Bayesian Networks in Reliability
The Good, the Bad, and the Ugly

Helge Langseth

Department of Computer and Information Science
Norwegian University of Science and Technology

MMR’07

NTNU
Outline

1 Introduction

2 The Good: Why Bayesian Nets are popular
   - Mathematical properties
   - Making decisions
   - Applications

3 The Bad: Building complex quantitative models
   - The model building process
   - The quantitative part
   - Utility theory

4 The Ugly: Continuous variables
   - Introduction
   - Approximations

5 Concluding remarks
A simple example: “Explosion”

\[ P(E, L, G, X, C) \]
A simple example: “Explosion”

\[ P(E, L, G, X, C) \]

\[ \text{pa}(X) = \{L, G\} \]
A simple example: “Explosion”

\[
\begin{align*}
E & : \text{Environment} \\
L & : \text{Leak} \\
G & : \text{GD failed} \\
X & : \text{Explosion} \\
C & : \text{Casualties}
\end{align*}
\]

\[
\begin{align*}
\text{pa}(X) &= \{L, G\} \\
\text{nd}(X) &= \{E, L, G\}
\end{align*}
\]

\[
P(E, L, G, X, C)
\]
A simple example: “Explosion”

\[ \begin{align*}
E & : \text{Environment} \\
L & : \text{Leak} \\
G & : \text{GD failed} \\
X & : \text{Explosion} \\
C & : \text{Casualties}
\end{align*} \]

\[ \begin{align*}
\text{pa}(X) &= \{L, G\} \\
\text{nd}(X) &= \{E, L, G\} \\
X & \perp\!
\!\!\!\perp E \mid \{L, G\}
\end{align*} \]

Other d-sep. rules: Jensen\&Nielsen (07)

\[ P(E, L, G, X, C) \]
A simple example: “Explosion”

\[ \begin{array}{c|cc}
G & E = \text{hostile} & E = \text{normal} \\
\hline
\text{yes} & \lambda_H \cdot \tau / 2 & \lambda_N \cdot \tau / 2 \\
\text{no} & 1 - \lambda_H \cdot \tau / 2 & 1 - \lambda_N \cdot \tau / 2 \\
\end{array} \]

\[ P(G \mid \text{pa}(G)) \]

\[ \text{pa}(X) = \{L, G\} \]

\[ \text{nd}(X) = \{E, L, G\} \]

\[ X \perp E \mid \{L, G\} \quad (\text{Hence, } P(X \mid E, L, G) = P(X \mid L, G)) \]

Other d-sep. rules: Jensen & Nielsen (07)

\[ P(E, L, G, X, C) = P(E) \cdot P(L \mid E) \cdot P(G \mid E, L) \cdot P(X \mid E, L, G) \cdot P(C \mid E, L, G, X) \]
\[ = P(E) \cdot P(L \mid E) \cdot P(G \mid E) \cdot P(X \mid L, G) \cdot P(C \mid X) \]

Markov properties \( \Leftrightarrow \) Factorization property
Outline

1 Introduction

2 The Good: Why Bayesian Nets are popular
   - Mathematical properties
   - Making decisions
   - Applications

3 The Bad: Building complex quantitative models
   - The model building process
   - The quantitative part
   - Utility theory

4 The Ugly: Continuous variables
   - Introduction
   - Approximations

5 Concluding remarks
What the mathematical foundation has to offer

**Intuitive representation:** Almost defined as “box-diagram with formal meaning”. Causal interpretation natural in many cases.

**Efficient representation:** The number of required parameters are reduced. If all variables are binary, the example requires 11 “local” parameters, compared to the 31 “global” parameters of the full joint.

**Efficient calculations:** Efficient calculations of any joint distribution $P(x_i, x_j)$ or conditional distribution $P(x_k \mid x_\ell, x_m)$.

**Model estimation:** Estimating parameters (fixed structure) via EM, estimating structure by discrete optimization techniques.
Influence diagrams: The "Explosion" example revisited

- $E$: Environment
- $L$: Leak
- $G$: GD failed
- $X$: Explosion
- $C$: Casualties
- $\text{Cost}_1$
- $\text{Cost}_2$
- $\text{SSM}$
Influence diagrams: The “Explosion” example revisited

- **$E$: Environment**
- **$L$: Leak**
- **$G$: GD failed**
- **$X$: Explosion**
- **$C$: Casualties**

Costs:
- **$Cost_1$**
- **$SSM$**
- **$F$: Effectiveness, SSM**
- **Test interval**
- **$Cost_2$**
- **$Cost_3$**
An application: Troubleshooting

**Solutions**
- Enable the Wireless network on your computer: 3.0
- Run your anti-virus software and your anti-spyware software to clean your computer: 0.0
- Configure your firewall correctly: 0.0

**Questions**
- Are you using a HUB or a switch to share the Internet connection with other computers: 342.3
- Which software firewall are you using? (Y): 486.5
- Are you using any: 0.0

**Causes**
- Computer is infected with spyware: 36.8
- Incorrect firewall configuration: 18.8
- Incorrect anti-virus configuration: 15.8
- Wireless configuration/radio not: 5.3

**Internet Connection Problems**

**Performed steps**
- What do you need help with?
  - My Internet connection seems incredibly slow
- How are you connected to the Internet?
  - ADSL
Underlying model

The diagram represents a Bayesian Network with the following nodes:

- **TOP**
- **C_1**, **C_2**, **C_3**, **C_4**
- **X_1**, **X_2**, **X_3**, **X_4**, **X_5**

The network is structured as follows:

- **TOP** is connected to **C_1**, **C_2**, **C_3**, and **C_4**.
- **C_1** connects to **X_1**.
- **C_2** connects to **X_2**.
- **C_3** connects to **X_3**.
- **C_4** connects to **X_4**.
- **X_5** is linked to the system-layer.
The Good: Why Bayesian Nets are popular

Applications

Underlying model

\begin{center}
\begin{tikzpicture}
  \node[shape=circle,draw=blue] (1) at (0,0) {$C_1$};
  \node[shape=circle,draw=blue] (2) at (1,0) {$C_2$};
  \node[shape=circle,draw=blue] (3) at (2,0) {$C_3$};
  \node[shape=circle,draw=blue] (4) at (3,0) {$C_4$};
  \node[shape=circle,draw=blue] (5) at (4,0) {TOP};
  \node[shape=circle,draw=blue] (6) at (0,-1) {$X_1$};
  \node[shape=circle,draw=blue] (7) at (1,-1) {$X_2$};
  \node[shape=circle,draw=blue] (8) at (2,-1) {$X_3$};
  \node[shape=circle,draw=blue] (9) at (3,-1) {$X_4$};
  \node[shape=circle,draw=blue] (10) at (4,-1) {$X_5$};
  \node[shape=circle,draw=red] (11) at (2,-2) {$E$};

  \path[->,blue]
    (5) edge (1)
    (5) edge (2)
    (5) edge (3)
    (5) edge (4);
  \path[->,blue]
    (1) edge (6)
    (1) edge (7)
    (2) edge (6)
    (2) edge (7)
    (3) edge (8)
    (4) edge (9)
    (4) edge (10);
  \path[->,red]
    (11) edge (5)
    (6) edge (11)
    (7) edge (11)
    (8) edge (11)
    (9) edge (11)
    (10) edge (11);
\end{tikzpicture}
\end{center}

system-layer
Underlying model

The Good: Why Bayesian Nets are popular

Applications

Bayesian Networks in Reliability

Helge Langseth
Underlying model

The Good: Why Bayesian Nets are popular

Applications

TOP

C₁  C₂  C₃  C₄

X₁  X₂  X₃  X₄  X₅

A₁  A₂  A₃  A₄  A₅

R₁  R₂  R₃  R₄  R₅

system-layer

action-layer

result-layer

Helge Langseth  Bayesian Networks in Reliability
Underlying model

- **Question-layer**: $Q_S$
- **System-layer**: $X_1, X_2, X_3, X_4, X_5$
- **Action-layer**: $A_1, A_2, A_3, A_4, A_5$
- **Result-layer**: $R_1, R_2, R_3, R_4, R_5$
Other applications

- Software reliability
- Modelling Organizational factors (e.g., the SAM-Framework)
- Explicit models of dynamics (e.g., repairable systems, phase-mission-systems, monitoring systems)
- Some of these can be seen at the Bayes net sessions later today
Outline

1. Introduction
2. The Good: Why Bayesian Nets are popular
   - Mathematical properties
   - Making decisions
   - Applications
3. The Bad: Building complex quantitative models
   - The model building process
   - The quantitative part
   - Utility theory
4. The Ugly: Continuous variables
   - Introduction
   - Approximations
5. Concluding remarks
Phases of the model building process

**Step 0 – Decide what to model:** Select the boundary for what to include in the model.

**Step 1 – Defining variables:** Select the important variables in the domain.

**Step 2 – The qualitative part:** Define the graphical structure that connects the variables.

**Step 3 – The quantitative part:** Fix parameters to specify each $P(x_i | pa(x_i))$. This is the ‘bad’ part.

**Step 4 – Verification:** Verification of the model.
Consider a binary node with $m$ binary parents. The CPT $P(y|z_1, \ldots, z_m)$ contains $2^m$ parameters.

**Naïve approach:** $2^m$ conditional probabilities:
All parameters are required if no other assumptions can be made.
Consider a binary node with \( m \) binary parents. The CPT \( P(y|z_1, \ldots, z_m) \) contains \( 2^m \) parameters.

**Naïve approach:** \( 2^m \) conditional probabilities

**Deterministic relations:** Parameter free:

\( Y \) considered a function of its parents, e.g.,

\[
\{Y = \text{fail}\} \iff \{Z_1 = \text{fail}\} \lor \{Z_2 = \text{fail}\} \lor \ldots \lor \{Z_m = \text{fail}\}.
\]
Consider a binary node with \( m \) binary parents. The CPT \( P(y|z_1, \ldots, z_m) \) contains \( 2^m \) parameters.

**Naïve approach:** \( 2^m \) conditional probabilities

**Deterministic relations:** Parameter free

**Noisy OR relation:** \( m + 1 \) conditional probabilities:

Independent inhibitors \( Q_1, \ldots, Q_m \); Assume
\[
\{ Q_1 = \text{fail} \} \lor \cdots \lor \{ Q_m = \text{fail} \} \implies \{ Y = \text{fail} \}.
\]

For each \( Q_i \) we have
\[
P(Q_i = \text{fail}|Z_i = \text{fail}) = q_i,
\]
\[
P(Q_i = \text{fail}|Z_i = \neg \text{fail}) = 0.
\]

“Leak probability”:
\[
P(Y = \text{fail}|Q_1 = \ldots = Q_m = \neg \text{fail}) = q_0.
\]
Consider a binary node with $m$ binary parents. The CPT $P(y|z_1, \ldots, z_m)$ contains $2^m$ parameters.

**Naïve approach:** $2^m$ conditional probabilities

**Deterministic relations:** Parameter free

**Noisy OR relation:** $m + 1$ conditional probabilities

**Logistic regression:** From $m + 1$ to $2^m$ regression parameters:

$Y$ is dependent variable in logistic regression with $Z_i$’s as “covariates”:

$$
\log \left( \frac{p_{z_1, \ldots, z_m}}{1 - p_{z_1, \ldots, z_m}} \right) = \eta_0 + \sum_j \eta_j z_j + \sum_i \sum_j \eta_{ij} z_i \cdot z_j + \ldots
$$
Consider a binary node with \( m \) binary parents. The CPT \( P(y|z_1, \ldots, z_m) \) contains \( 2^m \) parameters.

**Naïve approach:** \( 2^m \) conditional probabilities

**Deterministic relations:** Parameter free

**Noisy OR relation:** \( m + 1 \) conditional probabilities

**Logistic regression:** From \( m + 1 \) to \( 2^m \) regression parameters

**IPF procedure:** \( m + 1 \) marginal distributions, \( m \) CPRs:

Find a *joint* PT over \( Z_1, \ldots, Z_m, Y \) with given CPRs.

Assume \( m = 1 \), \( p_0(z, y) \) initialized to fit CPR.

\[ p'_k(z, y) \leftarrow p_{k-1}(z, y) \cdot p(z)/\sum_y p_{k-1}(z, y) \]

\[ p_k(z, y) \leftarrow p'_k(z, y) \cdot p(y)/\sum_z p'_k(z, y) \]
Consider a binary node with \( m \) binary parents. The CPT \( P(y|z_1, \ldots, z_m) \) contains \( 2^m \) parameters.

**Naïve approach:** \( 2^m \) conditional probabilities

**Deterministic relations:** Parameter free

**Noisy OR relation:** \( m + 1 \) conditional probabilities

**Logistic regression:** From \( m + 1 \) to \( 2^m \) regression parameters

**IPF procedure:** \( m + 1 \) marginal distributions, \( m \) CPRs

**Special structures:** From 2 to \( 2^m \) conditional probabilities:
- \( Y \) defined, e.g., by rules such as
  
  \[ P(Y = \text{fail}|Z_1 = \text{fail}, \ldots, Z_m = \text{fail}) = p_1, \text{ but} \]
  \[ P(Y = \text{fail}|z_1, \ldots, z_m) = p_2 \text{ for all other configurations } z. \]
The quantitative part: Defining $P(y|\text{pa}(y))$

Consider a binary node with $m$ binary parents. The CPT $P(y|z_1, \ldots, z_m)$ contains $2^m$ parameters.

Naïve approach: $2^m$ conditional probabilities

Deterministic relations: Parameter free

Noisy OR relation: $m + 1$ conditional probabilities

Logistic regression: From $m + 1$ to $2^m$ regression parameters

IPF procedure: $m + 1$ marginal distributions, $m$ CPRs

Special structures: From 2 to $2^m$ conditional probabilities

**Qualitative BNs:** No quantitative parameters:

Only *qualitative* effects modelled (and later calculated). From $m$ to $2^m$ qualitative effects (‘+’, ‘0’ or ‘−’).
Consider a binary node with $m$ binary parents. The CPT $P(y|z_1, \ldots, z_m)$ contains $2^m$ parameters.

**Naïve approach:** $2^m$ conditional probabilities

**Deterministic relations:** Parameter free

**Noisy OR relation:** $m + 1$ conditional probabilities

**Logistic regression:** From $m + 1$ to $2^m$ regression parameters

**IPF procedure:** $m + 1$ marginal distributions, $m$ CPRs

**Special structures:** From 2 to $2^m$ conditional probabilities

**Qualitative BNs:** No quantitative parameters

**Alternative solutions:** No conditional probabilities:
- Other frameworks (like vines), or parameter estimation.
Utility Theory
Utility Theory

Utility - 2

Pareto boundary

Utility - 1
Outline

1 Introduction

2 The Good: Why Bayesian Nets are popular
   - Mathematical properties
   - Making decisions
   - Applications

3 The Bad: Building complex quantitative models
   - The model building process
   - The quantitative part
   - Utility theory

4 The Ugly: Continuous variables
   - Introduction
   - Approximations

5 Concluding remarks
The original calculation procedure only supports a restricted set of distributional families:

- Continuous variables must have Gaussian distributions.
- Discrete variables should only have discrete parents.
- Gaussian parents of Gaussians are partial regression coefficients of their children.

These classes of distributions are not sufficient for reliability analysis. This is the ‘ugly’ part.
An example model: The THERP methodology

- Used to model human ability to perform in certain settings (measured as binary variables)
- Known environment variables, like "Level of feedback"

This is simple. The probability of a subject failing to perform task $T_i$ is:

$$P(T_i = t_i | z) = \left(1 + \exp \left(-w_i'z\right)\right)^{-1}$$
An example model: The THERP methodology

- Used to model human ability to perform in certain settings (measured as binary variables)
- Known environment variables, like "Level of feedback"

We can also have latent traits, which are unknown and vary between subjects (like "Omitting a step in a procedure").

In this case, the model is a "latent trait model" (similar to binary factory analyzer).

In the following we will focus on a situation with two latent "traits", and one "task".
Why is this difficult

Assume we have one observation $D_1 = \{1\}$, and parameters $w_1 = [1 \ 1]^T$.

The likelihood is given by

$$P(T = 1) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}^2} \frac{\exp \left( - \left\{ \frac{(z_1 - \mu_1)^2}{2\sigma_1^2} + \frac{(z_2 - \mu_2)^2}{2\sigma_2^2} \right\} \right)}{1 + \exp(-z_1 - z_2)} \, dz,$$

which has no known analytic representation in general. Hence, we cannot do the required calculations in this model.

Note! This is true even if we choose not to use Bayesian networks as our modelling language.
Attempts to find $f(z_1, z_2 \mid T = 1)$ and $P(T = 1)$

Numerical approximation:

$$P(T = 1) = 0.49945$$
$$\text{CPU} = 600 \text{ msec.}$$

$$f(z_1, z_2 \mid T = 1)$$

1000 × 1000 grid
Attempts to find \( f(z_1, z_2 | T = 1) \) and \( P(T = 1) \)

**Discretization:**

- Every continuous variable is “translated” into a discrete one.
- The more discrete states used the higher ...
  - approximation quality.
  - computational complexity.
- “Tricks” are available to find number of states and where to set split-points, including *dynamic* discretization.
Attempts to find $f(z_1, z_2 \mid T = 1)$ and $P(T = 1)$

Discretization:

$P(T = 1) = 0.49761$

CPU = 2 msec.

$f(z_1, z_2 \mid T = 1)$

$5 \times 5$ discretization grid
Mixtures of Truncated Exponentials:

- In standard discretization, the continuous variable is approximated by a step-function.

- Calculations are also possible if each ‘step’ is replaced by a truncated exponential.

- A single variable density is split into \( n \) intervals \( I_k \), \( k = 1, \ldots, n \), each approximated by

\[
    f^*(z) = a_0^{(k)} + \sum_{i=1}^{m} a_i^{(k)} \exp \left( b_i^{(k)} z \right) \quad \text{for } z \in I_k
\]

We typically see \( 1 \leq n \leq 4 \) and \( 0 \leq m \leq 2 \).

- Clever parameter choices are tabulated for many standard distributions.
Attempts to find $f(z_1, z_2 | T = 1)$ and $P(T = 1)$

Mixtures of Truncated Exponentials:

$$P(T = 1) = 0.49914$$

CPU = 4 msec.

$$f(z_1, z_2 | T = 1)$$

S. Acid et al.: ELVIRA
Attempts to find $f(z_1, z_2 \mid T = 1)$ and $P(T = 1)$

**Markov Chain Monte Carlo:**
- Works well with Bayesian Networks, as independence statements can be exploited for fast simulation:
  - Metropolis-Hastings works directly out-of-the-box.
  - Gibbs sampling might sometimes require clever adaptation.
Attempts to find $f(z_1, z_2 \mid T = 1)$ and $P(T = 1)$

Markov Chain Monte Carlo:

$P(T = 1) = \frac{0.49821}{0.49821}$

CPU $= 32 \cdot 10^3$ msec.

$f(z_1, z_2 \mid T = 1)$

W. Gilks et al.: BUGS
Attempts to find $f(z_1, z_2 | T = 1)$ and $P(T = 1)$

Variational Approximations:

$P(T = 1|\mathbf{x})$

Replace a tricky function $h(v)$ with family of simple functions indexed by $\xi$, $\hat{h}(v, \xi)$, such that $h(v) = \sup_{\xi} \hat{h}(v, \xi)$.

Note: $\log(P(T = 1|v)) = v/2 - \log(\exp(v/2) + \exp(-v/2))$, where the last term is convex in $v^2$. 
Attempts to find \( f(z_1, z_2 \mid T = 1) \) and \( P(T = 1) \)

Variational Approximations:

Define

- \( A(z_1, z_2) = z_1 + z_2 \)
- \( \lambda(\xi) = \frac{\exp(-\xi) - 1}{4\xi(1 + \exp(-\xi))} \)

Variational approximation

The variational approximation of \( P(T = 1 \mid z) \) is

\[
\tilde{P}(T = 1 \mid z, \xi) = \frac{\exp \left[ (A(z) - \xi)/2 + \lambda_i(\xi) \cdot (A(z)^2 - \xi^2) \right]}{1 + \exp(-\xi)}.
\]

Can be shown that the best choice is

\[ \xi \leftarrow \sqrt{\mathbb{E} \left[ (Z_1 + Z_2)^2 \, \mid \, T = 1 \right]} \]

\[ \implies \text{We need to iterate} \]
Attempts to find $f(z_1, z_2 \mid T = 1)$ and $P(T = 1)$

**Variational Approximations:**

$$P(T = 1) = 0.49828$$

CPU = 17 msec.

$$f(z_1, z_2 \mid T = 1)$$

J. M. Winn: VIBES
Attempts to find $f(z_1, z_2 | T = 1)$ and $P(T = 1)$

Other approaches:
A number of other approaches are also being examined

- Laplace-approximation
- Transformation into Mixture-of-Gaussian models
- Other frameworks, like Vines
- etc.
1. Introduction

2. The Good: Why Bayesian Nets are popular
   - Mathematical properties
   - Making decisions
   - Applications

3. The Bad: Building complex quantitative models
   - The model building process
   - The quantitative part
   - Utility theory

4. The Ugly: Continuous variables
   - Introduction
   - Approximations

5. Concluding remarks
Bayesian Networks’ popularity is increasing, also in the reliability community.

The main features (as seen from our community) are:
- Constitute an intuitive modelling ‘language’.
- High level of modelling flexibility.
- Efficient calculations based on utilization of the conditional independence structures encoded in the graph.
- Cost efficient representation.

Building models can still be time consuming.

Problem owners lack training in using BNs:
- Users more confident when using traditional frameworks, like, e.g., FT modelling.
- The calculations may be too complex to understand.

Most important research focus (for this community) is to find good approximations to handle continuous variables.
A number of people have helped or worked with me on the topics covered in this presentation:

**Bayesian Network Models:**
- Thomas D. Nielsen, Finn V. Jensen, Jiří Vomlel

**Bayesian Networks in Reliability:**
- Luigi Portinale, Claus Skaanning

**Continuous Variables:**
- Antonio Salmerón, Rafael Rumí

**Reliability Models:**
- Bo Lindqvist, Tim Bedford, Roger M. Cooke, Jørn Vatn
Further reading


