Bayesian probabilistic networks

Definition and use of Bayesian probabilistic networks

K. Nishijima
nishijima@ibk.baug.ethz.ch

Today’s agenda

Probabilistic representation of systems

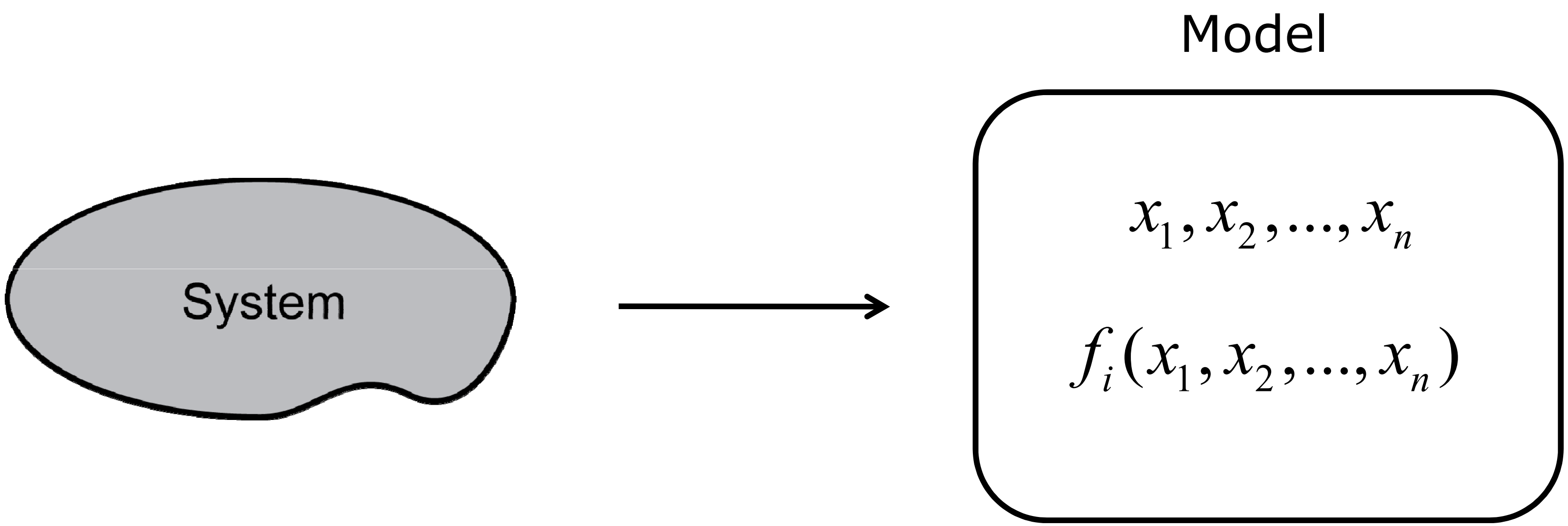
- Representation of systems with (random) variables
- Observable and unobservable variables
- joint probability, conditional probability, marginal probability, prior/posterior probability, predictive probability

Bayesian probabilistic networks (BPNs)

- Graphical representation / conditional probability (tables)
- Advantages of BPNs
- Example

Representation of systems

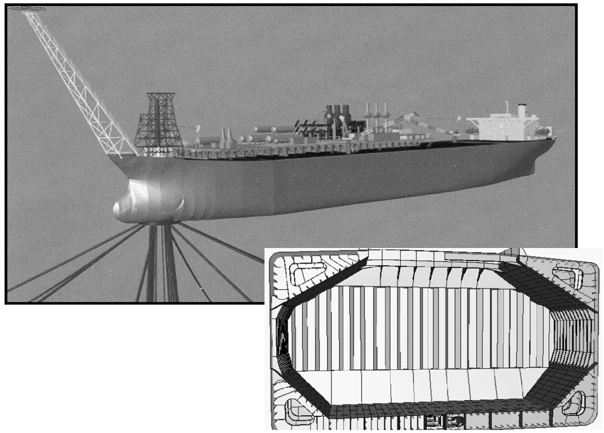
- Systems may be characterized by variables and the systems performance may be described by using the variables.



Representation of systems

- Systems may be characterized by variables and the systems performance may be described by using the variables.

Ship hull structure system



Corresponding model

Variables:
- Self weight
- Wave loads
- Thickness of plates
- Stiffness
- Corrosion, etc.

Representation of systems

- If the states of the systems are uncertain, the systems may be characterized by random variables.

Model

$$x_1, x_2, \ldots, x_n$$
$$f_i(x_1, x_2, \ldots, x_n)$$

→

Probabilistic model

$$X_1, X_2, \ldots, X_n$$
$$f_i(X_1, X_2, \ldots, X_n)$$

Task in modeling

- The main task in modeling is to identify the interrelations between all the variables involved in the models in logical and/or probabilistic manners.

Probabilistic model

$$X_1, X_2, \ldots, X_n$$

$$f_i(X_1, X_2, \ldots, X_n)$$

- Example of logical relation:

$$X_3 = X_1 + X_2$$

- Example of probabilistic relation:

$$P[X_3 \mid X_1, X_2]$$

(The logical relation is a special case of the probabilistic relation.)

Aim of analysis

- The aim of the analysis with models is to evaluate the quantities of interest, e.g. the probability of the occurrence of critical situations, using the logical and probabilistic relations in the models.

Probabilistic model

$$X_1, X_2, ..., X_n$$

$$f_i(X_1, X_2, ..., X_n)$$

- Example of quantities of interest:

$$P_F = P[X = 1]$$

where

$$X = I[X_1 < X_2]$$

Observable and unobservable

Observable variables

- The values of some of variables can be observed by e.g. measurements.

- An example is the annual maximum wind speed at a given location:

  10 m/s, 12 m/s, 29 m/s, 14 m/s, 35 m/s, 18 m/s.

Observable and unobservable

Unobservable variables

(latent variables, model parameters, etc.)

- Unobservable variables are not directly observed but rather inferred from the variables that are directly measured.

- An example is the mean value of the annual maximum wind speed at one location. The mean value may be inferred from the observations as:

$$(10+12+29+14+35+18)/6 = 19.67 \text{ m/s}$$

Observable and unobservable

Example

- The annual maximum wind speeds X_i ($i=1,2,3,4,5$) are assumed modeled as:

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$$

  Here, X_i 's are the examples of observable variables, and μ and σ^2 are the examples of unobservable variables.

- In Bayesian approaches, μ and σ^2 themselves may be considered as **random variables**.

Joint probability

- Any models that represent systems can be fully described by joint probabilities (probability densities or their combination):

$$P[X_1, X_2, ..., X_n]$$

- Only few parametric families of multivariate distributions are known, e.g., multivariate Normal distribution, Dirichlet distribution. They are often **less flexible**; consider the case X_1 is continuous and X_2 is discrete.

- Quite **lots of information are required** for characterizing the multivariate distributions; $(\mathrm{n}^2 + \mathrm{n})$ values are required for the multivariate Normal distribution. Also, think of a multivariate discrete distribution!

Conditional probability

- An efficient (compact) way of characterizing joint probabilities is to employ conditional probabilities.

- Any joint probability can be built up with conditional probabilities:

$$P[X_1, X_2, ..., X_n] = \prod_{i=1}^{n} P[X_i \mid pa(X_i)]$$

where $pa(X_i)$ represents the parents of X_i (which will be later explained in the context of Bayesian probabilistic networks.)

Conditional probability

Example

- Assume two situations where the joint probability of discrete random variables X_1, X_2 and X_3, each of which can take 5 possible values, are modeled in two different ways:
  - $P[X_1, X_2, X_3]$

    This *direct* representation requires 5^3=125 values to describe the joint probability.
  - $P[X_1, X_2, X_3] = P[X_3 \mid X_2] P[X_2 \mid X_1] P[X_1]$

    This *conditional independence* representation requires only 5+5^2+5^2=55 values, although this representation is less flexible than the above one.

Marginal probability

- Any marginal probability can be calculated from joint probabilities for discrete cases as:

$$P[X_i] = \sum_{1} \sum_{2} \cdots \sum_{i-1} \sum_{i+1} \cdots \sum_{n} P[X_1, X_2 .., X_n]$$

For continuous cases the summations may be replaced by appropriate integrals.

Marginal probability

Example

Consider a Bernoulli trial – throwing a coin 10 times. However, the probability that a head comes out is not known. Assume that X_1 represents the number that the head comes out in the 10 trials, and X_2 represents the probability that the head comes out in each trial and follows the uniform distribution $[0,1]$.
Calculate $P[X_1]$.

Marginal probability

Example (solution)

- Joint probability

$$P[X_1 = i, X_2 = p] = P[X_1 = i \mid X_2 = p] \cdot P[X_2 = p]$$

$$= \binom{10}{i} p^i (1-p)^{10-i} \cdot 1$$

- Marginal probability

$$P[X_1 = i] = \int_0^1 P[X_1 = i, X_2 = p] dp = \int_0^1 \binom{10}{i} p^i (1-p)^{10-i} dp$$

$$= \frac{10!}{(10-i)!i!} \frac{i!(10-i)}{11!} = \frac{1}{11}$$

Prior/posterior probability

- According to the Bayes' theorem the following relation holds:

$$P[X_2 \mid X_1 = x_1] = \frac{P[X_1 = x_1 \mid X_2]}{P[X_1 = x_1]} P[X_2]$$

Consider special cases where X_2 represents the parameter of the probability of X_1. In such cases, the relation above provides a rationale how the probability of X_2 may be **updated** with the observed data ($X_1 = x_1$). Thereby,

- $P[X_2]$: **prior** probability (distribution)
- $P[X_2 \mid X_1 = x_1]$: **posterior** probability (distribution).

Prior/posterior probability

Example

Consider again the Bernoulli trial – throwing a coin. Assume that 5 heads came out after 10 trials, i.e., $X_1 = 5$.
Calculate $P[X_2 | X_1 = 5]$, the posterior probability of X_2.

Prior/posterior probability

Example

- Posterior probability

$$P[X_2 = p \mid X_1 = 5] = \frac{P[X_1 = 5 \mid X_2 = p]}{P[X_1 = 5]} P[X_2 = p]$$

$$\propto p^5 (1-p)^5$$

Thus,

$$P[X_2 = p \mid X_1 = 5] = \frac{\Gamma(12)}{\Gamma(6)\Gamma(6)} p^5 (1-p)^5$$

Predictive probability

- Predictive probability is an alias of the marginal probability when the joint probabilities are marginalized in regards to the parameters of the probabilities:

$$P[X_1] = \sum_2 P[X_1, X_2] = \sum_2 P[X_1 \mid X_2] P[X_2]$$

where X_2 is the parameter of the probability of X_1.

  - $P[X_1]$: Predictive probability

Predictive probability

Example

Consider again the Bernoulli trial – throwing the coin. Calculate the predictive probability $P[X_1]$, using the updated (posterior) probability of X_2.

Predictive probability

Example (solution)

$$P[X_1 = i] = \int_0^1 P[X_1 = i, X_2 = p]dp$$

$$= \int_0^1 P[X_1 = i \mid X_2 = p]P[X_2 = p]dp$$

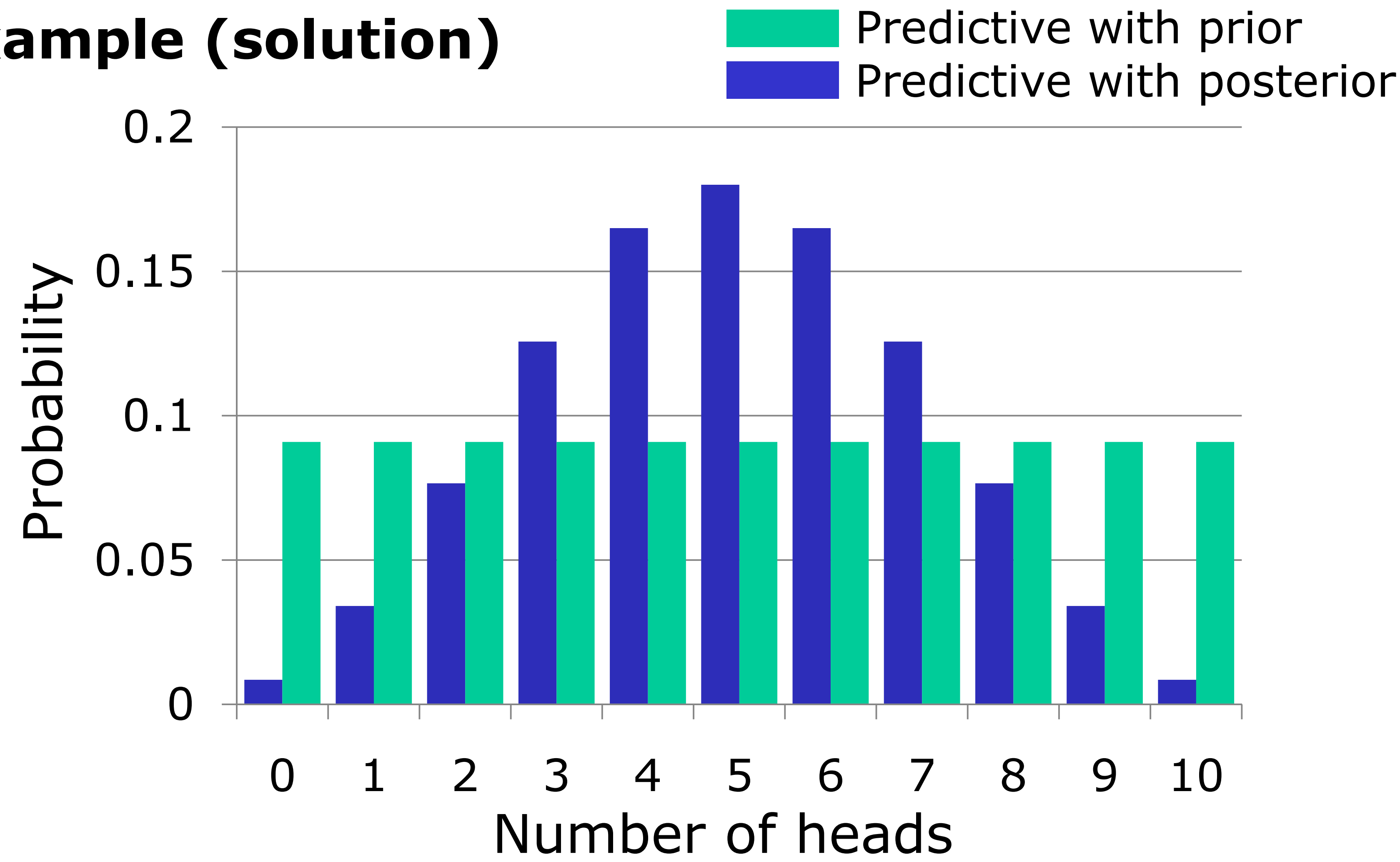
$$= \int_0^1 \binom{10}{i} p^i (1-p)^{10-i} \frac{\Gamma(12)}{\Gamma(6)\Gamma(6)} p^5 (1-p)^5 dp$$

$$= \binom{10}{i} \frac{\Gamma(12)}{\Gamma(6)\Gamma(6)} \int_0^1 p^{i+5} (1-p)^{15-i} dp$$

$$= \frac{\Gamma(11)}{\Gamma(i+1)\Gamma(11-i)} \frac{\Gamma(12)}{\Gamma(6)\Gamma(6)} \frac{\Gamma(i+6)\Gamma(16-i)}{\Gamma(22)}$$

Predictive probability

Example (solution)

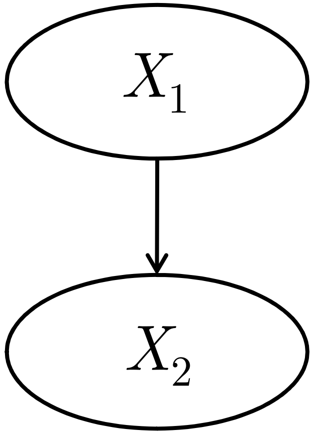


Bayesian probabilistic networks

- DAG (**D**irected **A**cyclic **G**raph)
- Conditional probability (tables)

Bayesian probabilistic networks

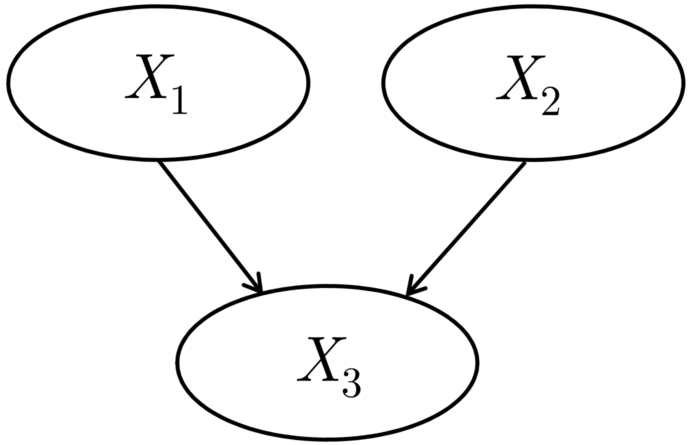
- Example of DAG (**D**irected **A**cyclic **G**raph)



$$P[X_2 \mid X_1]P[X_1]$$

Bayesian probabilistic networks

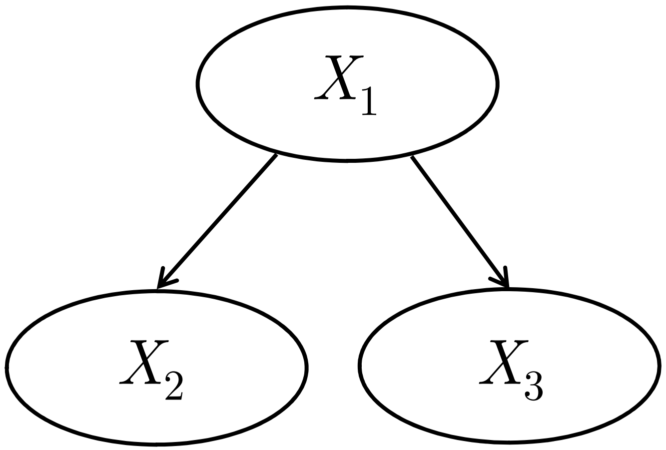
- Example of DAG (**D**irected **A**cyclic **G**raph)



$$P[X_3 \mid X_1, X_2]P[X_1]P[X_2]$$

Bayesian probabilistic networks

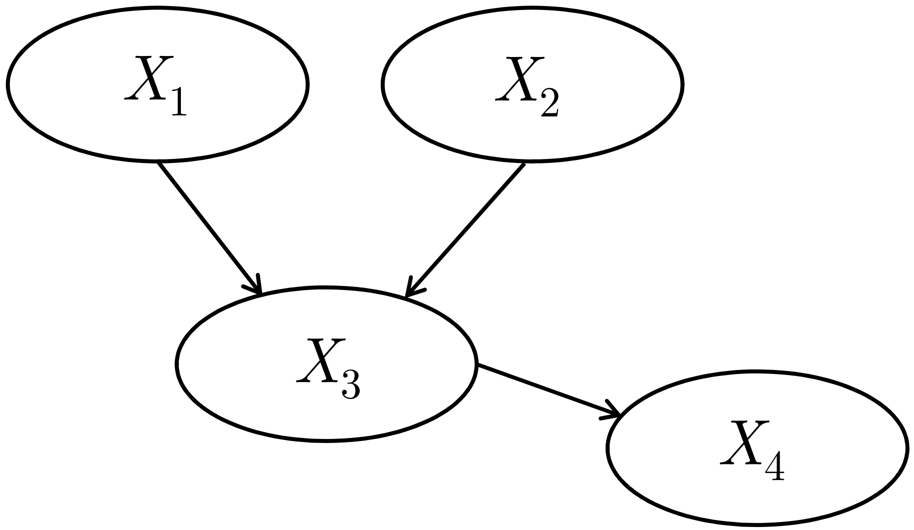
- Example of DAG (**D**irected **A**cyclic **G**raph)



$$P[X_2 \mid X_1]P[X_3 \mid X_1]P[X_1]$$

Bayesian probabilistic networks

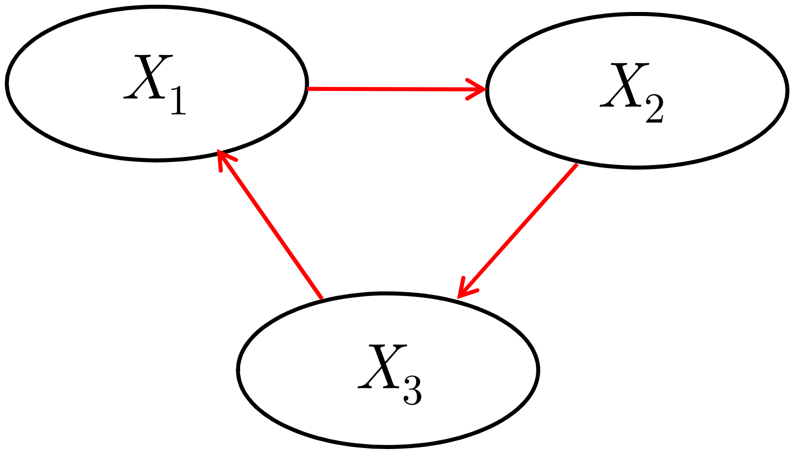
- Example of DAG (**D**irected **A**cyclic **G**raph)



$$P[X_4 \mid X_3]P[X_3 \mid X_1, X_2]P[X_1]P[X_2]$$

Bayesian probabilistic networks

- Counter-example of DAG (**D**irected **A**cyclic **G**raph)



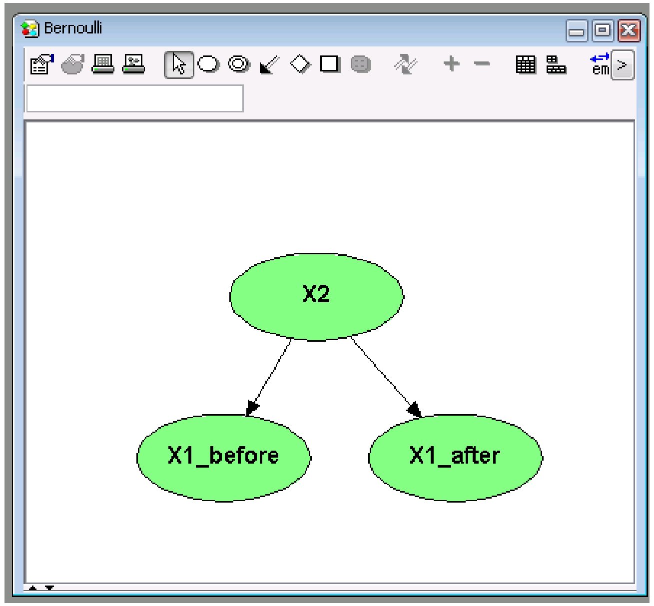
Why are BPNs useful?

- Easy way of mind mapping.
- Platform for integrating all relevant aspects.
- Generic algorithms as well as software tools are available for
  - creating BPNs
  - calculating probabilities of interest.

Bayesian probabilistic networks

Example

Consider the Bernoulli trial – throwing a coin 10 times.



- X1_before: number of heads at first trial
- X1_after: number of heads at second trial
- X2: the probability that a head comes out

← Will be demonstrated.

Modeling with BPNs

- Bottom-up approach

  The components of a BPN are built up by expert knowledge and analyses and are integrated into the BPN.

- Top-down approach

  The interrelation between the components in a BPN are fixed, but the probabilistic (quantitative) relations are estimated from observed data.

Let’s see these approaches with application examples later.