Introduction to Bayesian Networks

How We Can Use Them as Probabilistic and Causal Models

Marco Scutari
scutari@idsia.ch

Dalle Molle Institute for
Artificial Intelligence (IDSIA)

September 16, 2021

Bayesian Networks: a Graph and a Probability Distribution

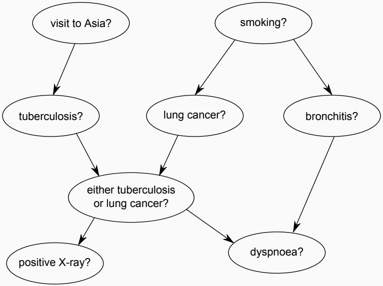
Bayesian networks (BNs) [6] are defined by:

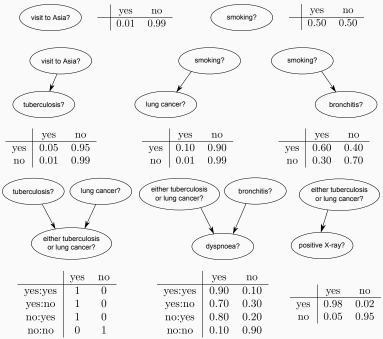
- a network structure, a directed acyclic graph $\mathcal{G}$ in which each node corresponds to a random variable X_i;
- a global probability distribution $\mathbf{X} = \{X_1, \dots, X_N\}$ which can be factorised into smaller local probability distributions according to the arcs present in the graph $\mathcal{G}$.

The main role of the network structure is to express the conditional independence relationships among the variables in the model through graphical separation, thus specifying the factorisation of the global distribution:

$$\mathrm{P}(\mathbf{X}) = \prod_{i=1}^{N} \mathrm{P}(X_i \mid \Pi_{X_i}; \Theta_{X_i}) \quad \text{where} \quad \Pi_{X_i} = \{\text{parents of } X_i\}.$$

Local distributions can be anything, so BNs can subsume many classical statistical models used in clinical practice.





visit to Asia?

yes	no
0.01	0.99

smoking?

yes	no
0.50	0.50

visit to Asia? → tuberculosis?

	yes	no
yes	0.05	0.95
no	0.01	0.99

smoking? → lung cancer?

	yes	no
yes	0.10	0.90
no	0.01	0.99

smoking? → bronchitis?

	yes	no
yes	0.60	0.40
no	0.30	0.70

tuberculosis?, lung cancer? → either tuberculosis or lung cancer?

	yes	no
yes:yes	1	0
yes:no	1	0
no:yes	1	0
no:no	0	1

either tuberculosis or lung cancer?, bronchitis? → dyspnoea?

	yes	no
yes:yes	0.90	0.10
yes:no	0.70	0.30
no:yes	0.80	0.20
no:no	0.10	0.90

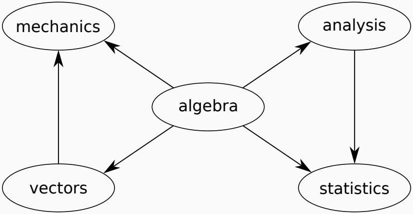
either tuberculosis or lung cancer? → positive X-ray?

	yes	no
yes	0.98	0.02
no	0.05	0.95

Discrete BNs subsume contingency table analysis and (multinomial) logistic regression: the local distributions $X_i \mid \Pi_{X_i}$ are conditional probability tables.

A classic example of discrete BN is the ASIA network from Lauritzen & Spiegelhalter (1988) [2], which includes a collection of binary variables. It describes a simple diagnostic problem for tuberculosis and lung cancer.

Gaussian Bayesian Networks

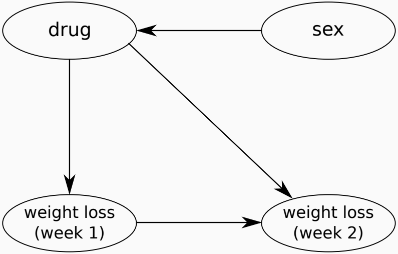


$$\text{ALG} = 50.60 + \varepsilon_{\text{ALG}}$$
$$\text{ANL} = -3.57 + 0.99\text{ALG} + \varepsilon_{\text{ANL}}$$
$$\text{MECH} = -12.36 + 0.54\text{ALG} + 0.46\text{VECT} + \varepsilon_{\text{MECH}}$$
$$\text{STAT} = -11.19 + 0.76\text{ALG} + 0.31\text{ANL} + \varepsilon_{\text{STAT}}$$
$$\text{VECT} = 12.41 + 0.75\text{ALG} + \varepsilon_{\text{VECT}}$$

Gaussian BNs subsume linear models and analysis of variance techniques. The local distributions $X_i \mid \Pi_{X_i}$ take the form of linear regression models with independent error terms in which the Π_{X_i} act as regressors.

A classic example of GBN is the MARKS networks from Mardia, Kent & Bibby (1979) [3], which describes the relationships between the marks on 5 math-related topics.

Conditional Linear Gaussian Bayesian Networks



Conditional Linear Gaussian BNs contain both discrete and continuous nodes, which are modelled using either conditional probability tables or mixtures of regression models.

$$\mathsf{W}_{2,\mathrm{D}_1} = 1.02 + 0.89\beta_{\mathrm{W}_1} + \varepsilon_{\mathrm{D}_1}$$
$$\mathsf{W}_{2,\mathrm{D}_2} = -1.68 + 1.35\beta_{\mathrm{W}_1} + \varepsilon_{\mathrm{D}_2}$$
$$\mathsf{W}_{2,\mathrm{D}_3} = -1.83 + 0.82\beta_{\mathrm{W}_1} + \varepsilon_{\mathrm{D}_3}$$

A classic example is the RAT WEIGHTS network from Edwards (1995) [1], which describes weight loss in a drug trial performed on rats.

More complex model setups are possible, but not often used, because they may be more difficult to interpret and for computational reasons.

Learning a Bayesian Network

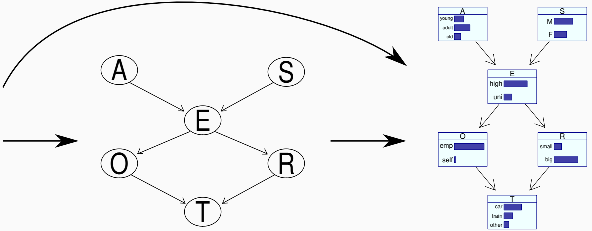
Model selection and estimation are collectively known as learning, and are usually performed as a two-step process:

1. Structure learning, learning the graph structure from the data.
2. Parameter learning, learning the local distributions implied by the graph structure learned in the previous step.

This workflow is implicitly Bayesian: given a data set $\mathcal{D}$ and if we denote the parameters of the global distribution as $\mathbf{X}$ with Θ, we have

$$\underbrace{\mathrm{P}(\mathcal{M} \mid \mathcal{D})}_{\text{learning}} = \underbrace{\mathrm{P}(\mathcal{G} \mid \mathcal{D})}_{\text{structure learning}} \cdot \underbrace{\mathrm{P}(\Theta \mid \mathcal{G}, \mathcal{D})}_{\text{parameter learning}} .$$

	A	R	E	O	S	T
1	adult	big	high	emp	F	car
2	adult	small	uni	emp	M	car
3	adult	big	uni	emp	F	train
4	adult	big	high	emp	M	car
5	adult	big	high	emp	M	car
6	adult	small	high	emp	F	train



Using Expert Knowledge to Construct a Bayesian Network

Learning a BN of any complexity is a data-driven exercise: usually there is not enough information available to build it from first principles. However, we can also incorporate information from expert knowledge along with that provided by the data in a Bayesian workflow. For instance:

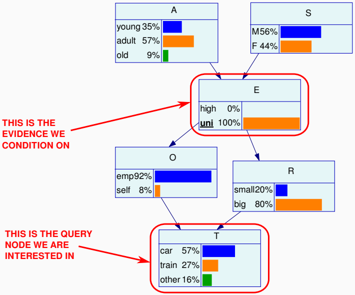
- Using whitelists and blacklists for specific arcs or arc patterns.
- Giving prior probabilities to particular patterns of arcs to encourage or discourage them from appearing in the BN.
- Using prior distributions on the values of specific parameters to drive their sign or their magnitude.

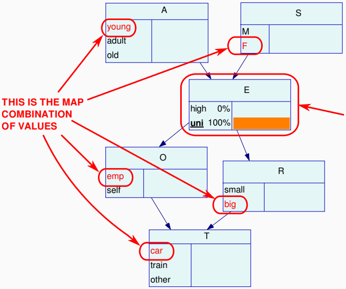
PROS: we get a BN that agrees with what we know, and we reduce the number of models we are considering.
CONS: we must trust expert knowledge to be correct, which is not trivial when there are multiple experts and they disagree with each other.

Using Bayesian Networks: Inference

A BN represents a working model of the world that a computer system can understand: we can ask it questions, and the computer system answers them algorithmically (no manual computations needed). This known as probabilistic inference in BNs. Commonly:

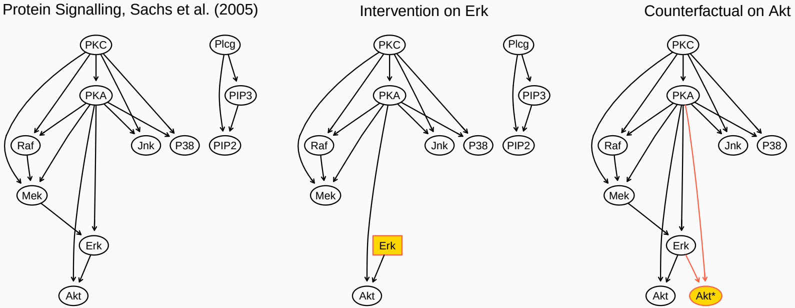
- Conditional probability queries: the probability of an event of interest given some evidence.
- Most probable explanation queries: the most probable value one or more variables will take given some evidence.





If we make some assumptions (no confounding), we can use BNs to perform causal inference [5] in a principled way:

- it improves our understanding of the underlying phenomenon;
- it allows us to target interventions to effect some desirable change to the underlying phenomenon;
- it allows us to reason about counterfactuals.



The link between BNs and causal inference also involves key concepts from experimental design [4]. Many key concepts from that field can be found in the network structure of a BN:

- Randomisation: randomising (say) a treatment means that the corresponding node in the BN cannot have any parents.
- Blocking (or stratification): adding one or more discrete nodes that do not have any parents but can be parents of outcomes.
- Confounding: unobserved variables that are parents of one or more nodes in the BN.
- Analysis of variance: explaining a statistically-significant proportion of variance implies a strong conditional association, which is what is used to perform structure learning.

Summary

- BNs provide an intuitive representation of the relationships linking heterogeneous sets of variables, which we can use for qualitative and quantitative reasoning.
- We can learn BNs from data while including prior knowledge as needed to improve the quality of the model.
- BNs subsume a large number of probabilistic models and thus can readily incorporate other techniques from statistics and experimental design.
- For most tasks we can start just reusing state-of-the-art, general purpose algorithms.
- Once learned, BNs provide a flexible tool for inference, both probabilistic and causal.

Thanks!

Any questions?

- D. I. Edwards.
  Introduction to Graphical Modelling.
  Springer, 2nd edition, 2000.
- S. L. Lauritzen and D. J. Spiegelhalter.
  Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion).
  Journal of the Royal Statistical Society: Series B (Statistical Methodology), 50(2):157–224, 1988.
- K. V. Mardia, J. T. Kent, and J. M. Bibby.
  Multivariate Analysis.
  Academic Press, 1979.
- D. C. Montgomery.
  Design and Analysis of Experiments.
  Wiley, 9th edition, 2017.
- J. Pearl, M. Glymour, and N. P. Jewell.
  Causal Inference in Statistics: A Primer.
  Wiley, 2016.
- M. Scutari and J.-B. Denis.
  Bayesian Networks with Examples in R.
  Chapman & Hall, 2nd edition, 2021.