Bayesian Networks in Credit Rating

Samuel Gerssen

March 12, 2004

Prof. Dr. H. Koppelaar
Drs. Dr. L.J.M. Rothkrantz
Ir. W. de Jong
Dr. K. van der Meer
Dr. D. Fokkema (ABN AMRO)
Dr. Ir. M.J. Druzdzel (University of Pittsburgh)

University of Pittsburgh

Bayesian Networks in Credit Rating

TU Delft

ABN·AMRO

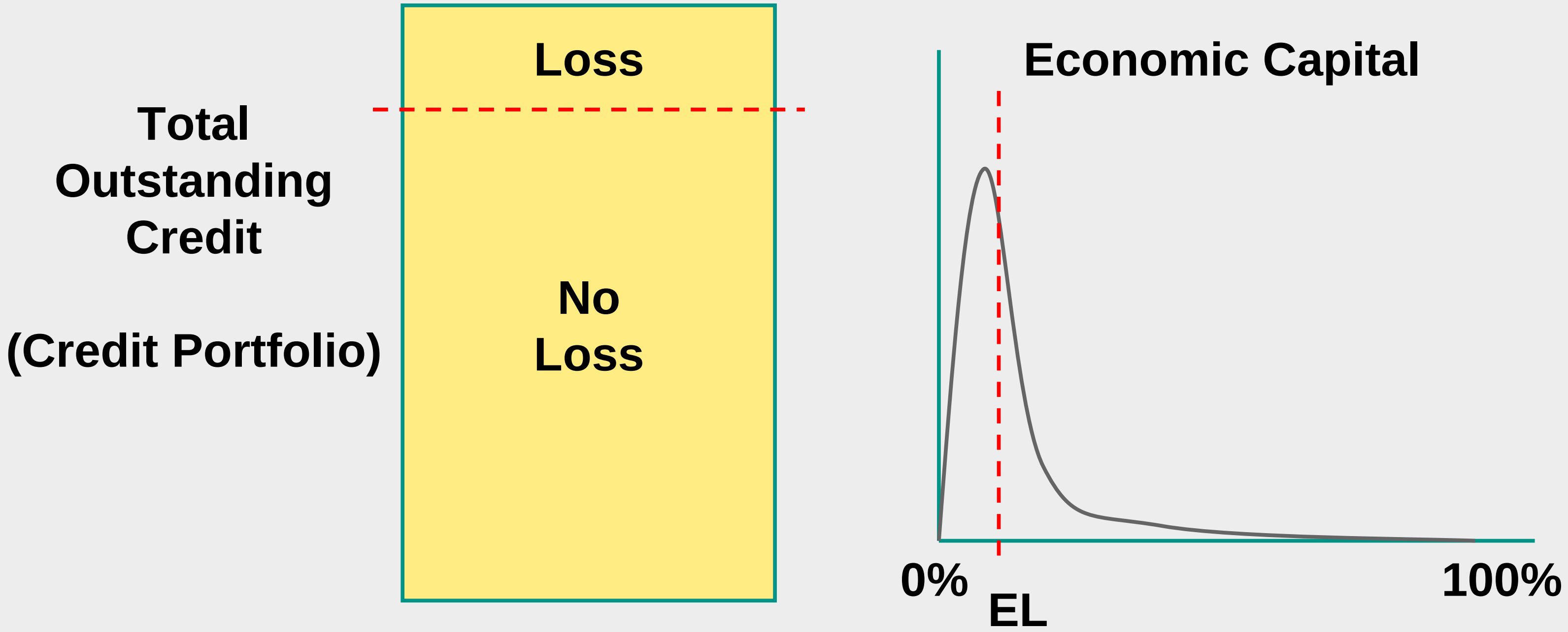
University of Pittsburgh

Contents

- Credit Rating
- Bayesian Networks
- Assignment
- Bayesian Networks Learning Research
- Bayesian Networks in Credit Rating
- Conclusions

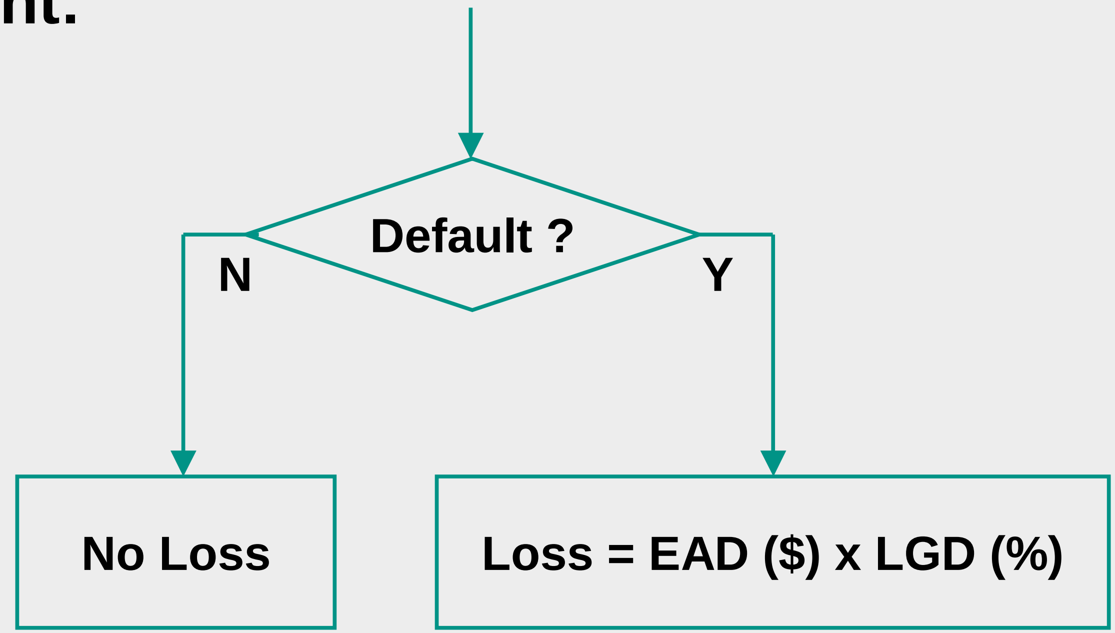
Credit Rating - Definition

Credit Rating = Assessment of risk on credit portfolios



Credit Rating – Expected Loss per client

Loss per client:



Expected Loss per client = PD (EAD x LGD)

Credit Rating – Expected Loss

EL = PD (EAD x LGD)

PD = Probability of Default

EAD = Exposure At Default

LGD = Loss Given Default

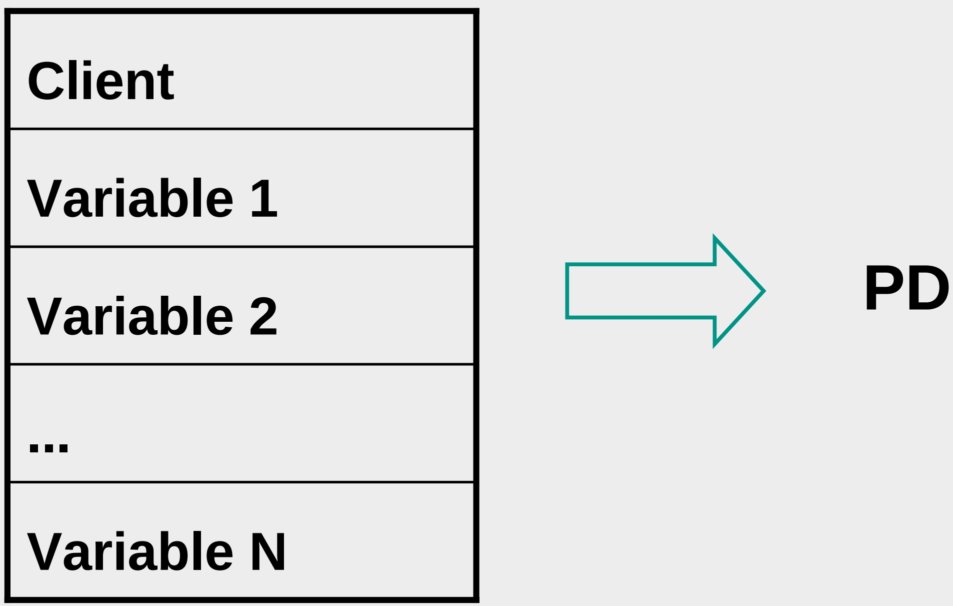
Credit Rating – Probability of Default

Probability of Default (PD)

= the probability that a client can not meet its repayment obligations between now and 1 year

Credit Rating – PD Model

PD calculation for each client



Credit Rating – Modeling steps

Construction of data set:

- **Select portfolio information from previous year**
- **Add default history from last 12 months**

Modeling technique: Binary logistic regression

Scoring: 5-fold cross validation

Credit Rating - Binary Logistic Regression

PD depends on variables (x_1 .. x_n) and parameters (β_0 .. β_n)

$$ln\left(\frac{PD}{1-PD}\right) = \beta_0 + \beta_1 x_1 + .. + \beta_n x_n$$

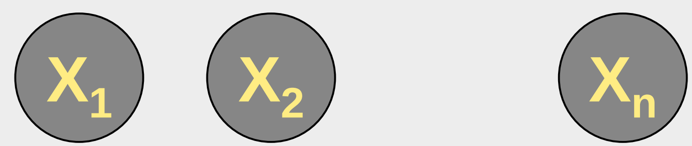
Calculation of (β_0 .. β_n) using maximum likelihood

Bayesian Networks - Definitions

A Bayesian Network (BN) is a probabilistic model of variables and their causal relations.

A BN is a directed acyclic graph where:

Variables are nodes..



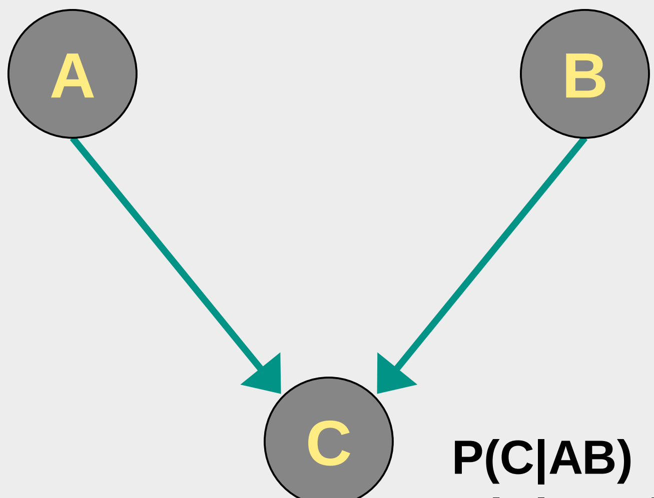
Causal relations are vertices...

Bayesian Networks – CPT

Conditional Probability Tables (CPT)

P(A)=0.02

P(B)=0.3



P(C|AB) = 0.95
P(C|A¬B) = 0.8
P(C|¬AB) = 0.1
P(C|¬A¬B) = 0.01

Bayesian Networks – Inference

- **Bayes' Theorem**

$$p(a|b)=\frac{p(b|a)p(a)}{p(b)}$$

- **Expansionrule**

$$p(a)=p(ab)+p(a\overline{b})=p(a|b)p(b)+p(a|\overline{b})p(\overline{b})$$

$$p(a)=p(abc)+p(ab\overline{c})+p(a\overline{b}c)+p(a\overline{b}\overline{c})$$

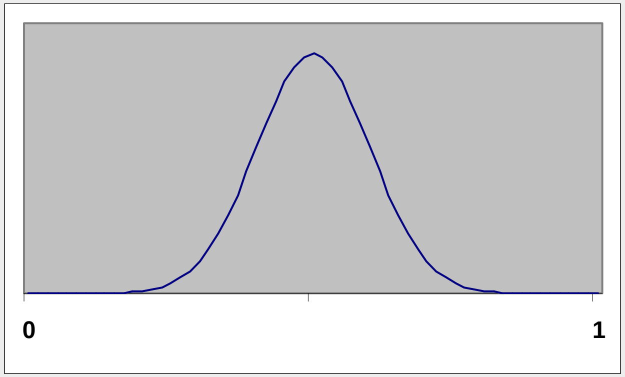
Bayesian Networks - Learning

θ = **probability of variable** X

$p(\theta)$ = **probability distribution of** θ

= **belief about** θ

Example of $p(\theta)$



Bayesian Networks – Belief Updating

Observation: O_i

The belief about θ **is now updated:**

$$p(\theta \mid O_i) = \begin{cases} c \cdot \theta \cdot p(\theta) & \text{if } \; O_i : X = 1 \\ c \cdot (1-\theta) \cdot p(\theta) & \text{if } \; O_i : X = 0 \end{cases}$$

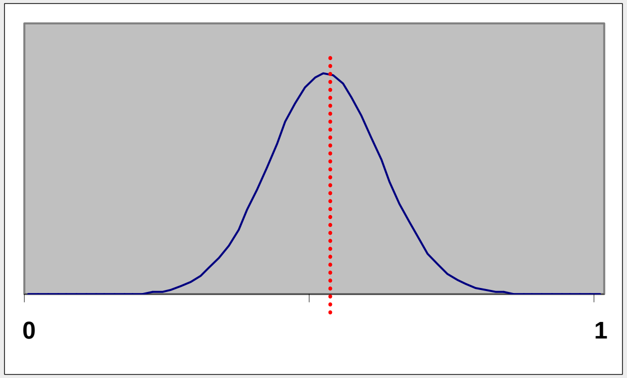
c **= normalization coefficient**

Bayesian Networks – Belief Updating

Observation O_i **:** X=1

Belief about θ **is updated**

Updated distribution $p(\theta | O_i) = p(\theta) \cdot \theta$



$$p(\theta | O_i)$$

Bayesian Networks – Beta distribution

Beta distribution with parameters α and β:

$$p(\theta \mid \alpha\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

α **= the number of observations** $X_i = 1$

β **= the number of observations** $X_i = 0$

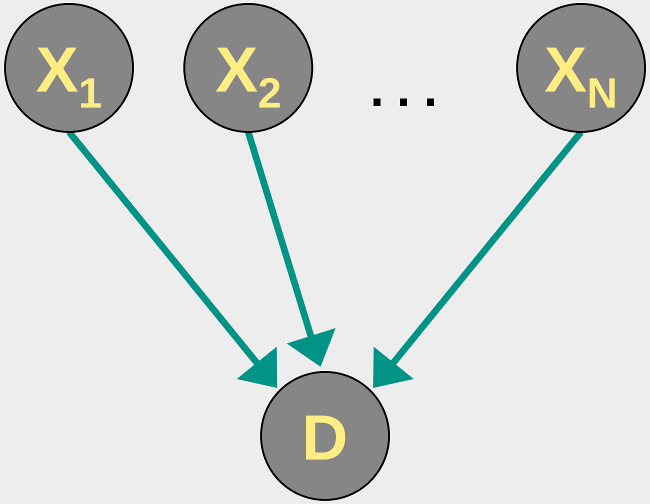
Property: after updating again Beta distribution

Bayesian Networks - Structures

Applied network structures in this research:

- **CPT model**
- **Naive Bayes model**
- **Noisy-MAX model**

Bayesian Networks – CPT model

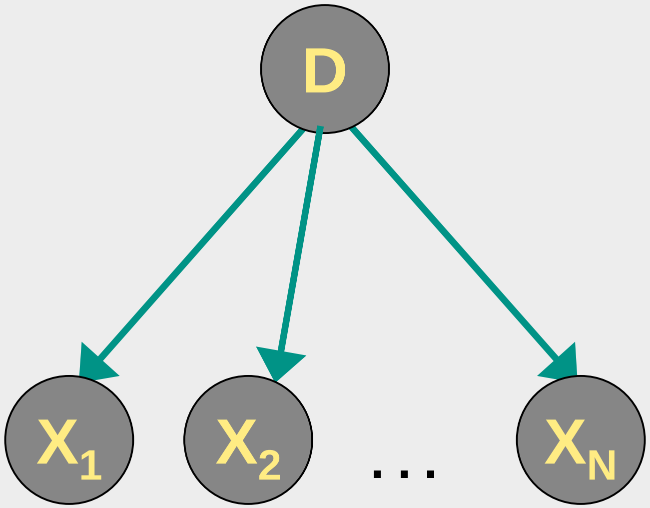


Size of CPT = $(s_p)^p$

s_p **= number of parent states**

p **= number of parents**

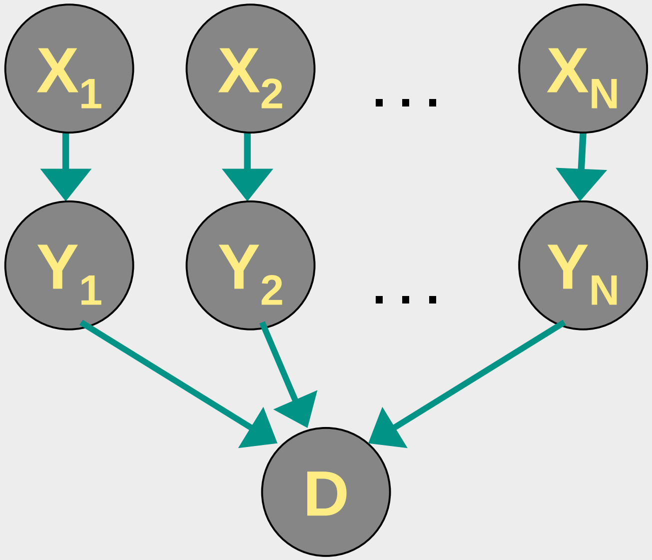
Bayesian Networks – Naive Bayes model



Unrealistic assumption:

Predictive variables dependent on effect variable.

Bayesian Networks – Noisy-MAX model



$P(Y_i \mid X_i) = c_i$

State of D **= logical MAX of** $Y_1..Y_N$

Assignment

- **Provide validated model for PD estimation for a credit portfolio. Logistic regression and Bayesian networks should be applied.**
- **Explore Bayesian network modeling and improve parameter learning techniques.**

Remainder of Presentation

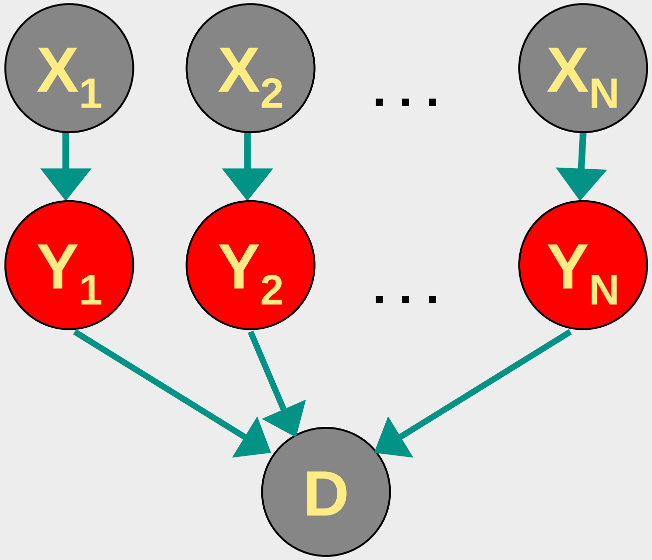
- Credit Rating
- Bayesian Networks
- Assignment
- **Bayesian Network Learning Research:**

  - **Noisy-MAX learning algorithm**

  - **Prior / Posterior learning**
- **Bayesian Networks in Credit Rating**
- **Conclusions**

Noisy-MAX Learning

Parameter learning based on observations not possible



$Y_1..Y_N$ **unknown**

Noisy-MAX Learning - Approach

- **Construct noisy-MAX CPT for node D, based on nodes $X_1..X_N$ and parameters $c_1..c_N$**
- **Learn normal CPT (not noisy-MAX) for node D, based on $X_1..X_N$**
- **Minimize difference between the two CPT's**

Noisy-MAX Learning – Noisy-MAX CPT

CPT for node Y_1

	$X_1 = 2$	$X_1 = 1$	$X_1 = 0$
$Y_1 = 2$	c_{122}	c_{121}	0
$Y_1 = 1$	c_{112}	c_{111}	0
$Y_1 = 0$	$1 - c_{112} - c_{122}$	$1 - c_{111} - c_{121}$	1

Noisy-MAX Learning – Noisy-MAX CPT

Cumulative CPT for node Y_1

	$X_1 = 2$	$X_1 = 1$	$X_1 = 0$
$Y_1 = 2$	c_{122}	c_{121}	0
$Y_1 = 1$	$c_{112} + c_{122}$	$c_{111} + c_{121}$	0
$Y_1 = 0$	1	1	1

$$p(Y_1 \geq y_1 \mid X_1 = x_1)$$

Noisy-MAX Learning – Noisy-MAX CPT

Transformation of cumulative into normal:

$$p(Y_1 = y_1 \mid X_1 = x_1)$$
$$= p(Y_1 \geq y_1 \mid X_1 = x_1) - p(Y_1 \geq y_1 + 1 \mid X_1 = x_1)$$

Noisy-MAX Learning – Noisy-MAX CPT

For MAX gates, the following holds:

$$
\begin{aligned}
& p(D \geq d \mid X_1 = x_1, .., X_N = x_N) \\
& \quad = p((Y_1 \geq d \mid X_1 = x_1) \vee (Y_2 \geq d \mid X_2 = x_2) \vee \ldots) \\
& \quad = 1 - \prod_{i=1}^{N} 1 - p(Y_i \geq d \mid X_i = x_i)
\end{aligned}
$$

→ Values for cumulative noisy-MAX CPT

Noisy-MAX Learning – Noisy-MAX CPT

Entries in noisy-MAX CPT defined as:

$$\theta_{ij} = \prod_{x_p^r \in \mathbf{X}} \sum_{k=1}^{j} q_{prk} - \prod_{x_p^r \in \mathbf{X}} \sum_{k=1}^{j-1} q_{prk}$$

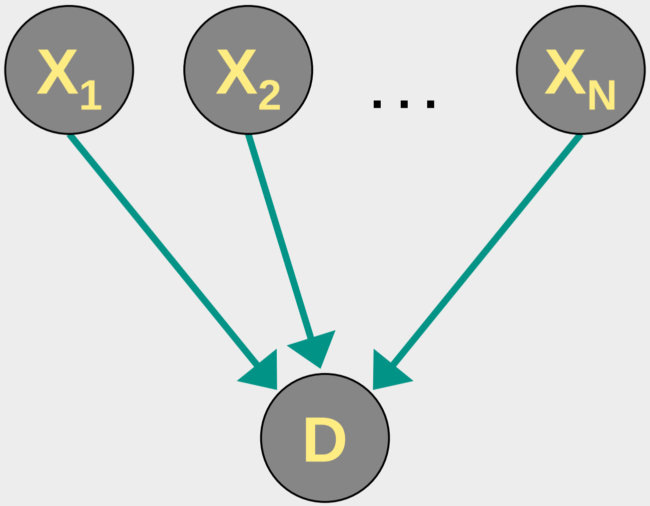
where

$$\theta_{ij} = p(D = d_j \,|\, \mathbf{X} = \mathbf{x_i})$$

$$q_{ijk} = p(Y_i = y_k \,|\, X_i = x_j)$$

Noisy-MAX Learning – Normal CPT

Normal CPT obtained by parameter learning



Noisy-MAX Learning – Distance

- **Euclidean Distance**

$$\sum_i \sum_j \left(\theta_{ij}^{CPT} - \theta_{ij}^{noisy-MAX}\right)^2$$

- **Kullback-Leibler Distance**

$$\sum_i \sum_j \theta_{ij}^{CPT} \, ln \frac{\theta_{ij}^{CPT}}{\theta_{ij}^{noisy-MAX}}$$

Noisy-MAX Learning – Optimization

Gradient Descent:

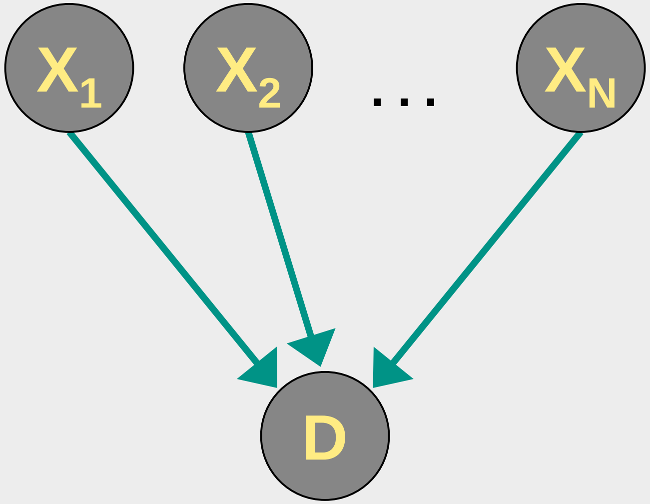
```
repeat
    for each noisy-MAX parameter c do
          c* = c +/- step
          d = CalculateDistance
          if d < d_min then
               d_min = d
               c = c*
          end if
    end for
until no distance reduction possible
```

Noisy-MAX Learning – Summary

- **Algorithm performs well**
- **Very fast**
- **Global minimum guaranteed**
- **EM performs better**
- **CPT not optimal starting point**

Prior / Posterior – Concept

CPT approach..



..but

- **prior is initial value in CPT**
- **posterior is prior updated by observations**

Prior / Posterior - Motive

CPT model has too many degrees of freedom

→ Rare entries have unreliable values

→ Resulting model is not stable

→ Setting priors can reduce instability

Prior / Posterior - Stages

1. **Prior assessment**
2. **Posterior learning**
3. **Merging**

Prior / Posterior – Posterior learning

CPT probabilities:

- **Beta distribution (binary variables)**

$$p(\theta \mid \alpha\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- **Dirichlet distribution (multinomial variables)**

$$p(\theta \mid \alpha_1..\alpha_N) = \frac{\Gamma\left(\sum_{i=1}^{N} \alpha_i\right)}{\prod_{i=1}^{N} \Gamma(\alpha_i)} \prod_{i=1}^{N} \theta^{\alpha_i - 1}$$

Prior / Posterior – Posterior learning

Expectation of Beta distribution:

$$\int \theta \; Beta(\alpha\beta) \; d\theta = \frac{\alpha}{\alpha + \beta}$$

Expectation of Dirichlet distribution:

$$\int \theta_i \; Dir(\alpha_1..\alpha_N) \; d\theta_i = \frac{\alpha_i}{\sum_j \alpha_j}$$

Prior / Posterior – Posterior learning

CPT entries for variable X with states $x_1..x_N$:

$$\theta_{ij} = p(x_j \mid \mathbf{p}_i) = \frac{o(\mathbf{p}_i, x_j)}{\sum_{k=1}^{N} o(\mathbf{p}_i, x_k)}$$

$\mathbf{p}_i$ **= i-th combination of parent states**

$o(\mathbf{p}_i, x_j)$ **= number of observations $(\mathbf{p}_i, x_j)$ in the data set.**

Prior / Posterior – Prior assessment

Goal: Stable priors for every entry in the CPT

Solution:

1. **Build naive Bayes model**
2. **Learn parameters of CPT's**
3. **Perform inference to obtain** $p(d|x_1..x_N)$

Prior / Posterior – Prior assessment

Inference in naive Bayes:

$$
\begin{aligned}
p(d \mid x_1..x_N) \\
&= \frac{p(x_1..x_N \mid d)\cdot p(d)}{p(x_1..x_N)} \\
&= \frac{\prod_i p(x_i \mid d)\cdot p(d)}{\prod_i p(x_i)} \\
&= \frac{\prod_i p(x_i \mid d)\cdot p(d)}{\prod_i \left(p(x_i \mid d)\cdot p(d) + p(x_i \mid \bar{d})\cdot p(\bar{d})\right)}
\end{aligned}
$$

Prior / Posterior – Merging

Merging of priors and posteriors: weighting

$$\theta_{ij} = \frac{(w \cdot p_{ij}) + o(\mathbf{p}_i, x_j)}{w + \sum_{k=1}^{j} o(\mathbf{p}_i, x_k)}$$

weight w = 'number of observations' for the prior p_{ij}

Prior / Posterior – Summary

+ Very stable for rare event data

+ Naive Bayes priors guarantee high performance

+ Posterior CPT learning provides flexibility

BN in Credit Rating – Project results

- Construction of data set
- Development of logistic regression model
- BN tests with C++ tool
- Development of BN models
- Algorithm for Noisy-MAX
- Design & implementation of prior / posterior learning
- Comparison of BN and logistic regression

BN in Credit Rating - Comparison

Model	GINI score
Logistic Regression	78.3%
BN – CPT	55.6%
BN – Noisy-MAX	75.4%
BN – Naive Bayes	77.5%
BN – Prior / Posterior	77.3%

BN in Credit Rating - Comparison

Logistic regression

- **Favored by data selection process**
- **Optimization technique is by nature bound to yield good results**
- **Provides the right number of degrees of freedom**

BN in Credit Rating - Comparison

Bayesian Networks

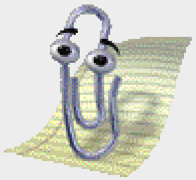
- Expert knowledge could not be incorporated
- CPT has too many degrees of freedom
- Naive Bayes performs well; apparently right number of degrees of freedom
- Logistic regression PD calculation and naive Bayes inference show similarities

Conclusions & Recommendations

BN is a very powerful modeling technique:

- **Expert knowledge combined with statistical data**
- **High acceptance due to transparency**
- **Good performance in diagnosis**

  (medical, trouble shooting, Microsoft Office Assistant)

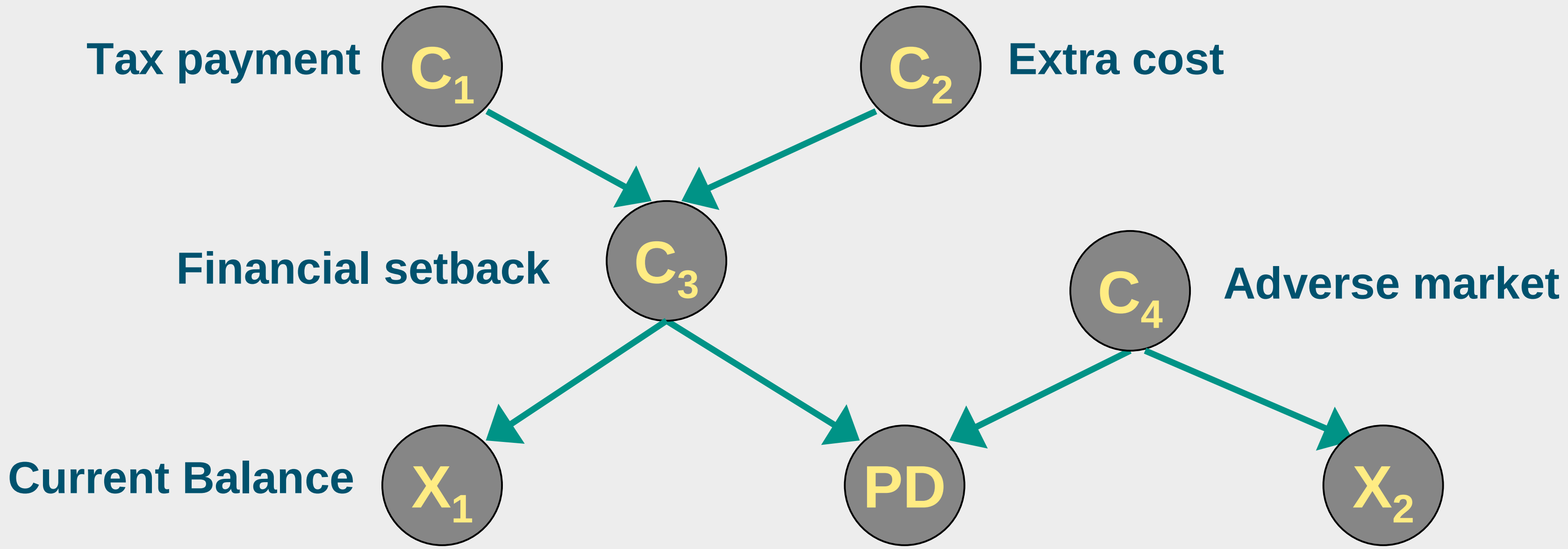


Conclusions & Recommendations

Improvements for BN results:

→ **Data selection process (stepwise BN)**

→ **High quality expert knowledge**



Conclusions & Recommendations

CPT approximation for Noisy-MAX learning

- **Noisy-MAX is very good alternative to CPT for rare event data**
- **CPT is probably not the best starting point for optimal parameters**

Conclusions & Recommendations

Prior / Posterior learning

- **Priors provide very stable networks**
- **High performance, promising technique**

→ **More testing on data**

TU Delft

Bayesian Networks in Credit Rating

Samuel Gerssen

March 12, 2004

Prof. Dr. H. Koppelaar
Drs. Dr. L.J.M. Rothkrantz
Ir. W. de Jong
Dr. K. van der Meer
Dr. D. Fokkema (ABN AMRO)
Dr. Ir. M.J. Druzdzel (University of Pittsburgh)

ABN·AMRO

University of Pittsburgh