CS 284 Programming Assignment 1

Credit-Worthiness Predictor using Bayesian Networks

Isabelle Tingzon*

Department of Computer Science

University of the Philippines-Diliman

02 October 2016

Contents

1 Introduction **2**

1.1 Bayesian Networks 2

1.2 Credit-Worthiness Predictor 2

1.3 SAMIAM Software 3

2 Implementation **3**

2.1 Variable Description 3

2.2 Network Construction 4

2.3 SAMIAM Network Screenshot 4

3 Test Cases **5**

4 Conclusion **10**

*ibtingzon@up.edu.ph

1 Introduction

1.1 Bayesian Networks

A Bayesian network (BN) is a probabilistic graphical model (PGM) that represents the joint probability distributions of a set of random variables and describes their conditional dependencies via directed acyclic graphs (DAGs), whose nodes denote random variables and whose edges correspond to the dependence (or influence) of one node on another [2].

More formally, a Bayesian network is defined as a pair $B = (G, P)$ where G is a BN graph, and P is a set of conditional probability distributions (CPD) associated with G's nodes, and P *factorizes* according to G, i.e. P can be expressed as a product:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i}^{G})$$

where $Pa_{X_i}^{G}$ represents the parent nodes of X_i in graph G.

1.2 Credit-Worthiness Predictor

The goal of this project is to develop a predictor for credit-worthiness based on a a set of observations made from an expert's experience in evaluating customers' credit-worthiness. The following can be observed about the customer: income, amount of assets, debts to income ratio, payment history, and age. Moreover, credit worthiness is ultimately dependent on debts to income ratio, future income and reliability.

The following observations are specified as follows:

1. The better a person's payment history, the more likely the person is to be reliable.
2. The older a person is, the more likely the person is to be reliable.
3. Older people are more likely to have an excellent payment history.
4. People who have a high ratio of debts to income are likely to be in financial hardship and hence less likely to have a good payment history.
5. The higher a person's income, the more likely it is for the person to have many assets.
6. The more assets a person has and the higher the person's income, the more likely the person is to have a promising future income.
7. Reliable people are more likely to be credit-worthy than unreliable people. People who have promising future incomes, or who have low ratios of debts to income, are more likely to be credit-worthy than people who do not.

1.3 SAMIAM Software

The Bayesian network constructed for the credit-worthiness predictor was implemented using the SAMIAM modelling tool developed by the Automated Reasoning Group of Professor Adnan Darwiche at UCLA [1].

2 Implementation

2.1 Variable Description

Based on the observable features described in Section 1.2, there are eight variables: income, assets, debt-to-income ratio, payment history, age, reliability, future income, and credit-worthiness. Table 2.1 describes the random variables used to denote each of the eight variables.

Variable	**Random Variable**	**Values**	**Description**
income	I	$Val(I) = \{i^0, i^1, i^2\}$	i^0 - low income i^1 - middle income i^2 - high income
assets	S	$Val(S) = \{s^0, s^1\}$	s^0 - few assets s^1 - many assets
debt-to-income ratio	D	$Val(D) = \{d^0, d^1\}$	d^0 - low DTI ratio d^1 - high DTI ratio
payment history	H	$Val(H) = \{h^0, h^1, h^2\}$	h^0 - unacceptable h^1 - good/acceptable h^2 - excellent
age	A	$Val(A) = \{a^0, a^1\}$	a^0 - age <30 y.o. (young) a^1 - age >30 y.o. (old)
reliability	R	$Val(R) = \{r^0, r^1\}$	r^0 - not reliable r^1 - reliable
future income	F	$Val(F) = \{f^0, f^1\}$	f^0 - bleak future income f^1 - promising future income
credit-worthiness	C	$Val(C) = \{c^0, c^1\}$	c^0 - not credit-worthy c^1 - credit-worthy

Table 1: A description of the variables and their corresponding random variables.

2.2 Network Construction

Based on the observations stated in Section 1.2, we can infer the following:

- From observations 1 and 2, reliability depends on payment history and age.
- From observation 3 and 4, payment history is dependent on age and debt-to-income ratio.
- From observation 5, assets depends solely on income.
- From observation 6, future income is dependent on assets and income.
- From observation 7, credit-worthiness is dependent on reliability, future income, and debt-to-income ratio.

Based on these observations, we construct the joint probability distribution as follows:

$$P(I, S, F, C, D, H, R, A) = P(I)P(D)P(A)P(A|I)P(F|A, I)P(H|D, A)P(R|H, A)P(C|F, D, R)$$

The complete Bayesian network is presented in Fig. 1. This network structure is most natural as its edges encode intuition about credit-worthiness based on the specified observations.

2.3 SAMIAM Network Screenshot

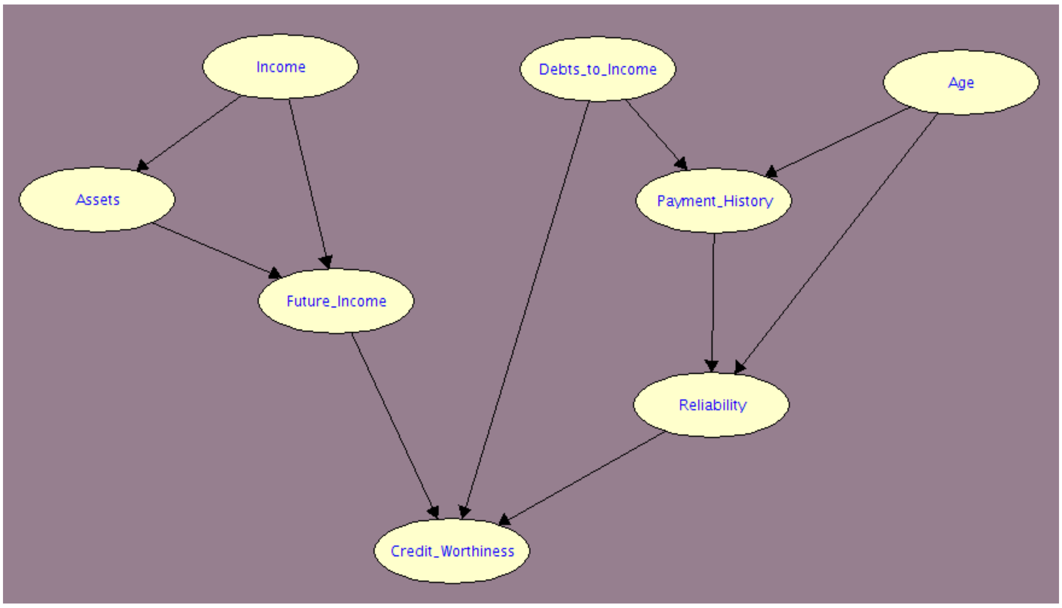

Figure 1: Bayesian network for credit-worthiness predictor.

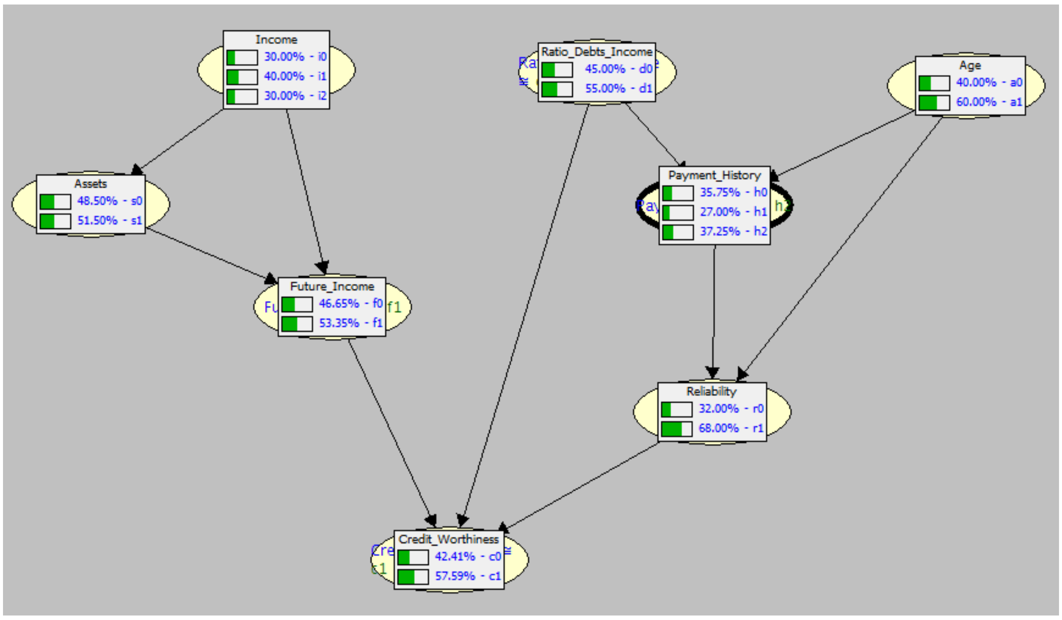

Figure 2: (Query mode) Conditional probability distributions.

3 Test Cases

In order to verify whether the network is consistent with the desired behavior, we check that it satisfies a set of conditions based on the observations.

- **Observation 1:** *The better a person's payment history, the more likely the person is to be reliable.* For the network to be consistent with observation 1, the following condition should be satisfied:

$$P(R = r^1 | H = h^2) > P(R = r^1 | H = h^1) > P(R = r^1 | H = h^0)$$

As shown in Fig. 3, we get

$$P(R = r^1 | H = h^2) = 0.8965 > P(R = r^1 | H = h^1) = 0.7566 > P(R = r^1 | H = h^0) = 0.3983$$

Thus, the condition is satisfied.

- **Observation 2:** *The older a person is, the more likely the person is to be reliable.* For the network to be consistent with observation 2, the following condition should be satisfied:

$$P(R = r^1 | A = a^1) > P(R = r^1 | A = a^0)$$

As shown in Fig. 4, we get

$$P(R = r^1 | A = a^1) = 0.8266 > P(R = r^1 | A = a^0) = 0.4595$$

Thus, the condition is satisfied.

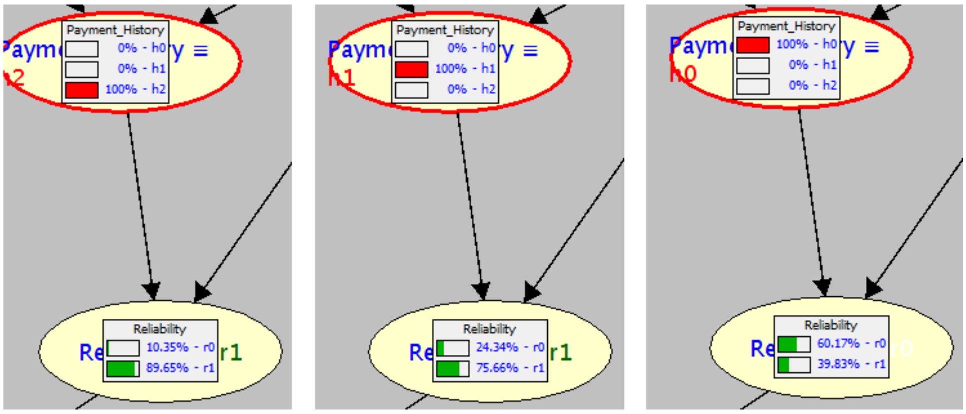

Figure 3: Conditioning R on $H = h^2$, $H = h^1$, and $H = h^0$

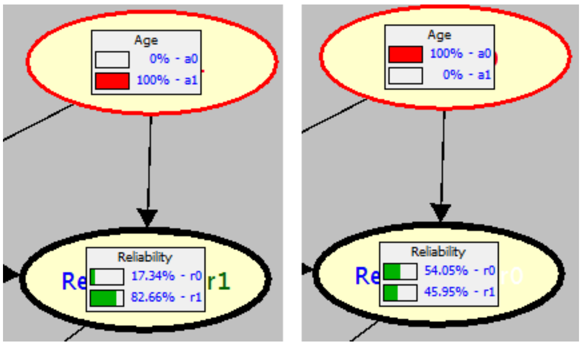

Figure 4: Conditioning R on $A = a^1$, $A = a^0$

- **Observation 3:** *Older people are more likely to have an excellent payment history.* For the network to be consistent with observation 3, the following condition should be satisfied:

$$P(H = h^2 | A = a^1) > P(H = h^2 | A = a^0)$$

As shown in Fig. 5, we get

$$P(H = h^2 | A = a^1) = 0.4525 > P(H = h^2 | A = a^0) = 0.1850$$

Thus, the condition is satisfied.

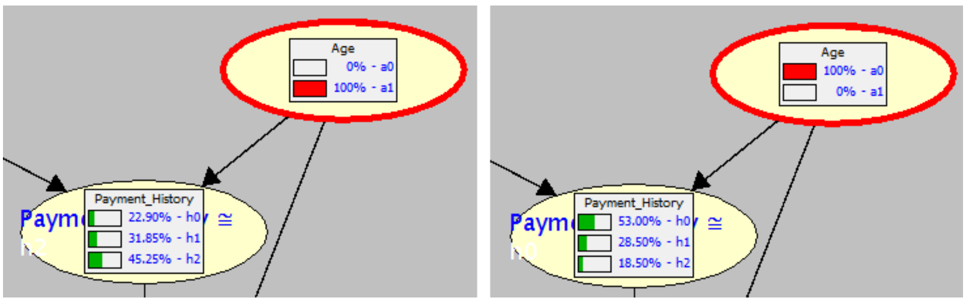

Figure 5: Conditioning H on $A = a^1$, $A = a^0$

- **Observation 4:** *People who have a high ratio of debts to income are likely to be in financial hardship and hence less likely to have a good payment history.* For the network to be consistent with observation 4, the following condition should be satisfied:

$$P(H = h^1|D = d^1) < P(H = h^1|D = d^0)$$

As shown in Fig. 6, we get

$$P(H = h^1|D = d^1) = 0.27 < P(H = h^1|D = d^0) = 0.348$$

Thus, the condition is satisfied.

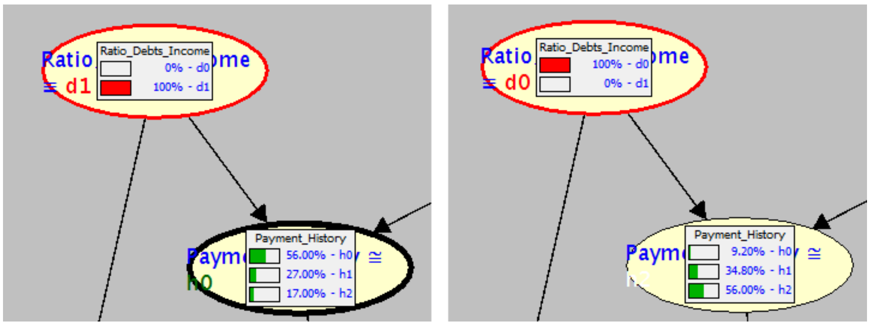

Figure 6: Conditioning H on $D = d^1$, $D = d^0$

- **Observation 5:** *The higher a person's income, the more likely it is for the person to have many assets.* For the network to be consistent with observation 5, the following condition should be satisfied:

$$P(S = s^1|I = i^2) > P(S = s^1|I = i^1) > P(S = s^1|I = i^0)$$

As shown in Fig. 7, we get,

$$P(S = s^1|I = i^2) = 0.8 > P(S = s^1|I = i^1) = 0.65 > P(S = s^1|I = i^0) = 0.05$$

Thus, the condition is satisfied.

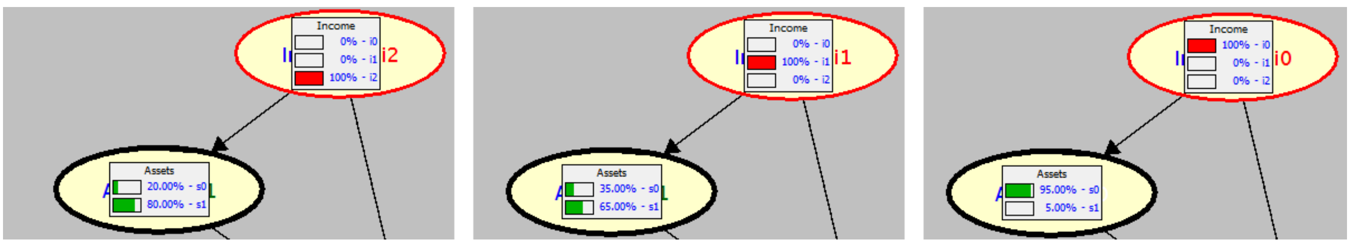

Figure 7: Conditioning S on $I = i^2$, $I = i^1$, $I = i^0$

- **Observation 6:** *The more assets a person has and the higher the person's income, the more likely the person is to have a promising future income.* For the network to be consistent with observation 6, the following conditions should be satisfied:

$$P(F = f^1|I = i^2, S = s^1) > P(F = f^1|I = i^1, S = s^1) > P(F = f^1|I = i^0, S = s^1)$$

$$P(F = f^1|I = i^2, S = s^0) > P(F = f^1|I = i^1, S = s^0) > P(F = f^1|I = i^0, S = s^0)$$

$$P(F = f^1|I = i^2, S = s^1) > P(F = f^1|I = i^2, S = s^0)$$

$$P(F = f^1|I = i^1, S = s^1) > P(F = f^1|I = i^1, S = s^0)$$

$$P(F = f^1|I = i^0, S = s^1) > P(F = f^1|I = i^0, S = s^0)$$

As shown in Fig. 8, we get,

$$P(F = f^1|I = i^2, S = s^1) = 0.9 > P(F = f^1|I = i^1, S = s^1) = 0.65 > P(F = f^1|I = i^0, S = s^1) = 0.55$$

$$P(F = f^1|I = i^2, S = s^0) = 0.7 > P(F = f^1|I = i^1, S = s^0) = 0.6 > P(F = f^1|I = i^0, S = s^0) = 0.05$$

$$P(F = f^1|I = i^2, S = s^1) = 0.9 > P(F = f^1|I = i^2, S = s^0) = 0.7$$

$$P(F = f^1|I = i^1, S = s^1) = 0.65 > P(F = f^1|I = i^1, S = s^0) = 0.6$$

$$P(F = f^1|I = i^0, S = s^1) = 0.55 > P(F = f^1|I = i^0, S = s^0) = 0.05$$

Thus, the condition is satisfied.

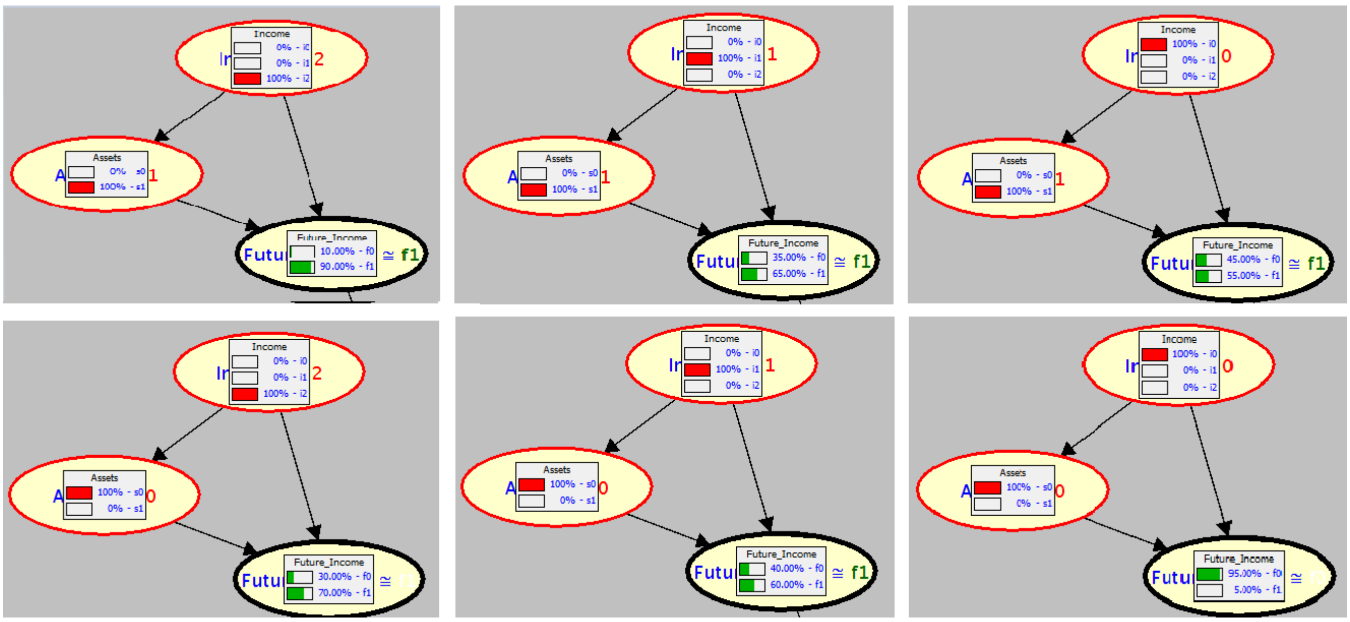

Figure 8: Conditioning F on $I = i^2$, $I = i^1$, $I = i^0$, $S = s^0$, $S = s^1$

- **Observation 7:** *Reliable people are more likely to be credit-worthy than unreliable people. People who have promising future incomes, or who have low ratios of debts to income, are more likely to be credit-worthy than people who do not.* For the network to be consistent with observation 7, the following conditions should be satisfied:

$$P(C = c^1|R = r^1) > P(C = c^1|R = r^0)$$

$$P(C = c^1|F = f^1) > P(C = c^1|F = f^0)$$

$$P(C = c^1|D = d^0) > P(C = c^1|D = d^1)$$

As shown in Fig. 9, we get,

$$P(C = c^1|R = r^1) = 0.6733 > P(C = c^1|R = r^0) = 0.3690$$

As shown in Fig. 10, we get,

$$P(C = c^1|F = f^1) = 0.6925 > P(C = c^1|F = f^0) = 0.4424$$

As shown in Fig. 11, we get,

$$P(C = c^1|D = d^0) = 0.7413 > P(C = c^1|D = d^1) = 0.4405$$

Thus, the condition is satisfied.

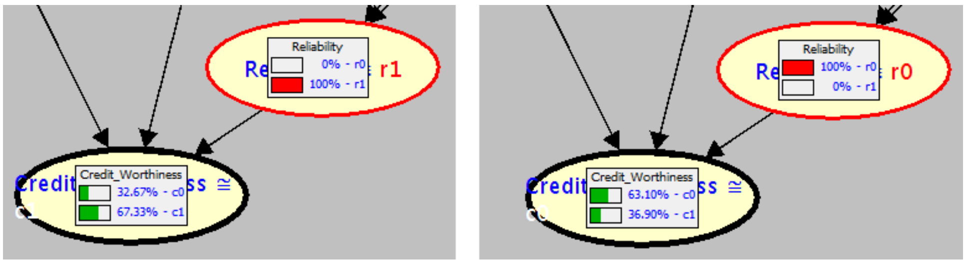

Figure 9: Conditioning C on $R = r^1$, $R = r^0$

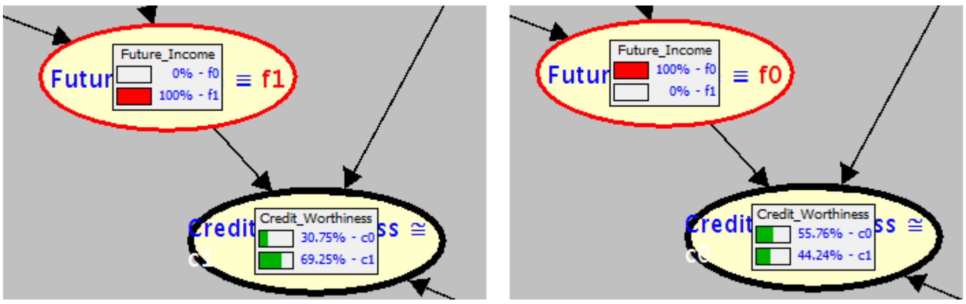

Figure 10: Conditioning C on $F = f^1$, $F = f^0$

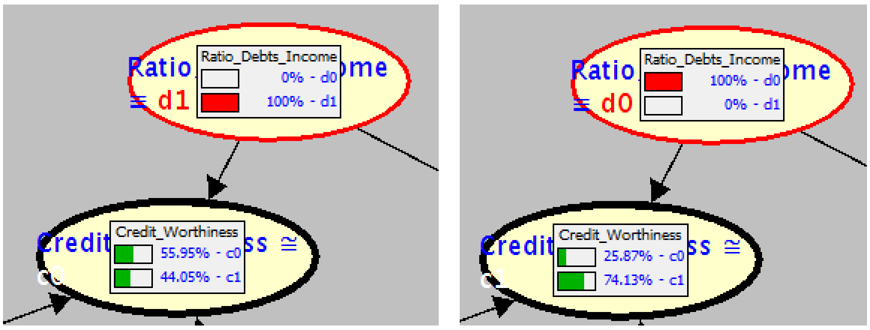

Figure 11: Conditioning C on $D = d^1$, $D = d^0$

4 Conclusion

This work demonstrates the use of Bayesian networks as a graphical modelling tool for predicting customer credit-worthiness. The BN was constructed using the SAMIAM software and evaluated based on the consistency of the produced marginals with the actual observations. The results show that the constructed Bayesian network was successfully able to satisfy all conditions and is therefore an effective tools for solving the customer credit-worthiness prediction problem.

Thus, we have successfully demonstrated in this project the ability of Bayesian networks to mirror intuition and real world observations. In particular, BN's are an effective tool for solving prediction problems that involve having expert knowledge on hand. In general, Bayesian networks can be effectively applied to a range of applications such as medical diagnosis, threat modelling, and decision support systems.

References

[1] Keith Cascio et al. SAMIAM. http://reasoning.cs.ucla.edu/samiam. Accessed 02-10-2016.

[2] Prospero Naval. Bayesian Networks (CS 284 Lecture 2). University Lecture, 2016.