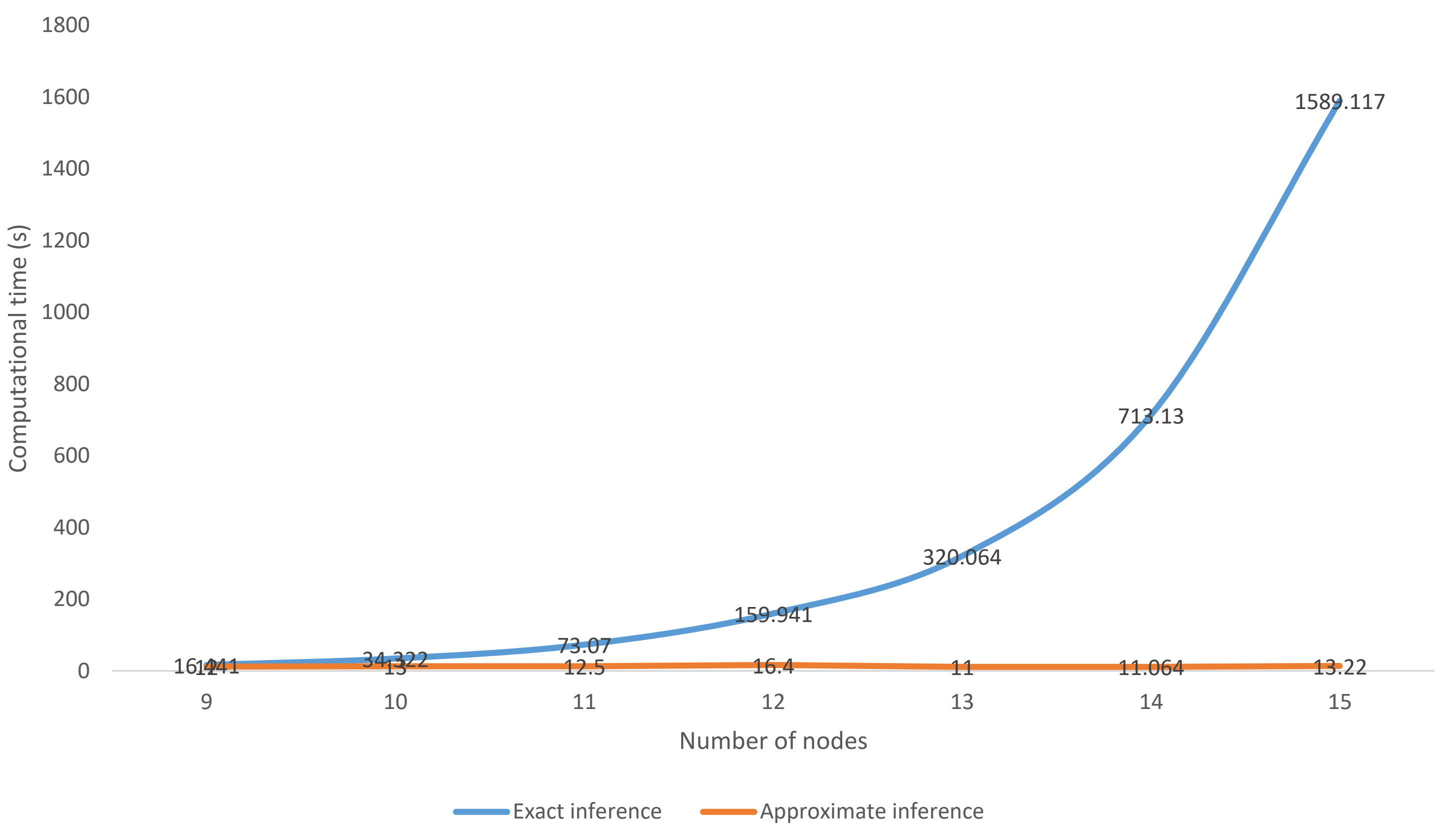
- Embankment slope over which the rail tracks are placed.
- Terrain quality depending on:
  - Earthworks
  - Cut slopes
  - Embankment slope steepness
- Derailment, due to factors:
  - Final train speed
  - Track obstructions
  - Track defects



Computational time

Results

- Obstructions in the railway due to:
  - Earthworks
  - Terrain
- Train speed.
- Damage in the tracks.



Method 1: Naïve inference

- ✓This method is computationally cheap.
- ✓ Reliable when **extreme scenarios** are of the interest.
- ✓ Real probability values will be enclosed inside the bounds.
- ✓ Uncertainty attached to the bounds provided.
- ✓ No need for inference computation on node-state combination irrelevant.

- o Boolean variables.
- o Overestimation of upper bounds.
- o Underestimation of lower bounds.
- o Not suitable for large networks, number of inference computations increase as 2^n.

Method 2: Approximate inference

- ✓ Does not suffer from **large credal sets**.
- ✓ Follows the same topology of BN.
- ✓ Does not requires to indicate the extreme points.
- ✓ It can be used with variables with **many states and/or parents**.
- ✓ Provides inner approximate solutions.
- ✓ **Fast and accurate**.

- Local credal sets specified by lean constraints.*
- Not for local credal sets given by explicit enumeration of the extreme points.
- Outer approximations are currently excluded.
- A combination of inner with outer approximations can bring reliable inferences.

Conclusions

- Two different inference computation methods were tested to compare their performance.
- The use of interval probabilities allows to consider a broader range data types (imprecise data sets).
- Imprecise probabilities allows to take into account epistemic uncertainty due to the vagueness or lack of data.
- This model can be applicable to different complex technological facilities.
- Work is carried out to provide a reliable method to provide an outer approximation of the probability bounds and study convergence.

CONACYT
Consejo Nacional de Ciencia y Tecnología

Thank you for your attention

WPMSIIP 2018,
Oviedo, Spain

Questions?

Approximate inference methods for Advanced Bayesian networks

Hector Diego Estrada Lugo

h.d.estrada-lugo@liverpool.ac.uk

UNIVERSITY OF LIVERPOOL
Institute for Risk and Uncertainty

https://www.liverpool.ac.uk/risk-and-uncertainty/