On the errors introduced by the naive Bayes independence assumption

Author
Matthijs de Wachter
3671100

Utrecht University

Master Thesis
Artificial Intelligence

Supervisor
Dr. Silja Renooij
Department of Information and Computing Sciences
Utrecht University

The Netherlands
June 1, 2018

Contents

Abstract This research seeks to shed light on how well a naive Bayesian approach functions when we cannot be sure of independent evidence. This research will show just how large the error in likelihood and posterior probability of a class variable given certain evidence can possibly get when using a naive Bayesian approach. I will prove that complete dependency among the evidence variables is the worst case scenario for the error in likelihood when we have 2 or 3 pieces of evidence. Based on these results, this research introduces an equation to calculate the maximum error in likelihood under complete dependency, based on the number of observed evidence variables. I will also show that there is no real bound on the error in posterior, except that it cannot become equal to one, where the worst case scenario is when the class variable deterministically follows from the observed evidence when dependencies among evidence is considered. This research will present some experimental results on how large the error in both likelihood and posterior will typically be, and how this error correlates with various dependency measures. These experimental results support my claim that complete dependency is the worst case scenario for the error in likelihood, and determinism is the worst case scenario for the error in posterior.

1. Introduction

The field of Artificial Intelligence aims to automate or assist people with tasks that require some form of intelligence. One of these tasks is reasoning or decision making under uncertainty. For these tasks we often use models of probability distributions, like Bayesian networks [1]. Implementations of these models typically take a set of observations, or evidence, as an input and use these to calculate the probabilities of certain variables of interest. This process is called probabilistic inference.

At the heart of probablistic inference is Bayes' rule [2], which is used to calculate the probabilities of some outcome based on some amount of evidence found supporting or opposing this outcome. Bayes' rule is given by the following formula

$$\Pr(w|\boldsymbol{x}) = \frac{\Pr(\boldsymbol{x}|w) \cdot \Pr(w)}{\Pr(\boldsymbol{x})} \tag{1}$$

Here w is a specific value for the variable W we are interested in, which we will call the class variable, and $\boldsymbol{x}$ is the observed evidence, which is a specific combination of values for a set of evidence variables $\boldsymbol{X}$. $\Pr(w|\boldsymbol{x})$ is the probability for the class variable to have the value w given the set of evidence $\boldsymbol{x}$, which is called the *posterior* probability. $\Pr(\boldsymbol{x}|w)$ is the probability for the set of evidence $\boldsymbol{x}$ to be true, given that the class has a value of w, and is called the *likelihood*. $\Pr(w)$ and $\Pr(\boldsymbol{x})$ are the probabilities of w and $\boldsymbol{x}$ being true regardless of any other values. These are called the class *prior* and *marginal* probability of $\boldsymbol{x}$, respectively. Throughout this thesis I will often refer to likelihood and posterior, which refers to the concepts described here.

$\boldsymbol{X}$ consists of multiple evidence variables X, for which the specific values are denoted as x, all of which can be dependent on each other. However, in practice, often only the dependencies between the class variable and each evidence variable are considered. The approach in which all pieces of evidence are considered independent given the class variable is known as the *naive Bayesian* approach, or just naive Bayes. This independency assumption does not always hold, however, and can introduce errors upon inference.

This naive Bayesian method is often used in the context of classification, which is classifying which of the values w of class variable W has the highest posterior probability given a certain set of observed evidence x. The error in classification that arises from wrongly assuming independence has been studied to some extent. The classifier is concerned only with which of the various values for W has the highest posterior probability. This means the classifier could be correct, even though the probabilities used to determine which value is most likely may be wrong. Hand and Yu [3] show many examples of cases where the evidence is clearly not independent, but the naive Bayesian classifier still has surprisingly good results. Rish et al [4] show that the naive Bayesian classifier performs optimally in the case of complete dependency among evidence variables and equiprobable priors. Here, a classifier performs optimally when the value selected by the classifier is the same value that would have been selected by a classifier that does not assume independence. Domingos and Pazzani [5] show some necessary and some sufficient conditions under which the naive Bayesian classifier performs optimally, where optimal has the same meaning as before. Kuncheva and Hoare [6] propose the Q statistic as a measure of dependency to account for the difference in dependency distributions between the classes. Although the degree of dependency among evidence and the accuracy of Naive Bayes are not directly correlated, the Q statistic does seem to correlate with the accuracy of the naive Bayesian classifier.

In some cases, however, we would like to be more sure about a certain value than just knowing it is the most likely one. For example, in healthcare, where the probability of a certain ailment influences whether the patient needs surgery or not. In these cases we would like to calculate the actual likelihoods and posteriors. There seems to have been little research about the error introduced in likelihood and posterior probability by assuming independence. Renooij [7] seems to be the first to consider likelihood and likelihood ratios, instead of just the classifier output. Renooij established a theoretical bound on the error in overall likelihood of 0.25 when 2 pieces of evidence are dependent and an empirical bound of 0.30 with more than 2 pieces of dependent evidence. This research only concerned cases with at most 4 pieces of evidence.

In this research I further examine the relation between the error in naive Bayes and dependency. I attempt to determine just how high this error can and typically will get, for both likelihood and posterior. The error I consider in this thesis is simply the absolute difference between the likelihood and posterior probability obtained from using a naive Bayesian method and the likelihood and posterior probability when we do consider dependencies among evidence variables.

Section 2 will introduce some terminology used in this thesis. In Section 3 I will prove for cases with two pieces of evidence that complete dependency is the worst case scenario for the error in likelihood. Assuming that this holds true for a greater number of evidence variables as well, I introduce an equation to calculate the maximum error in likelihood for any number of evidence variables. Section 4 shows that the error in posterior can be any value up to, but not including, 1, and that the largest errors are achieved when the class value deterministically follows from the evidence. Section 5 describes a pilot experiment, which was set up to empirically determine the maximum error in likelihood, as well as the relation between this error and dependency. Section 6 describes the main experiment, which is an extension of the pilot experiment and most notably includes the error in posterior as well. Section 7 concludes this thesis.

2. Preliminaries

In this section, some terminology will be established, which will be used throughout this thesis. This thesis is concerned with variables that have a probability distribution over a set of discreet variables, for this we will use standard probability theory rules [2].

The variable of interest, for which we want to calculate the likelihood and posterior probabilities, is referred to as the *class* variable. This class variable is dependent on a certain number of other variables referred to as the *evidence* variables.

The evidence variables will be denoted as X_i, for which the possible values are x_i for the specific value of interest and $\neg x_i$ for all other values. The class variable will be denoted as W, for which the values are again w for a specific value and $\neg w$ for any other value. This means we use a binary representation, but the theoretical results are also true for non-binary cases. A full set of evidence variables is denoted as $\boldsymbol{X}$ and a specific set of values as $\boldsymbol{x}$. $\boldsymbol{X}$ is also used to denote the set of every possible set of evidence values, and X and W are also used for the set of every possible value of X and W, respectively. Which of these usages is intended will be clear from context. The posterior probability of any value of the class variable can be calculated by using its prior probability and the conditional probabilities from the evidence variables, using Bayes' rule, which can be found in Equation 1. The number of evidence variables in a network is denoted by n.

The errors used in this thesis are simply the absolute differences between the naive approach to compute likelihood and posterior probability and the likelihood and posterior probabilities when dependencies among evidence variable are considered, which we will call the exact approach. So, for likelihood and posterior the equations for error are as follows.

$$Err(\boldsymbol{x}|w) = |\Pr_{exact}(\boldsymbol{x}|w) - \Pr_{naive}(\boldsymbol{x}|w)| \tag{2}$$

$$Err(w|\boldsymbol{x}) = |\Pr_{exact}(w|\boldsymbol{x}) - \Pr_{naive}(w|\boldsymbol{x})| \tag{3}$$

Since we want to find out how well a naive Bayesian approach to computing likelihoods and posterior probabilities works when its underlying independence assumption does not hold, we would like to relate the error due to using a naive approach to dependency. Dependency is difficult to quantify however. But we can define *complete dependency*, which is the case where any single piece of evidence determines the value of all other evidence variables.

Definition 1. *Let $\boldsymbol{X}$ be a set of variables. Now, $\boldsymbol{X}$ is* completely dependent *when for all combinations of evidence $\boldsymbol{x}$ the following holds:*

$$\forall \boldsymbol{x} \in \boldsymbol{X} \quad (\Pr(\boldsymbol{x}) = 0 \quad \vee \quad \forall x, y \in \boldsymbol{x} \Pr(y|x) = 1) \tag{4}$$

From this definition, the following property follows.

Lemma 2.1. *Let $\boldsymbol{X}$ be a completely dependent set of variables. Now, the following equations holds*

$$\forall \boldsymbol{x} \in \boldsymbol{X} \quad (\Pr(\boldsymbol{x}) = 0 \quad \vee \quad \forall x, y \in \boldsymbol{x} \Pr(x) = \Pr(y)) \tag{5}$$

Proof. Let $\boldsymbol{X}$ be as before and let x and y be random values from $\boldsymbol{X}$. Now from the definition of complete dependency follows that either $\Pr(x, y) = 0$ or that both $\Pr(x|y) = 1$ and $\Pr(y|x) = 1$. If $\Pr(x, y) = 0$, it follows directly from Definition 1 that Lemma 2.1 holds. When $\Pr(x, y) \neq 0$, we have that $\Pr(x|y) = 1$ and $\Pr(y|x) = 1$. Using Bayes' rule we have that

$$\Pr(x|y) \cdot \Pr(y) = \Pr(y|x) \cdot \Pr(x) \tag{6}$$

And since both $\Pr(x|y)$ and $\Pr(y|x)$ equal 1, we have that

$$\Pr(y) = \Pr(x) \tag{7}$$

□

Another definition related to dependency is *determinism.* A class variable can be considered deterministic when every complete combination of evidence results in a 0 or 1 probability for either value for the class variable. More generally, this can be defined as follows.

Definition 2. *Let $\boldsymbol{X}$ be a set of variables and let W be a variable not included in $\boldsymbol{X}$. Now, W* deterministically *depends on $\boldsymbol{X}$ when for all combinations of evidence $\boldsymbol{x}$ the following holds:*

$$\forall \boldsymbol{x} \in \boldsymbol{X}, w \in W \quad \Pr(w|\boldsymbol{x}) = 0 \quad \vee \quad \Pr(w|\boldsymbol{x}) = 1 \tag{8}$$

When the probability of a specific value w of class variable W is either 0 or 1 given only a specific set of evidence $\boldsymbol{x}$, we speak of context specific determinism.

Definition 3. *Let $\boldsymbol{x}$ be a set of values for a set of variables $\boldsymbol{X}$ and let w be an arbitrary value for variable W. Now, w is* deterministic *in the context of $\boldsymbol{x}$ when* $\Pr(w|\boldsymbol{x})$ *is either 0 or 1.*

From these definitions it follows that every possible value w for a variable W that is deterministically dependent on a set of variables $\boldsymbol{X}$, is deterministic in the context of any set of values $x \in \boldsymbol{X}$.

Apart from Bayes' rule, which was already mentioned in Section 1, I will use the probabilistic chain rule [8]. The probabilistic chain rule follows from the repeated application of the definition for conditional probability, which states that a joint probability can be calculated by multiplying the prior probability of one of the values by the conditional probability of the other value given the first value:

$$\Pr(x_1, x_2, ..., x_n) = \Pr(x_1|x_2, ..., x_n) \cdot \Pr(x_2, ..., x_n) \tag{9}$$

Repeatedly applying this definition makes sure we no longer have any joint probabilities, but only conditional probabilities remain.

$$\Pr(x_1, x_2, x_3, ..., x_n) = \Pr(x_1|x_2, x_3, ..., x_n) \cdot \Pr(x_2|x_3, ..., x_n) \cdot ... \cdot \Pr(x_n) \tag{10}$$

3. Worst case scenario for likelihood

When dependencies among evidence variables are ignored while calculating certain probabilities, it would make sense that the error in these calculations, as defined in Equation 2, would be largest when the variables are *completely dependent* as defined in Section 2. In this section I prove the largest error in likelihood $\Pr(\boldsymbol{x}|w)$ is indeed achieved under complete dependency when we consider two or three evidence variables. Furthermore, I introduce an equation to compute the maximum error in likelihood for any number of variables under complete dependency. To achieve this, we will first consider the formulas for the likelihood for both the exact and naive Bayesian approaches. Now using the probabilistic chain rule we have that

$$\Pr_{exact}(\boldsymbol{x}|w) = \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot ... \cdot \Pr(x_n|x_1, x_2, ..., x_{n-1}, w) \tag{11}$$

And since the naive Bayesian approach assumes independency among evidence variables, we have that every conditional probability $\Pr(x_i|x_1, x_2, ..., x_{i-1}, w)$ is equal to $\Pr(x_i|w)$. So for the naive Bayesian apporach the likelihood is

$$\Pr_{naive}(\boldsymbol{x}|w) = \prod_{i=1}^{n} \Pr(x_i|w) \tag{12}$$

3.1. Two evidence variables

When we consider only two evidence variables, Equations 11 and 12 become

$$\Pr_{exact}(x_1, x_2|w) = \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \tag{13}$$

$$\Pr_{naive}(x_1, x_2|w) = \Pr(x_1|w) \cdot \Pr(x_2|w) \tag{14}$$

Now, since we are interested in the error, as defined in Equation 2, we want to know the difference between these two.

$$\begin{aligned} Err(\boldsymbol{x}|w) = |\Pr_{exact}(\boldsymbol{x}|w) - \Pr_{naive}(\boldsymbol{x}|w)| \\ = |\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) - \Pr(x_1|w) \cdot \Pr(x_2|w)| \\ = |\Pr(x_1|w) \cdot (\Pr(x_2|x_1, w) - \Pr(x_2|w))| \end{aligned} \tag{15}$$

Since $\Pr(x_1|w)$ can not be negative, this can be written as

$$Err(\boldsymbol{x}|w) = \Pr(x_1|w) \cdot |\Pr(x_2|x_1, w) - \Pr(x_2|w)| \tag{16}$$

Since we want to know under what circumstances the error is largest, we want to maximize this function. These probabilities cannot take just any value between zero and one, since there is a dependency between them given by Bayes' rule. So we would need certain constraints on these values. The first constraint can be derived from Bayes' rule:

$$\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) = \Pr(x_2|w) \cdot \Pr(x_1|x_2, w) \tag{17}$$

Since we know that $\Pr(x_1|x_2, w)$ must have a probability between 0 and 1, we know that $\Pr(x_1|w) \cdot \Pr(x_2|x_1, w)$ has to be either smaller than $\Pr(x_2|w)$ (in the case that $\Pr(x_1|x_2, w)$ is not equal to 1), or equal to $\Pr(x_2|w)$ (in the case that $\Pr(x_1|x_2, w)$ is equal to 1). This leads us to our first constraint:

$$\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \leq \Pr(x_2|w) \tag{18}$$

Similarly, we get the following constraint, by observing that $\Pr(\neg x_2|x_1, w) = 1 - \Pr(x_2|x_1, w)$ and $\Pr(\neg x_2|w) = 1 - \Pr(x_2|w)$.

$$\Pr(x_1|w) \cdot (1 - \Pr(x_2|x_1, w)) \leq 1 - \Pr(x_2|w) \tag{19}$$

Theorem 3.1. *The error from Equation 16 under the constraints from Equations 18 and 19 is maximized when the evidence variables X_1 and X_2 are completely dependent as defined in Definition 1, when conditioned on w.*

Lemma 3.2. *Let x_1 and x_2 be random values for evidence variables X_1 and X_2, and let w be a random value for class variable W. Consider the following optimization problem.*

$$\begin{aligned} \underset{a,b,c}{\text{maximize}} \quad & f(a, b, c) = a \cdot |b - c| \\ \text{subject to} \quad & a \cdot b \leq c, \\ & a \cdot (1 - b) \leq 1 - c \end{aligned} \tag{20}$$

Where $a = \Pr(x_1|w)$, $b = \Pr(x_2|x_1, w)$ and $c = \Pr(x_2|w)$. The following are solutions to this problem: $a = 0.5$, $b = 0$, $c = 0.5$ and $a = 0.5$, $b = 1$, $c = 0.5$.

To prove this I first consider three distinct cases: the case where b is equal to c, the case where b is less than c and the case where b is greater than c.

Lemma 3.3. *Let f, a, b and c be as before and let $b = c$. Now, f will always be equal to* 0.

Lemma 3.4. *Let f, a, b and c be as before, under the constraints from Lemma 3.2, and let $b < c$. Now, f will be maximized when $b = 0$ and $a = c = 0.5$.*

Proof. Since $0 \leq a \leq 1$ and $b < c$, $a \cdot b$ can never be greater than c, the first constraint in Lemma 3.2 always holds. To prove f is maximized when $b = 0$ and $a = c = 0.5$, we consider separate cases.

The greatest result for f would be achieved when the difference between b and c is greatest. This would be the case when $b = 0$ and $c = 1$. However, when we consider the second constraint this would mean $a \cdot (1 - 0) \leq 1 - 1$, which would mean $a = 0$ and in turn $f = 0$.

When $0 < b < 1$ and $c = 1$ the second constraint becomes $a \cdot (1 - b) \leq 1 - 1$, which either means that $a = 0$ or $(1 - b) = 0$, so $b = 1$. Both of these would result in $f = 0$.

When $b = 0$ and $0 < c < 1$, we get that $a \cdot (1-0) \leq 1-c$ and $f = a \cdot |0-c|$, so $a \leq 1-c$ and $f = a \cdot c$. Multiplying both sides of this constraint by c gives us $a \cdot c \leq c \cdot (1 - c)$. This means the maximum of f can not be higher than the maximum of $c \cdot (1-c)$. So, to maximize f, a needs to be equal to $1-c$ and we need to maximize $c \cdot (1-c)$. To maximize $c \cdot (1-c)$ we first take the derivative, which is $1-2c$. The maximum of $c \cdot (1-c)$ is found when the derivative equals to 0, which is the case when $c = 0.5$. Since $a = 1-c$ we have that $a = 1 - 0.5 = 0.5$. So, maximizing $f = a \cdot c$ under the constraint of $a \leq 1 - c$ gives us $a = c = 0.5$, so $f = 0.5 \cdot 0.5 = 0.25$.

The last possibility is when both $0 < b < 1$ and $0 < c < 1$. Now, for every value of a, b and c that do not violate the second constraint, we can lower the value of b by δ and raise the value of c by $\frac{\delta}{a}$, without violating the second constraint. Since a always has a value between 0 and 1, $\frac{\delta}{a}$ will always be larger than δ. This means that lowering b and raising c will lead to a larger value for f. Which means that the value of f can not be higher in this case than in the case where $b = 0$ and $0 < c < 1$.

□

Lemma 3.5. *Let f, a, b and c be as before, under the constraints from Theorem 3.2, and let $b > c$. Now, f will be maximized when $b = 1$ and $a = c = 0.5$.*

Proof. Since $0 \leq a \leq 1$ and $b > c$, $a \cdot (1 - b)$ can never be greater than $1 - c$, the second constraint in Theorem 3.2 always holds. To prove f is maximized when $b = 1$ and $a = c = 0.5$, we, again, consider separate cases.

Again, the greatest result for f would be achieved when the difference between b and c is greatest. Now, this would be the case when $b = 1$ and $c = 0$. However, when we consider the first constraint this would mean $a \cdot 1 \leq 0$, which would mean $a = 0$ and in turn $f = 0$.

When $0 < b < 1$ and $c = 0$ the second constraint becomes $a \cdot b \leq 0$, which either means that $a = 0$ or $b = 0$. Both of these would result in $f = 0$.

When $b = 1$ and $0 < c < 1$, we get that $a \cdot 1 \leq c$ and $f = a \cdot |1 - c|$, so $a \leq c$ and $f = a \cdot (1-c)$. Multiplying both sides of this constraint by $1-c$ gives us $a \cdot (1-c) \leq c \cdot (1-c)$. Analogously to the third possibility in Lemma 3.4, maximizing f gives us $a = c = 0.5$, so $f = 0.5 \cdot 0.5 = 0.25$.

The last possibility is when both $0 < b < 1$ and $0 < c < 1$. Now, for every value of a, b and c that do not violate the first constraint, we can raise the value of b by δ and lower the value of c by $\frac{\delta}{a}$, without violating the first constraint. Since a always has a value between 0 and 1, $\frac{\delta}{a}$ will always be larger than δ. This means that raising b and lowering c will lead to a larger value for f. Which means that the value of f can not be higher in this case than in the case where $b = 1$ and $0 < c < 1$. □

Proof of Lemma 3.2. The maximum error found in both Lemmas 3.4 and 3.5 is equal to 0.25, so both cases are solutions to this problem. So f is maximized when $\Pr(x_2|x_1, w)$ is equal to either 0 or 1 and both $\Pr(x_1|w)$ and $\Pr(x_2|w)$ are equal to 0.5. □

Proof of Theorem 3.1. The largest errors for likelihood are achieved when $\Pr(x_2|x_1, w)$ is equal to either 0 or 1 and both $\Pr(x_1|w)$ and $\Pr(x_2|w)$ are equal to 0.5. To prove X_1 and X_2 are completely dependent given w in both cases, we consider every possible combination of evidence for both cases.

When $\Pr(x_2|x_1, w) = 0$ and $\Pr(x_1|w) = \Pr(x_2|w) = 0.5$, we can use the equations from Appendix A to obtain the following joint probabilities: $\Pr(x_1, x_2|w) = 0$, $\Pr(\neg x_1, \neg x_2|w) = 0$, $\Pr(\neg x_1, x_2|w) = 0.5$ and $\Pr(x_1, \neg x_2|w) = 0.5$. This proves that the definition for complete dependency holds for $\boldsymbol{x} = \{x_1, x_2\}$ and $\boldsymbol{x} = \{\neg x_1, \neg x_2\}$. From Appendix A we also get that $\Pr(x_1|\neg x_2, w) = 1$, $\Pr(\neg x_2|x_1, w) = 1$, $\Pr(x_2|\neg x_1, w) = 1$ and $\Pr(\neg x_1|x_2, w) = 1$. Which means the definition holds for $\boldsymbol{x} = \{x_1, \neg x_2\}$ and $\boldsymbol{x} = \{\neg x_1, x_2\}$ as well.

Similarly, when $\Pr(x_2|x_1, w) = 1$ and $\Pr(x_1|w) = \Pr(x_2|w) = 0.5$, we get the following joint probabilities: $\Pr(x_1, x_2|w) = 0.5$, $\Pr(\neg x_1, \neg x_2|w) = 0.5$, $\Pr(\neg x_1, x_2|w) = 0$ and $\Pr(x_1, \neg x_2|w) = 0$. This proves that the definition for complete dependency holds for $\boldsymbol{x} = \{\neg x_1, x_2\}$ and $\boldsymbol{x} = \{x_1, \neg x_2\}$. We also have that $\Pr(x_1|x_2, w) = 1$, $\Pr(x_2|x_1, w) = 1$, $\Pr(\neg x_2|\neg x_1, w) = 1$ and $\Pr(\neg x_1|\neg x_2, w) = 1$. Which means the definition holds for $\boldsymbol{x} = \{x_1, x_2\}$ and $\boldsymbol{x} = \{\neg x_1, \neg x_2\}$ as well. □

This proves that for cases with 2 pieces of evidence, the maximum error in likelihood can only be reached under complete dependency. However, complete dependency is no guarantee for a large error, for example when $\Pr(x_2|x_1, w) = \Pr(x_1|w) = \Pr(x_2|w) = 1$ we still have complete dependency, while the error is 0.

3.2. Three evidence variables

When we consider three pieces of evidence, we get

$$\Pr_{exact}(\boldsymbol{x}|w) = \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot \Pr(x_3|x_1, x_2, w) \tag{21}$$

$$\Pr_{naive}(\boldsymbol{x}|w) = \Pr(x_1|w) \cdot \Pr(x_2|w) \cdot \Pr(x_3|w) \tag{22}$$

Now, the error will be, analogously to Equation 16

$$Err_{likelihood} = \Pr(x_1|w) \cdot |\Pr(x_2|x_1, w) \cdot \Pr(x_3|x_1, x_2, w) - \Pr(x_2|w) \cdot \Pr(x_3|w)| \tag{23}$$

In addition to the constraints we had in 18 and 19, we also get two additional constraints, analogously to the constraints from equation 18 and 19. These are

$$\Pr(x_1, x_2|w) \cdot \Pr(x_3|x_1, x_2, w) \leq \Pr(x_3|w) \tag{24}$$

$$\Pr(x_1, x_2|w) \cdot (1 - \Pr(x_3|x_1, x_2, w)) \leq 1 - \Pr(x_3|w) \tag{25}$$

Which can be rewritten as

$$\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot \Pr(x_3|x_1, x_2, w) \leq \Pr(x_3|w) \tag{26}$$

$$\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot (1 - \Pr(x_3|x_1, x_2, w)) \leq 1 - \Pr(x_3|w) \tag{27}$$

Theorem 3.6. *The error from Equation 23 under the constraints from Equations 18, 19, 26 and 27 is maximized when the evidence variables X_1, X_2 and X_3 are completely dependent as defined in Definition 1, when conditioned on w.*

Lemma 3.7. *Let x_1, x_2 and x_3 be random values for evidence variables X_1, X_2 and X_3, and let w be a random value for class variable W. Consider the following optimization problem.*

$$\begin{aligned}
\underset{a,b,c,d,e}{\text{maximize}} \quad & f(a,b,c,d,e) = a \cdot |b \cdot d - c \cdot e| \\
\text{subject to} \quad & a \cdot b \leq c, \\
& a \cdot (1 - b) \leq 1 - c, \\
& a \cdot b \cdot d \leq e, \\
& a \cdot b \cdot (1 - d) \leq (1 - e)
\end{aligned} \tag{28}$$

Where $a = \Pr(x_1|w)$, $b = \Pr(x_2|x_1, w)$, $c = \Pr(x_2|w)$, $d = \Pr(x_3|x_1, x_2, w)$ and $e = \Pr(x_3|w)$. The following is a solution to this problem: $a = 0.57735$, $b = 1$, $c = 0.57735$, $d = 1$, $e = 0.57735$.

This is too complex to solve algebraically. However, numerically solving this, using the FindMaximum function from Mathematica, gives us that $\Pr(x_1|w) = \Pr(x_2|w) = \Pr(x_3|w) = 0.57735$ and $\Pr(x_2|x_1, w)$ and $\Pr(x_3|x_1, x_2, w)$ both equal 1. The error itself will be 0.3849 in this case. This represents another case of complete dependency among evidence variables, since all conditional probabilities conditioned on more than just w equal 1, so when a single evidence value is x_i, all evidence values are x_i.

3.3. Theoretical bound on the error in likelihood

Given the results of this section, it seems likely that complete dependency among evidence variables is the worst case scenario for the error in likelihood regardless of the number of evidence variables. This means it is worth knowing what the maximum error for likelihood is for n evidence variables under complete dependency.

Theorem 3.8. *Consider a combination of values $\boldsymbol{x}$ for a set of n evidence variables $\boldsymbol{X}$, and a class variable W with value w. Given that we have complete dependency among evidence variables as defined in Definition 1, the maximum error in likelihood as defined in Equation 2 is given by $MaxErr(n) = n^{\frac{-1}{n-1}} - n^{\frac{-n}{n-1}}$.*

Proof. Assume that the probability of a single evidence variable given w equals y, that is

$$\Pr(x_i|w) = y \tag{29}$$

As defined in Definition 1, we have that every conditional probability conditioned on x_i is equal to 1. This means that the likelihood for the complete set evidence should be y as well, so

$$\Pr_{exact}(x_1, ..., x_n|w) = y \tag{30}$$

However, since the naive approach ignores dependencies and every piece of evidence has the same probability given class w as was proven in Lemma 2.1, we have that

$$\Pr_{naive}(x_1, ..., x_n|w) = y^n \tag{31}$$

This means that the error in likelihood is given by

$$Err(x_1, ..., x_n|w) = y - y^n \tag{32}$$

To establish the maximum error, we will take the derivative of the error and compute the value of y when this derivative equals zero.

$$Err'(\boldsymbol{x}|w) = 1 - ny^{n-1} = 0 \tag{33}$$

Solving this for y gives us:

$$y = n^{\frac{-1}{n-1}} \tag{34}$$

Now we replace y in Equation 32 with the result we obtained in Equation 34 to get a function for the theoretical bound on likelihood with completely dependent pieces of evidence.

$$MaxErr(\boldsymbol{x}|w) = n^{\frac{-1}{n-1}} - n^{\frac{-n}{n-1}} \tag{35}$$

□

Table 1 shows the maximum error for several numbers of observed evidence variables and the corresponding value of $\Pr(x_i|w)$, or y. For $n = 2$ and $n = 3$ these values are indeed equal to the values found earlier in this section. The limit of Equation 35 to infinity approaches 1, which means that when we have an infinite number of completely dependent pieces of evidence, the naive Bayesian approach could have an error of 1 in likelihood. Figure 1 shows the rate at which the error in likelihood increases as more pieces of observed evidence are added.

Table 1: Maximum error in likelihood under complete dependency, together with the $\Pr(x_i|w)$ that results in the maximum error.

n	2	3	4	5	10
$\Pr(x_i\|w)$	0.5	0.57735	0.629961	0.66874	0.774264
$MaxErr(\boldsymbol{x}\|w)$	0.25	0.3849	0.47247	0.534992	0.696837

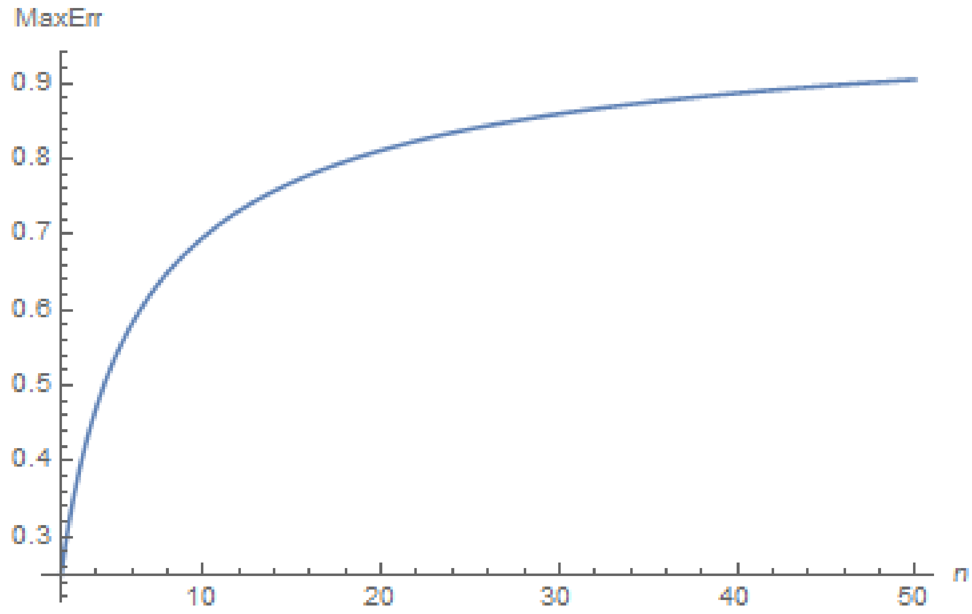


Figure 1: Maximum errors in likelihood under complete dependency for values of n ranging from 2 to 50.

4. Worst case scenario for posterior

Finding out how large errors can get for posterior might be more interesting than the error in likelihood, since the posterior is used for classification. This section shows that the maximum error in posterior can get arbitrarily close to 1, but will never reach 1, and shows the circumstances under which this is possible.

By Bayes' rule, we have that the equation for the posterior probability of a value w class variable W given a set observations $\boldsymbol{x}$ for a set of n evidence variables $\boldsymbol{X}$ is

$$\Pr(w|\boldsymbol{x}) = \frac{\Pr(w) \cdot \Pr(\boldsymbol{x}|w)}{\Pr(\boldsymbol{x})} \tag{36}$$

Where

$$\Pr(\boldsymbol{x}) = \sum_{w_i \in W} \Pr(w_i) \cdot \Pr(\boldsymbol{x}|w_i) \tag{37}$$

In the latter equation, every possible value of the class variable W is used for w_i.

As we have already seen in Equations 11 and 12, we have that

$$\Pr_{exact}(\boldsymbol{x}|w) = \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot \ldots \cdot \Pr(x_n|x_1, x_2, ..., x_{n-1}, w) \tag{38}$$

$$\Pr_{naive}(\boldsymbol{x}|w) = \prod_{k=1}^{n} \Pr(x_k|w) \tag{39}$$

So now, for the error in posterior due to ignoring possible dependencies among evidence variables we have that

$$Err(w|\boldsymbol{x}) = \left| \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) \cdot ... \cdot \Pr(x_n|x_1, x_2, ..., x_{n-1}, w)}{\sum_{w_i \in W}(\Pr(w_i) \cdot \Pr(x_1|w_i) \cdot \Pr(x_2|x_1, w_i) \cdot ... \cdot \Pr(x_n|x_1, x_2, ..., w_i))} - \frac{\Pr(w) \cdot \prod_{k=1}^{n} \Pr(x_k|w)}{\sum_{w_i \in W}(\Pr(w_i) \cdot \prod_{k=1}^{n} \Pr(x_k|w_i))} \right| \quad (40)$$

4.1. Two pieces of evidence

To simplify Equation 40 for the error in posterior, we will first consider the case where we only have two pieces of evidence. In this case the equation for the error reduces to

$$Err(w|\boldsymbol{x}) = \left| \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w)}{\sum_{w_i \in W}(\Pr(w_i) \cdot \Pr(x_1|w_i) \cdot \Pr(x_2|x_1, w_i))} - \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)}{\sum_{w_i \in W}(\Pr(w_i) \cdot \Pr(x_1|w_i) \cdot \Pr(x_2|w_i))} \right| \quad (41)$$

Without loss of generality, we can use binary values for W, where w is the value for which we calculate the error in posterior, while $\neg w$ consists of all other possible values. This gives us the following equation:

$$Err(w|\boldsymbol{x}) = \left| \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w)}{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) + \Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|x_1, \neg w)} - \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)}{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w) + \Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)} \right| \quad (42)$$

Like in the previous section, we want to maximise this error. Now, in addition to the constraints for likelihood from Equations 18 and 19, we have the same constraints for the other values of class variable W.

$$\Pr(x_1|\neg w) \cdot \Pr(x_2|x_1, \neg w) \leq \Pr(x_2|\neg w) \quad (43)$$

$$\Pr(x_1|\neg w) \cdot (1 - \Pr(x_2|x_1, \neg w)) \leq 1 - \Pr(x_2|\neg w) \quad (44)$$

To find out what the maximum error for the error in posterior from Equation 40 is, we first consider the largest error theoretically possible. Since both the naive posterior probability and the exact posterior probability need to have a value between 0 and 1, the largest difference between these two we could possibly reach is 1 [1]. To reach an error of 1, we would need either the exact posterior probability to be 1 and the naive posterior

[1] Using the NMaximize function from Mathematica finds a maximum error of 0.999829, with $\Pr(w) = 0.112312$, $\Pr(x_1|w) = 0.00191407$, $\Pr(x_2|x_1, w) = 0.553218$, $\Pr(x_2, w) = 0.110328$, $\Pr(x_1|\neg w) = 0.302287$, $\Pr(x_2|x_1, \neg w) = 0.$ and $\Pr(x_2|\neg w) = 0.517867$.

probability to be 0 or vice versa. To get an error approaching 1, either one or both of the posterior probabilities only need to approach an extreme value, rather than equal the extreme value.

Theorem 4.1. *Let $f(w|\boldsymbol{x})$ be the equation for the error in posterior for 2 evidence variables from Equation 42, under the constraints from Equations 18, 19, 43 and 44. Now, $f(w|\boldsymbol{x})$ will be maximized when w is deterministic in the context of $\boldsymbol{x}$ as defined in Definition 3. In this case $f(w|\boldsymbol{x})$ can get arbitrarily close to 1, but will never become equal to 1.*

To prove this we split the problem in several cases. First we prove that achieving an error of 1 is impossible, since the exact approach cannot have 1 as a result when the naive approach has 0 as a result and vice versa.

Lemma 4.2. *Let f be as before, under the same constraints as before, and let*

$$\Pr_{naive}(w|\boldsymbol{x}) \in \{0, 1\} \tag{45}$$

Now, $f(w|\boldsymbol{x}) = 0$.

Proof. $\Pr_{naive}(w|\boldsymbol{x}) = 0$ only holds when $\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w) = 0$, which means that at least one of the probabilities in this product equals 0. If $\Pr(w) = 0$ or $\Pr(x_1|w) = 0$, then $\Pr_{exact}(w|\boldsymbol{x}) = 0$. If $\Pr(x_2|w) = 0$, then Equation 18 dictates that $\Pr(x_1|w) \cdot \Pr(x_2|x_1, w) = 0$, which again results in $\Pr_{exact}(w|\boldsymbol{x}) = 0$. In all these cases $f(w|\boldsymbol{x}) = 0$.

$\Pr_{naive}(w|\boldsymbol{x}) = 1$ only holds when $\Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w) = 0$, which means that at least one of the probabilities in this product equals 0. If $\Pr(\neg w) = 0$ or $\Pr(x_1|\neg w) = 0$, then $\Pr_{exact}(w|\boldsymbol{x}) = 1$. If $\Pr(x_2|\neg w) = 0$, then Equation 43 dictates that $\Pr(x_1|\neg w) \cdot \Pr(x_2|x_1, \neg w) = 0$, which again results in $\Pr_{exact}(w|\boldsymbol{x}) = 1$. All these cases also result in $f(w|\boldsymbol{x}) = 0$. □

Corollary 4.2.1. *The error in posterior for two evidence variables cannot equal 1.*

Lemma 4.3. *Let f be as before, under the same constraints as before, and let*

$$\Pr_{exact}(w|\boldsymbol{x}) \in \{0, 1\} \tag{46}$$

Now, either $f(w|\boldsymbol{x}) = 0$ or $f(w|\boldsymbol{x})$ can get arbitrarily close to 1.

Proof. $\Pr_{exact}(w|\boldsymbol{x}) = 0$ only holds when $\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) = 0$, which means that at least one of the probabilities in this product equals 0. If $\Pr(w) = 0$ or $\Pr(x_1|w) = 0$, then $\Pr_{naive}(w|\boldsymbol{x}) = 0$. In both of these cases $f(w|\boldsymbol{x}) = 0$, so we will focus further on $\Pr(x_2|x_1, w)$. If $\Pr(x_2|x_1, w) = 0$ we get

$$f(w|\boldsymbol{x}) = \left| 0 - \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)}{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w) + \Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)} \right| \tag{47}$$

Now, since $\Pr_{exact}(w|\boldsymbol{x}) = 0$, $\Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|x_1, \neg w) > 0$. If $\Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|x_1, \neg w)$ were 0, $\Pr(\boldsymbol{x})$ would be 0 as well, which means we would have an impossible combination of evidence. As a result of this, $\Pr(w) = 1 - \Pr(\neg w) < 1$ and both $\Pr(x_1|\neg w)$ and $\Pr(x_2|x_1, \neg w)$ are non-zero. Since $\Pr(x_2|x_1, \neg w) > 0$, $\Pr(x_2|\neg w)$ has to be non-zero as well.

To maximize Equation 47 we need $\Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)$ to approach 0, since this fraction approaches a value of 1 when the numerator and denominator are roughly equal. This can be achieved by either taking the limit of $\Pr(w)$ to 1 or the limit of $\Pr(x_1|\neg w)$ or $\Pr(x_2|\neg w)$ to 0. Here we choose the first option:

$$MaxErr(w|\boldsymbol{x}) = \lim_{\Pr(w)\to 1} \left| 0 - \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)}{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w) + (1 - \Pr(w)) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)} \right| = 1 \quad (48)$$

where $\Pr(x_1|\neg w)$ and $\Pr(x_2|\neg w)$ can have any arbitrary value. Each additional limit brings the value for f closer to 1.

When $\Pr_{exact}(w|\boldsymbol{x}) = 1$, we analogously have that the only case that does not result in $f(w|\boldsymbol{x}) = 0$ is when $\Pr(x_2|x_1, \neg w) = 0$. Since $\Pr_{exact}(w|\boldsymbol{x}) = 1$, we need that $\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|x_1, w) > 0$, and thus all these probabilities are non-zero, which means $\Pr(x_2|w)$ is non-zero as well. Now, in order to maximize the error, we need to minimize $\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)$, which can be achieved by taking the limit of any of these to 0. Again, we choose the first option:

$$MaxErr(w|\boldsymbol{x}) = \lim_{\Pr(w)\to 0} \left| 1 - \frac{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w)}{\Pr(w) \cdot \Pr(x_1|w) \cdot \Pr(x_2|w) + (1 - \Pr(w)) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)} \right| = 1 \quad (49)$$

□

Proof of Theorem 4.1. Theorem 4.1 follows directly from Lemma 4.3 and Corollary 4.2.1, since Lemma 4.3 uses context specific determinism as a premise and proves the error can get arbitrarily close to 1, and Corollary 4.2.1 shows that the error in posterior cannot become 1. □

This proves that the error for posterior can get arbitrarily close to 1, when we have two evidence variables, but will never reach 1. These results also show that while *context dependent determinism* as defined in Definition 3 is necessary for an error approaching 1, *complete dependency* as defined in Definition 1 is not required. The worst case scenario for the error in likelihood from Section 3, which was achieved under complete dependency, where $\Pr(x_1|w) = \Pr(x_2|w) = 0.5$ and $\Pr(x_2|x_1, w) = 0$ is a possibility for a worst case scenario for the error in posterior as well, but only if $\Pr(\neg w) \cdot \Pr(x_1|\neg w) \cdot \Pr(x_2|\neg w)$ is extremely low. This means that complete dependency is neither a guarantee nor a requirement for a maximum error for posterior.

Because the error in posterior can already get arbitrarily close to 1 when we only have 2 evidence variables, the same has to be true when we have more than 2 variables, since we can introduce evidence variables that do not add any new information and we will have the same results. For example we can add a third evidence variable X_3 that has $\Pr(x_3) = 1$ as a prior. In this case, any set of evidence that includes x_3 has the same results as the results with only 2 evidence variables, while each set of evidence that includes $\neg x_3$ is impossible.

5. Pilot experiment

To get some more insight in the circumstances that lead to larger errors and just how large these errors get, I have first performed a small pilot experiment. In this experiment we will only consider likelihoods. Since the prior probability of the class variable does not influence likelihood at all, we will only consider evidence variables in this experiment. The goal of this experiment is to find the highest possible error in likelihood, caused by wrongly assuming independence among evidence variables. Moreover, we will analyze the relation between this error and the actual dependency among evidence variables.

5.1. Method

5.1.1. Computational structure

For this experiment I generate data structures containing probabilities and conditional probabilities which are used to calculate both the naive and exact likelihoods. I will refer to a single instantiated case of such a data structure as a *network*. An example of such a data structure is shown in Figure 2.

I construct two different types of network.

Assuming a total order on evidence variables, the first is a network where each evidence variable k is only dependent on the variable $k-1$, which I will refer to as *chain networks*[2]. The second is a network where each evidence variable is dependent on all predecessors, which I will call *complete networks*.

These networks serve to compute $\Pr(\boldsymbol{x}|w)$ for some assignment $\boldsymbol{x}$ to evidence variables $\boldsymbol{X}$, and some value w for class variable W. Without loss of generality, we assume each X_i to be instantiated to x_i. The error in likelihood caused by wrongly assuming the evidence variables to be independent given the class is then defined as

$$Err = |\Pr(\boldsymbol{x}|w)_{exact} - \Pr(\boldsymbol{x}|w)_{naive}| \tag{50}$$

where

$$\Pr(\boldsymbol{x}|w)_{naive} = \Pr(x_1|w) \prod_{k=2}^{n} \Pr(x_k|w) = \prod_{k=1}^{n} \Pr(x_k|w) \tag{51}$$

is called the naive likelihood.

[2]Note that these networks are graph structures that provide for computing probabilities, but they are not a member of the family of probabilistic graphical models.

The computation of the exact likelihood $\Pr(\boldsymbol{x}|w)_{exact}$ differs for the two types of network. For chain networks we have

$$\Pr(\boldsymbol{x}|w)_{exact} = \Pr(x_1|w) \prod_{k=2}^{n} \Pr(x_k|x_{k-1}, w) \tag{52}$$

whereas for complete networks we use

$$\Pr(\boldsymbol{x}|w)_{exact} = \prod_{k=1}^{n} \Pr(x_k|x_{k-1}, \ldots, x_1, w) \tag{53}$$

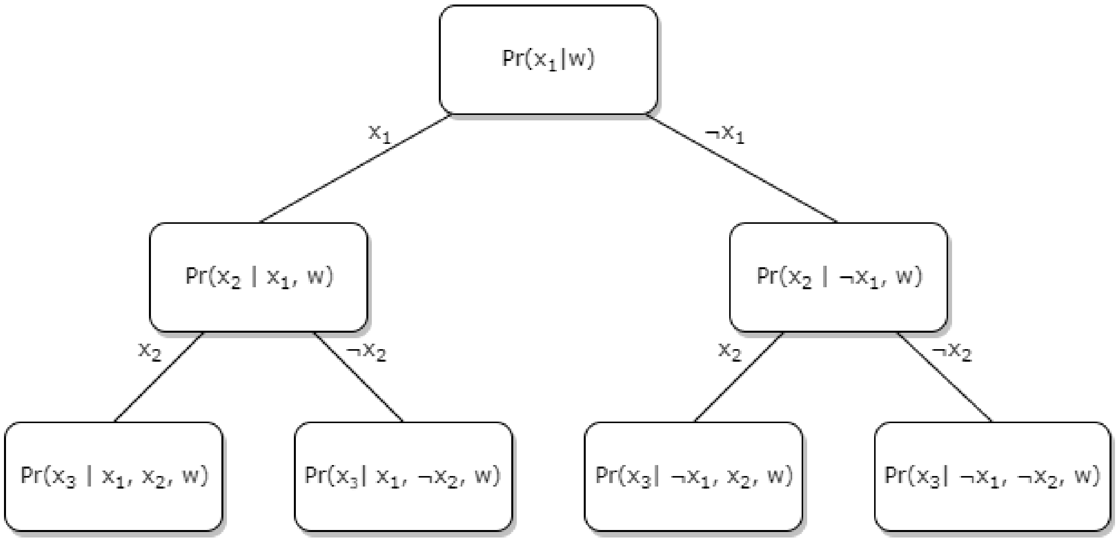
For chain networks, a much simpler structure than the data structure shown in Figure 2 suffices, since only a single value of each X_i is considered. The probabilities used in *chain networks* are created by first generating a random probability for the first evidence variable, $\Pr(x_1|w)$. Then, for each additional evidence variable k, a random conditional probability, $\Pr(x_k|x_{k-1}, w)$, and the corresponding probability, $\Pr(x_k|w)$ independent of x_{k-1} will be generated. As has already been established in Section 3, these probabilities have certain constraints, so the latter will be generated under the constraints from Equations 18 and 19.

For complete networks, the complete structure as shown in Figure 2 is required. The probabilities associated with the nodes in this structure are all generated randomly. From this structure, the exact likelihood can be computed by multiplying the probability contained in the leftmost node in each level of the structure.

The probabilities $\Pr(x_k|w)$ necessary for computing the naive likelihood can be calculated by multiplying every probability at the kth level with its parent if it is the left child, or by one minus its parent when it is the right child, until we reach the top, and then summing all these values. For example, to calculate $\Pr(x_2|w)$ we get

$$\Pr(x_2|w) = \Pr(x_2|x_1, w) \cdot \Pr(x_1|w) + \Pr(x_2|\neg x_1, w) \cdot (1 - \Pr(x_1|w)) \tag{54}$$

Figure 2: Probability structure for $n = 3$



5.1.2. Dependency

Since we are interested in the effect of dependency among the evidence variables on the error, we want some way to quantify dependency. This way we can determine the correlation between the error and the dependency measure. For this pilot experiment I decided to use four different dependency measures.

Yule's Q statistic [9] or *Q-Score* measures the strength of association between two binary variables. Q-Score values range from -1 to 1, where 0 means independency, and -1 and 1 indicate maximum dependency.

$$Q(X_a, X_b) = \frac{\Pr(x_a, x_b) \cdot \Pr(\neg x_a, \neg x_b) - \Pr(x_a, \neg x_b) \cdot \Pr(\neg x_a, x_b)}{\Pr(x_a, x_b) \cdot \Pr(\neg x_a, \neg x_b) + \Pr(x_a, \neg x_b) \cdot \Pr(\neg x_a, x_b)} \tag{55}$$

Mutual Information [10] is a measure of dependency between two variables. Unlike Q-Score, however, these variables do not have to be binary. Mutual information can not be lower than 0, and higher values means more dependency.

$$MI(X_a, X_b) = \sum_{x_a \in X_a} \sum_{x_b \in X_b} \Pr(x_a, x_b) \cdot \log \frac{\Pr(x_a, x_b)}{\Pr(x_a) \cdot \Pr(x_b)} \tag{56}$$

Since in our case the variables are binary, X_a only contains x_a and $\neg x_a$ and X_b contains only x_b and $\neg x_b$.

The *Simple* dependency measure I introduce in this thesis, is based on the main difference between dependency and independency of two variables. When two variables X_a and X_b are independent, then by definition the conditional probability $\Pr(x_b|x_a)$ is equal to $\Pr(x_b)$. When the variables are dependent, however, having evidence for x_a would affect the probability for x_b, so the difference between $\Pr(x_b|x_a)$ and $\Pr(x_b|)$

would become larger when the variables are more dependent. The Simple dependency quantifier is simply the absolute difference between $\Pr(x_b|x_a)$ and $\Pr(x_b)$.

$$Simple(X_a, X_b) = |\Pr(x_b|x_a) - \Pr(x_b)| \tag{57}$$

Since this measure only considers x_a and x_b instead of $\neg x_a$ and $\neg x_b$ as well, this measure is biased towards the problem at hand, where we only consider cases where all given evidence has value x_k. Since both probabilities are limited to be between 0 and 1, and since we use absolute values, the range of this dependency measure is between 0 and 1 as well, where 0 is independent, and higher values are more dependent.

The last quantifier I introduce is simply the *Pairwise error* in likelihood for two variables. Similarly to the simple score, this measure is biased to the problem at hand, since it ignores the cases where $\neg x_a$ or $\neg x_b$. The range of this measure is again between 0 and 1, where 0 means independency and 1 is most dependent.

$$PairwiseError(X_a, X_b) = \Pr(x_a) \cdot |\Pr(x_b|x_a) - \Pr(x_b)| \tag{58}$$

Due to the bias in both the simple dependency measure and pairwise error, I expect stronger correlations between these measures and errors found in this experiment than between the other measures and the error.

Since these measures are all defined only for two variables, the average of the absolute values for these measures for all pairs of dependent evidence variables will be used to capture dependence among all variables, and I will refer to these measures as pairwise dependency measures. For chain networks these will just be between each evidence variable and the next, while for complete networks this will be every possible combination of two variables. So for chain networks the average pairwise dependency is given by

$$APD_{Chain} = \frac{1}{n-1} \sum_{i=1}^{n-1} |f(i, i+1)| \tag{59}$$

And for complete networks this is

$$APD_{Complete} = \frac{2}{n \cdot (n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |f(i,j)| \tag{60}$$

Where f is the dependency measure used.

All probabilities needed to calculate a single pairwise dependency between variables a and b, can be calculated using just the following probabilities: $\Pr(x_a|w)$, $\Pr(x_b|w)$ and $\Pr(x_b|x_a, w)$, how all other probabilities can be calculated can be found in Appendix A. For chain networks the three initial probabilities needed are directly available. For complete networks $\Pr(x_a|w)$ and $\Pr(x_b|w)$ can be calculated using the manner described in Section 5.1.1. $\Pr(x_b|x_a, w)$ can be calculated by first using a similar method for $\Pr(x_b|w)$, but ignoring all values where $\neg x_a$. This gives us $\Pr(x_b, x_a|w)$. Dividing this by $\Pr(x_a|w)$ results in $\Pr(x_b|x_a, w)$. For example when we want to calculate $\Pr(x_3|x_1, w)$

we get

$$\Pr(x_3, x_1|w) = \Pr(x_3|x_1, x_2, w) \cdot \Pr(x_2|x_1, w) \cdot \Pr(x_1|w) + \Pr(x_3|x_1, \neg x_2, w) \cdot (1 - \Pr(x_2|x_1, w)) \cdot \Pr(x_1|w) \quad (61)$$

$$\Pr(x_3|x_1, w) = \frac{\Pr(x_3, x_1|w)}{\Pr(x_1|w)} \quad (62)$$

5.1.3. Experimental set-up

A single run of this experiment generates 100.000 networks with n evidence variables. For chain networks I have done runs from 2 to 50 variables, and for complete networks I have done runs from 2 to 15 variables, since these are computationally more demanding.

For each value of n, and for both types of networks, the highest and average error found is recorded, and the correlation between this error and the average pairwise dependency as defined in Equations 59 and 60 for each measure is calculated. Moreover, for each type of network and number of evidence variables, I performed a run where I enforced strong dependencies among evidence variables by forcing specific probabilities to be at least ρ. For ρ I use 0.9 and 0.999; note that $\rho = 0$ is similar to complete randomness.

In chain networks, I force dependency by constraining $\Pr(x_k|x_{k-1}, w)$ to be between ρ and 1, instead of completely random.

Forcing dependency in complete networks constrains all probabilities in the *main path* to be between ρ and 1. The main path is composed of every leftmost node on each level, so each node where the evidence is x_k.

5.2. Results

5.2.1. Random networks

The results of the experiments on the chain networks and complete networks for $n \leq 10$ are shown in Tables 2 and 3, respectively. This shows a maximum error that peaks at 3 evidence variables for the chain network, and at 5 variables for the complete network, and it subsequently decreases when more evidence variables are introduced. The average error always decreases with the addition of more evidence variables. The correlation between the error and the dependency as quantified by the different measures decreases with the addition of evidence as well, where the correlations with Simple dependency measure and Q-score even become negative. The results for networks with more than 10 evidence variables have been omitted from these tables, since they followed the same pattern and did not add any new information.

Table 2: Error in likelihood and correlation with average pairwise dependency for different measures in chain networks for $\rho = 0$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.2401	0.05579	0.5692	1	0.6917	0.9294
3	0.3119	0.04561	0.2336	0.7045	0.4403	0.6455
4	0.2957	0.03019	0.04818	0.5704	0.3007	0.4864
5	0.2730	0.01814	-0.03667	0.5065	0.2123	0.4021
6	0.2756	0.01042	-0.1127	0.4384	0.1228	0.3249
7	0.1397	0.005742	-0.1336	0.3858	0.06250	0.2729
8	0.1817	0.003047	-0.1399	0.3509	0.03488	0.2385
9	0.1141	0.001669	-0.1398	0.2934	-0.001517	0.1854
10	0.09392	0.0008864	-0.1442	0.2676	-0.02524	0.1692

Table 3: Error in likelihood and correlation with average pairwise dependency for different measures in complete networks for $\rho = 0$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.244	0.05614	0.5563	1	0.6810	0.9318
3	0.2868	0.05256	0.03386	0.4701	0.2823	0.4417
4	0.2877	0.03785	-0.1971	0.2522	0.1338	0.2341
5	0.3231	0.02438	-0.2651	0.1438	0.06301	0.1323
6	0.2909	0.01393	-0.2754	0.09714	0.03981	0.08302
7	0.2447	0.008079	-0.2625	0.05172	0.02302	0.05026
8	0.2571	0.004283	-0.2417	0.04580	-0.003744	0.03400
9	0.09496	0.002304	-0.2426	0.008203	-0.01523	0.003243
10	0.09763	0.001240	-0.2123	0.003738	-0.003088	0.0009956

5.2.2. Forced dependency $\rho = 0.9$

The results for chain networks and complete networks with $\rho = 0.9$ and $n \leq 15$ are shown in Tables 4 and 5, respectively. The highest maximum error in likelihood is found at 11 variables for chain networks and 14 for complete networks. The highest average error is found at 12 for chain networks and at 10 for complete networks. The relation between number of evidence nodes and errors found in these tables seems less clear than in the tables for random networks. These errors still seem to increase up to a certain number of evidence variables, and decrease when even more are added. Except for Q-score, all correlations with the dependency as quantified by the measures decrease with the addition of more evidence variables. The correlation between error and the Q-score measure seems to improve when the number of evidence variables is increased.

Table 4: Error in likelihood and correlation with average pairwise dependency for different measures in chain networks for $\rho = 0.9$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.2466	0.07483	0.5056	1	0.5851	0.9547
3	0.3492	0.1002	0.4415	0.9891	0.6316	0.9431
4	0.41	0.1128	0.3671	0.9742	0.635	0.9316
5	0.4236	0.1239	0.3046	0.9639	0.6464	0.9213
6	0.4621	0.1288	0.2418	0.9518	0.6478	0.9087
7	0.4361	0.1338	0.1868	0.9443	0.6628	0.8938
8	0.4063	0.1354	0.1306	0.9338	0.6754	0.882
9	0.4216	0.1375	0.1051	0.9258	0.6925	0.8671
10	0.4372	0.1385	0.0418	0.9221	0.7044	0.8592
11	0.4685	0.1378	-0.005102	0.9102	0.7017	0.8398
12	0.42	0.1386	-0.03722	0.9054	0.7257	0.8301
13	0.4118	0.1354	-0.06882	0.8947	0.7132	0.8096
14	0.3765	0.1346	-0.1004	0.8913	0.7223	0.799
15	0.3464	0.1312	-0.1295	0.8823	0.7128	0.7861

Table 5: **Errors and correlations with dependency measures** in complete networks for $\rho = 0.9$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.2465	0.07517	0.5164	1	0.5898	0.9533
3	0.3493	0.1313	0.2171	0.8996	0.371	0.8512
4	0.42	0.1742	-0.02109	0.7901	0.3313	0.7694
5	0.4468	0.207	-0.2345	0.6766	0.39	0.6969
6	0.4498	0.2305	-0.3922	0.5963	0.5284	0.6668
7	0.484	0.2452	-0.5036	0.517	0.6572	0.6272
8	0.5092	0.257	-0.5761	0.4806	0.7588	0.6325
9	0.4996	0.2597	-0.6168	0.4718	0.8393	0.6431
10	0.5045	0.2614	-0.6429	0.4668	0.8949	0.6584
11	0.52	0.2606	-0.6358	0.4912	0.9268	0.6868
12	0.5078	0.256	-0.6349	0.5003	0.9495	0.7122
13	0.52	0.2478	-0.6135	0.5551	0.9644	0.7478
14	0.5223	0.2407	-0.5817	0.5909	0.9744	0.7748
15	0.5102	0.2348	-0.5454	0.604	0.9797	0.7871

5.2.3. Forced dependency $\rho = 0.999$

The results for chain networks and complete networks with $\rho = 0.999$ are shown in Tables 6 and 7, respectively. The maximum error for chain networks peaks at 14 evidence variables, while the maximum error found for complete networks is found at 15 variables.

The highest average error is found at 12 evidence variables for chain networks. For complete networks the average error keeps increasing when more variables are added for $n \leq 15$. Again all correlations with dependency measures seem to decrease with a higher number of evidence variables, except Q-score. For chain networks the correlation between error and Q-score peaks at 7 variables and for complete networks the correlation between error and Q-score keeps increasing with more variables for $n \leq 15$. For complete networks, the correlation between the error and mutual information starts increasing as well, at 7 evidence variables.

Table 6: Error in likelihood and correlation with average pairwise dependency for different measures in chain networks for $\rho = 0.999$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.2477	0.08302	0.4873	1	0.1962	0.9617
3	0.373	0.1104	0.4375	0.9896	0.2604	0.9589
4	0.4304	0.1251	0.4071	0.978	0.3028	0.9557
5	0.4837	0.1289	0.4026	0.9677	0.3328	0.9508
6	0.5174	0.1299	0.3746	0.9626	0.3517	0.9502
7	0.5293	0.1319	0.3861	0.9597	0.3642	0.9487
8	0.5415	0.1327	0.3913	0.9574	0.3495	0.9473
9	0.4912	0.1331	0.3958	0.9601	0.3451	0.9497
10	0.5254	0.1329	0.3922	0.9563	0.3011	0.9471
11	0.512	0.1305	0.3946	0.957	0.3061	0.9492
12	0.5315	0.1341	0.403	0.9594	0.2758	0.9513
13	0.5046	0.1334	0.3795	0.9565	0.2462	0.948
14	0.5526	0.1334	0.3919	0.959	0.2427	0.9498
15	0.5254	0.133	0.3846	0.9579	0.2313	0.9481

Table 7: Error in likelihood and correlation with average pairwise dependency for different measures with dependency measures in complete networks for $\rho = 0.999$

Evidence #	Max error	Avg error	Simple corr	Pair err corr	Q-score corr	MI corr
2	0.2489	0.08319	0.4862	1	0.1991	0.9617
3	0.3759	0.1445	0.263	0.9145	0.08199	0.8815
4	0.4459	0.1955	0.08605	0.8411	0.1133	0.846
5	0.5024	0.2313	-0.05096	0.7778	0.1922	0.8289
6	0.5379	0.2596	-0.1263	0.7306	0.29	0.8295
7	0.5536	0.2852	-0.1623	0.7049	0.3718	0.8396
8	0.5805	0.304	-0.2338	0.6743	0.4869	0.8554
9	0.5939	0.3194	-0.2252	0.658	0.5125	0.8659
10	0.6262	0.3367	-0.2776	0.6319	0.6145	0.8744
11	0.6227	0.3487	-0.2629	0.6134	0.6232	0.8858
12	0.6669	0.3589	-0.2577	0.6081	0.6651	0.8932
13	0.6607	0.37	-0.2884	0.5823	0.7059	0.9015
14	0.6694	0.3751	-0.2772	0.568	0.7191	0.9053
15	0.681	0.3834	-0.2669	0.5554	0.7402	0.9107

Figure 3 shows the relations between the likelihood error and Q-Score, respectively Mutual Information, for complete networks with $n = 5$ and $\rho = 0$, $\rho = 0.9$ and $\rho = 0.999$. In Figure 4, the graphs for all values of ρ for Q-Score and Mutual Information are combined in a single graph. Each point shows the error and the value of the dependency measure for a single network. We can use these graphs to examine the correlation between the error and Q-score or mutual information.

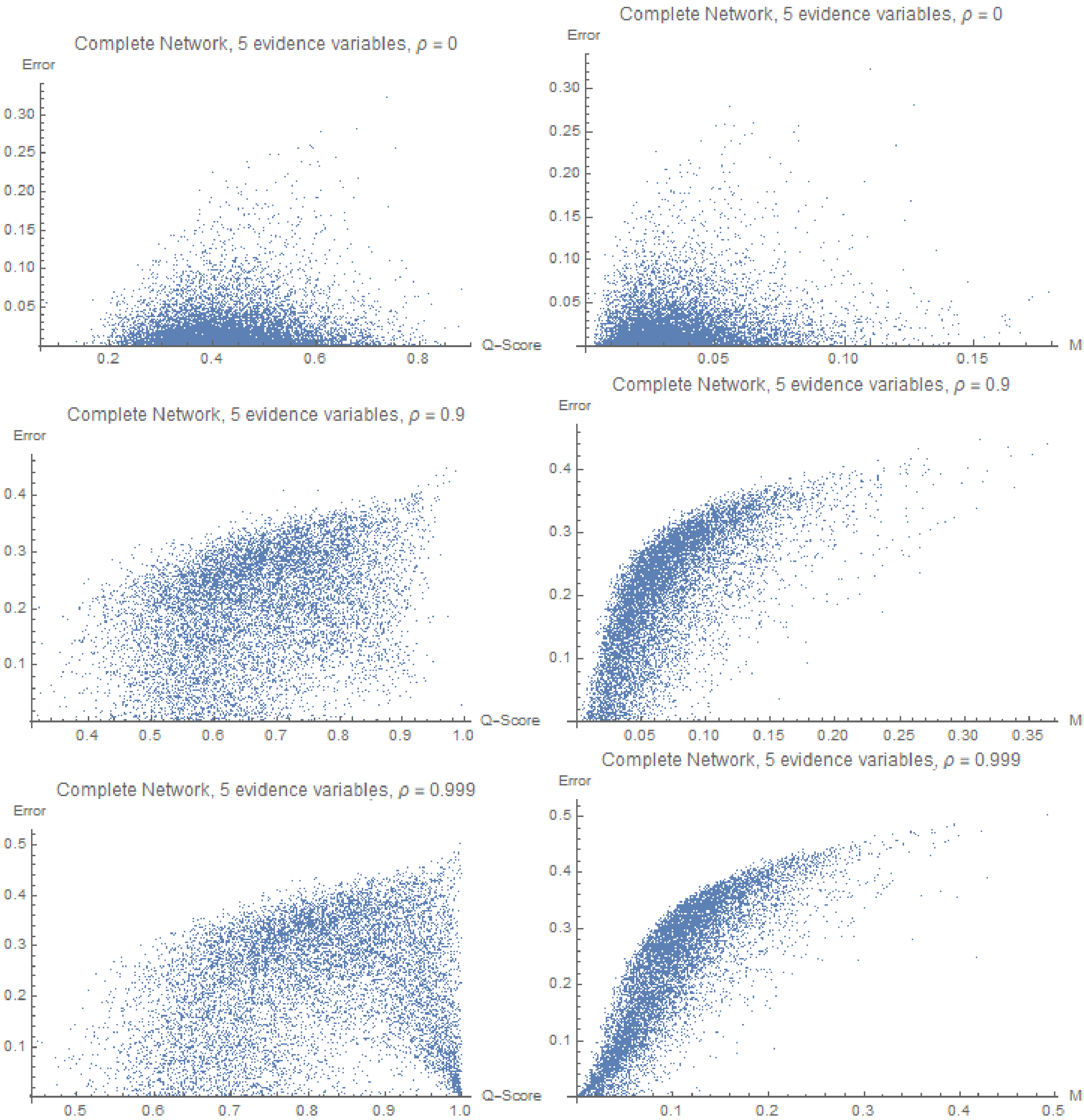


Figure 3: These graphs show all combinations of errors and Q-score, and errors and mutual information found for $n = 5$ and $\rho = 0$, $\rho = 0.9$ and $\rho = 0.999$. Every point represents the results for a single network.

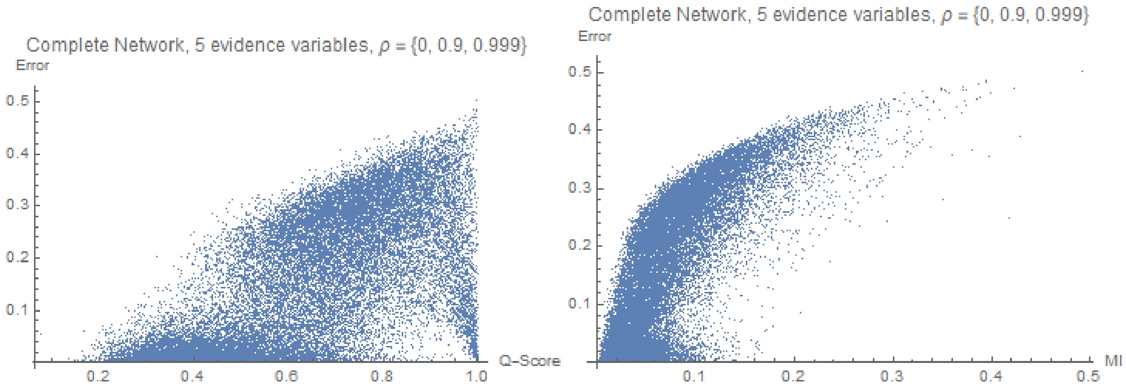


Figure 4: These graphs show the graphs from figure 3 combined in a single image for both Q-score and mutual information.

5.3. Analysis

None of the maximum errors found in this experiment exceeded the theoretically established errors found in Table 1, which provides additional proof that complete dependency is the worst case scenario for likelihood. It is also quite clear that forcing higher dependency results in larger errors. Correlation between the error and dependency as quantified by the various measures, however, seems to be quite low in networks without forced dependency, while forcing dependency increases this correlation for Q-Score and Mutual Information. The simple dependency measure seems to perform poorly for every case. Q-Score correlation actually seems to improve when more evidence variables are introduced in the cases with forced dependency. Pairwise likelihood error only seems to perform well in chain networks. This can be explained by the lower number of connections between evidence variables, which causes this measure to resemble the calculation for error more.

The poor correlation between error and dependency as quantified by the various measures for networks without forced dependency can be explained by the low average error found in those networks. And as we can see in the two top graphs in Figure 3, there are many results that have low values for both error and Q-Score or Mutual Information. Forcing dependency gives us both larger errors and higher Q-Scores or Mutual Information, as we can see in the other graphs in Figure 3. Figure 4 shows this in more detail. All values found in the case without forced dependency for Q-Score are at the bottom of this graph, and for Mutual Information all these values are in the bottom left. This graph shows that for a high error, a high Q-Score seems to be a requirement. A high Q-Score is, however, no guarantee for a high error. Mutual Information seems to correlate better with error, since all points with a high error also have a high Mutual Information and vice versa.

All of these results seem to imply that higher dependency leads to larger errors and that complete dependency is indeed the worst case scenario for likelihood.

6. Experiment

Since the pilot experiment only considered likelihoods, an additional experiment that also considered the error in posterior is in order. This final experiment is an extended version of the pilot experiment. The biggest difference between the pilot experiment and this experiment is the addition of posterior. Other differences include additional errors, apart from just the main path, and additional dependency measures. The chain networks from the pilot experiment have been omitted, since the additional data from these experiments did not add anything to the experiment, and these networks are a subset of the complete networks.

The goals of this experiment are as follows.

The first goal is finding what the maximum error and average error is we can find for both likelihood and posterior. Again, we define the error in likelihood and posterior as the absolute difference between the likelihood and posterior in the naive case and the exact likelihood and posterior. We also want to find out under what circumstances we get a larger error, and more specifically how dependency and determinism, as defined in Section 2, influence the error in both likelihood and posterior. This will be achieved by both forcing dependency and determinism and calculating the correlation between the errors and the different dependency measures. We also want to find out the relation between the error in likelihood and the error in posterior. Lastly we want to examine the relation between the prior of the class variable and the error in posterior. The relation between the prior of the class variable and the error in likelihood is of no interest, since the error in likelihood is not affected by the prior at all.

6.1. Method

6.1.1. Computational structure

The most important extension to the experiment is the addition of calculations for the error in posterior, rather than just for likelihood. This means that now we also need the class variable and conditional probabilities when the value of W is $\neg w$.

The network is now formed by generating a prior probability for the class, which will serve as the root of this structure, and generating a probability structure, similar to the structure from Section 5.1.1, for both values of the class as children.

These networks serve to compute both $\Pr(\boldsymbol{x}|w)$ and $\Pr(w|\boldsymbol{x})$ for any assignment $\boldsymbol{x}$ to evidence variables $\boldsymbol{X}$ and any value w for class variable W. The error in likelihood is again defined by

$$Err_{likelihood} = |\Pr(\boldsymbol{x}|w)_{exact} - \Pr(\boldsymbol{x}|w)_{naive}| \tag{63}$$

and is calculated in the same way as in Section 5.1.1. The error in posterior caused by wrongly assuming independence among evidence variables is defined as

$$Err_{posterior} = |\Pr(w|\boldsymbol{x})_{exact} - \Pr(w|\boldsymbol{x})_{naive}| \tag{64}$$

where

$$\Pr(w|\boldsymbol{x}) = \frac{\Pr(\boldsymbol{x}|w) \cdot \Pr(w)}{\Pr(\boldsymbol{x}|w) \cdot \Pr(w) + \Pr(\boldsymbol{x}|\neg w) \cdot \Pr(\neg w)} \tag{65}$$

$\Pr(w|\boldsymbol{x})_{exact}$ is now obtained by using exact likelihoods from Equation 53 for $\Pr(\boldsymbol{x}|w)$ and $\Pr(\boldsymbol{x}|\neg w)$, and $\Pr(w|\boldsymbol{x})_{naive}$ can be obtained by using the naive likelihoods from Equation 51.

These networks can be used to calculate various probabilities.

Simple joint probabilities for complete sets of evidence $\boldsymbol{x}$ and a value w for W can be calculated similarly to how the exact likelihood was calculated in Equation 53, except now we also have to include the class node

$$\Pr(w, \boldsymbol{x}) = \Pr(w) \prod_{k=1}^{n} \Pr(x_k | x_{k-1}, ..., x_1, w) \tag{66}$$

Any other joint probability for (sub)set of evidence $\boldsymbol{x}$ can be calculated by summing the joint probability of all possible combinations of evidence that matches the (sub)set of evidence used as input. For example

$$\Pr(x_1, x_3) = \Pr(w, x_1, x_2, x_3) + \Pr(\neg w, x_1, x_2, x_3) + \Pr(w, x_1, \neg x_2, x_3) + \Pr(\neg w, x_1, \neg x_2, x_3) \tag{67}$$

Note that we only have to use the first $k+1$ levels of the probability structure, where k is the highest index used in (sub)set $\boldsymbol{x}$.

Any conditional probability $\Pr(\boldsymbol{x}_a|\boldsymbol{x}_b)$ can be calculated with

$$\Pr(\boldsymbol{x}_a|\boldsymbol{x}_b) = \frac{\Pr(\boldsymbol{x}_a, \boldsymbol{x}_b)}{\Pr(\boldsymbol{x}_b)} \tag{68}$$

where $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ can be any subset of $\boldsymbol{x} \cup \{w\}$.

6.1.2. Dependency

To determine the correlation between errors and dependency, we again compute some dependency measures for each network. The dependency measures can be divided into two categories: pairwise dependency measures and network dependency measures. Where the pairwise dependency measures are averaged over every combination of two evidence variables, as explained in Section 5.1.2, and the network dependency measures are properties of the entire network.

The pairwise dependency measures used in this experiment are the same as the ones used in the pilot experiment. However, since we now have the entire network, instead of just the part conditioned on w, the probabilities used for the pairwise dependency measures also will not be conditioned on w.

Apart from only calculating the pairwise dependency between the evidence variables, we also calculate the average pairwise dependencies between the class and each evidence variable.

The second category of dependencies are dependencies that concern the entire network. These are all generalizations of mutual information, called total correlation, interaction information [11] and their conditional variants.

Total correlation is defined as

$$TC(X_1, X_2, ..., X_n) = \sum_{\substack{x_1 \in X_1, \\ x_2 \in X_2, \\ ..., \\ x_n \in X_n}} \Pr(x_1, x_2, ..., x_n) \log \frac{\Pr(x_1, x_2, ..., x_n)}{\prod_{i=1}^{n} \Pr(x_i)} \tag{69}$$

In our case, each variable can have only two values, so X_i can have x_i and $\neg x_i$ as possible values.

Interaction information is mostly similar to total correlation. The difference is that instead of using just the complete joint probability and the priors, interaction information uses every possible combination of evidence, where all joint probabilities with an even amount of variables are multiplied and serve as the numerator, and all joint probabilities with an even amount of variables are multiplied and serve as the denominator. So for three variables this would be defined as

$$II(X_1, X_2, X_3) = \sum_{\substack{x_1 \in X_1, \\ x_2 \in X_2, \\ x_3 \in X_3}} \Pr(x_1, x_2, x_3) \log \frac{\Pr(x_1, x_2) \cdot \Pr(x_1, x_3) \cdot \Pr(x_2, x_3)}{\Pr(x_1) \cdot \Pr(x_2) \cdot \Pr(x_3) \cdot \Pr(x_1, x_2, x_3)} \tag{70}$$

The total correlation that is computed in this implementation is $TC(W, X_1, X_2, ...X_n)$ and the interaction information is $II(W, X_1, X_2, ...X_n)$. The conditional variants of these can be obtained by multiplying the multivariate mutual information for the first tree with the class prior and adding the multivariate mutual information of the second tree multiplied by one minus the class prior to it. So for conditional total correlation we get

$$CTC(X_1, X_2, ..., X_n|W) = \sum_{w \in W} \Pr(w) \sum_{\substack{x_1 \in X_1, \\ x_2 \in X_2, \\ ..., \\ x_n \in X_n}} \Pr(x_1, x_2, ..., x_n|w) \log \frac{\Pr(x_1, x_2, ..., x_n|w)}{\prod_{i=1}^{n} \Pr(x_i|w)} \tag{71}$$

And for interaction information, for three variables, this would be

$$\begin{aligned} CII(X_1, X_2, X_3|W) = &\sum_{w \in W} \Pr(w) \\ &\sum_{\substack{x_1 \in X_1, \\ x_2 \in X_2, \\ x_3 \in X_3}} \Pr(x_1, x_2, x_3|w) \log \frac{\Pr(x_1, x_2|w) \cdot \Pr(x_1, x_3|w) \cdot \Pr(x_2, x_3|w)}{\Pr(x_1|w) \cdot \Pr(x_2|w) \cdot \Pr(x_3|w) \cdot \Pr(x_1, x_2, x_3|w)} \end{aligned} \tag{72}$$

6.1.3. Experimental set-up

A single run of this experiment generates 100.000 networks with n evidence variables. I have done runs from 2 to 10 variables.

For each value of n, the highest and average errors in both likelihood and posterior are recorded. These errors are recorded for the case where all evidence is x_k, and for the case where each variable has a random value of either x_k or $\neg x_k$. For $n \leq 5$ we also calculate the average error for every possible set of evidence for each network and the average error for every possible subset of evidence for each network.

For each error we record, we also calculate the Pearson correlation coefficient between that error and every other error and dependency measure. These measures are the 4 pairwise dependency measures from Section 5.1.2 used to calculate the average pairwise dependency among evidence variables and the same 4 measures used to calculate the average pairwise dependency between the class node and each evidence variable. For networks where $n \leq 5$ the 4 network measures as defined in Section 6.1.2 will also be used.

Moreover, for each number of evidence variables, I did some runs where I constrained specific probabilities to enforce certain properties in the network.

The first constraint attempts to make networks more dependent by making the network more similar to complete dependency as defined in Section 2. This is best explained by dividing the probability structure into a separate class prior probability and two substructures. The first substructure is the left child of the class node, where each probability is conditioned on w, and the second substructure is the right child, where each probability is conditioned on $\neg w$.

Now, to enforce dependency we constrain the leftmost node of each level of both substructures to be higher than ρ, except for the root of each substructure. We also constrain the rightmost node of each level of both substructures to be lower than $1 - \rho$, except for the root of each substructure. This means that for a network with enforced dependency, the following equation hold true.

$$\begin{aligned}\forall k \in \{2, ..., n\} \, &\Pr(x_k | x_{k-1}, ..., x_1, w) \geq \rho \wedge \Pr(x_k | x_{k-1}, ..., x_1, \neg w) \geq \rho \\ &\wedge \Pr(x_k | \neg x_{k-1}, ..., \neg x_1, w) \leq (1 - \rho) \wedge \Pr(x_k | \neg x_{k-1}, ..., \neg x_1, \neg w) \leq (1 - \rho)\end{aligned} \tag{73}$$

Apart from forcing dependency, we also force networks to be more deterministic. Determinism will be enforced by constraining all leaves in one substructure to be higher than ρ or lower than $1 - \rho$, which will be randomly decided. The same probabilities in the other substructure will then be constrained by the opposite constraint. If the same constraint is used for every leaf in a single tree, one random probability will be reset using the other constraint. This makes sure we can't have a probability of 0 or 1 for $\Pr(w)$, which would make the class variable deterministic even without any observed evidence, leading to an error of 0. This means that for a network with enforced determinism, the following equations hold true.

$$\begin{aligned}\forall \boldsymbol{x} \in \boldsymbol{X}_{1,...,n-1} (&\Pr(x_n | \boldsymbol{x}, w) \leq 1 - \rho \wedge \Pr(x_n | \boldsymbol{x}, \neg w) \geq \rho) \\ &\vee (\Pr(x_n | \boldsymbol{x}, w) \geq \rho \wedge \Pr(x_n | \boldsymbol{x}, \neg w) \leq 1 - \rho)\end{aligned} \tag{74}$$

$$\exists \boldsymbol{x} \in \boldsymbol{X}_{1,...,n-1} \Pr(x_n | \boldsymbol{x}, w) \leq 1 - \rho \tag{75}$$

$$\exists \boldsymbol{x} \in \boldsymbol{X}_{1,...,n-1} \Pr(x_n | \boldsymbol{x}, w) \geq \rho \tag{76}$$

For both forced dependency and forced determinism I use $\rho = \{0.5, 0.9, 0.999, 1\}$.

Lastly I will use networks with fixed prior probabilities for the class node, but no forced dependency or determinism. This is achieved by setting the root of the structure to a fixed value Ω. The priors I use for this are $\Omega = \{0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.999\}$, 0 and 1 will not be used, since these priors do not allow any error.

6.2. Results

In this section, all relevant results from the experiments are presented. More detailed results can be found in the appendix. All results are rounded to four significant digits. In the tables found in this section *error* refers to the error found when all evidence is x_i and *random errors* are the errors found using random sets of evidence. For average errors in entire networks, *full errors* refer to the average errors found using only complete sets of evidence, and *all errors* refer to average errors found using every possible (sub)set of evidence. These results will be analyzed in more detail in Section 6.3.

6.2.1. Errors

Recall that the goal of this research was to find out how large the error in both likelihood and posterior can get, and under what circumstances these errors are largest. Figure 5 shows the maximum and average errors found in completely random networks and Figure 6 shows the averages of the average errors found in entire networks, for both every complete set of evidence and for every possible (sub)set of evidence. In these graphs we can see that all errors in posterior seem to increase with more evidence variables, while average likelihood errors only decrease and the maximum error in likelihood tends to increase until about 6 evidence variables, after which it decreases as well. These graphs use values from Tables 24 and 25, which are found in the appendix.

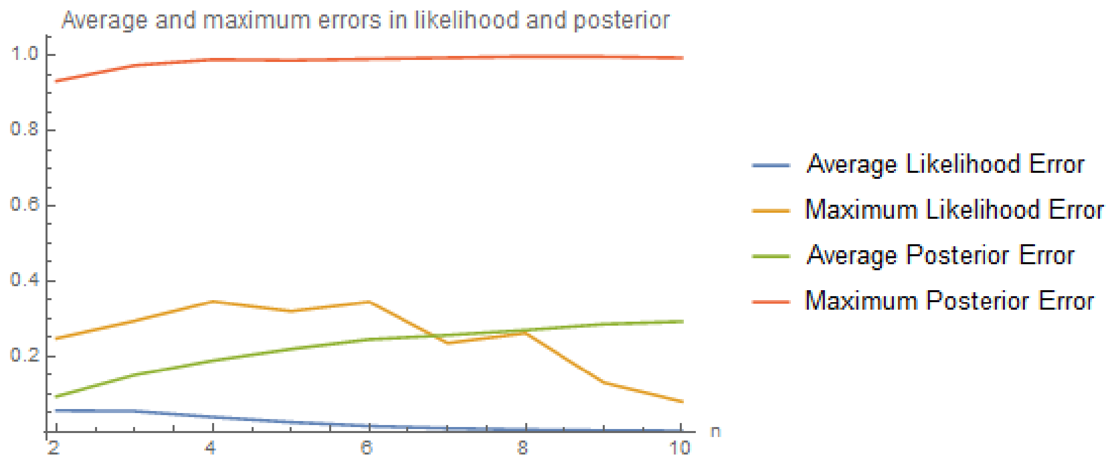


Figure 5: Average and Maximum errors found for likelihood and posterior for completely random networks

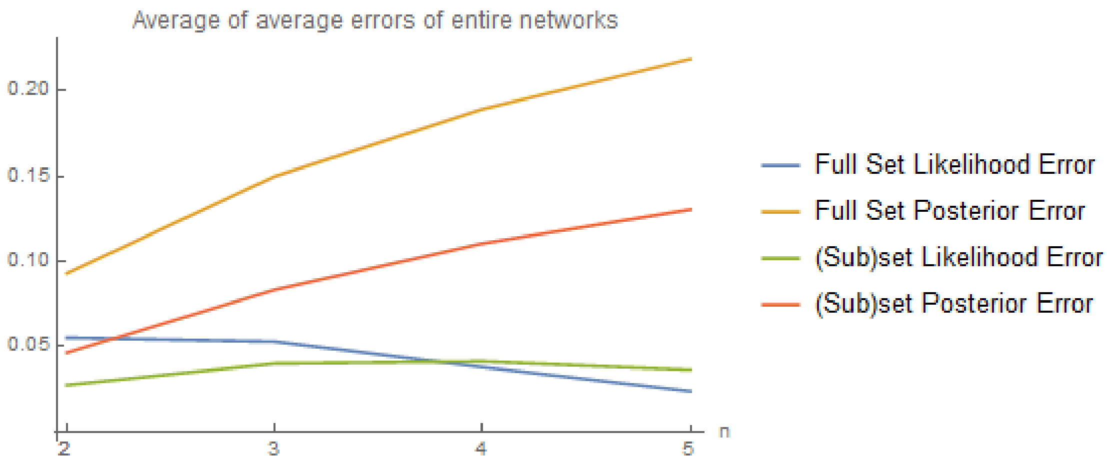


Figure 6: Average of average errors found in entire networks for likelihood and posterior for completely random networks

Tables 8 and 9 show the relation of maximum and average error in both likelihood and posterior to forced *dependency* and forced *determinism* as described in Section 6.1.3. These tables can be used to determine the circumstances under which the error is largest. Table 8 shows an increased error in both likelihood and posterior for higher values of ρ, although this effect is stronger for likelihood. Table 9 shows a more significant increase in posterior errors and a smaller increase in likelihood error. These tables show only the results for $n = 5$, since other values of n have similar results.

Table 8: Average and Maximum errors found for forced dependency networks for $n = 5$

Dependency ρ	0	0.5	0.9	0.999	1
Average Likelihood error	0.02415	0.102	0.2805	0.3326	0.3322
Maximum Likelihood error	0.3194	0.3898	0.4992	0.5344	0.535
Average Posterior error	0.2191	0.121	0.188	0.2037	0.2061
Maximum Posterior error	0.9883	0.8356	0.9379	0.9712	0.9893
Average Random Likelihood error	0.02403	0.02569	0.03599	0.04255	0.04195
Maximum Random Likelihood error	0.3294	0.356	0.4845	0.5344	0.535
Average Random Posterior error	0.2182	0.2054	0.26	0.286	0.4806
Maximum Random Posterior error	0.9972	0.978	0.9955	0.9998	1

Table 9: Average and Maximum errors found for forced determinism networks for $n = 5$

Determinism ρ	0	0.5	0.9	0.999	1
Average Likelihood error	0.02415	0.02437	0.03075	0.03236	0.03352
Maximum Likelihood error	0.3194	0.3023	0.2975	0.3736	0.3683
Average Posterior error	0.2191	0.2331	0.3138	0.4524	0.4799
Maximum Posterior error	0.9883	0.9816	0.9947	0.9998	1
Average Random Likelihood error	0.02403	0.02433	0.03046	0.0338	0.03338
Maximum Random Likelihood error	0.3294	0.2827	0.317	0.3835	0.3622
Average Random Posterior error	0.2182	0.2307	0.3182	0.4598	0.4801
Maximum Random Posterior error	0.9972	0.9983	0.9951	0.9998	1

Table 10 shows the highest errors found in this experiment. For likelihood these are all errors found in networks with forced *dependency* where $\rho = 1$. For posterior these errors are all found in networks with forced *determinism* where $\rho = 1$, however, similar errors where found for forced dependency where $\rho = 1$, except for $n = 2$, where the highest error found for forced dependency was 0.9999.

Table 10: Highest maximum errors found across all runs for every number of evidence variables

n	2	3	4	5	6	7	8	9	10
Likelihood	0.25	0.3849	0.4725	0.535	0.5824	0.6197	0.6501	0.6754	0.6968
Posterior	1	1	1	1	1	1	1	1	1

Figure 7 shows the relation between fixed priors as described in Section 6.1.3 and the errors found from these networks, to help examine the relation between prior and error in posterior. Since the prior does not affect likelihoods at all, only the posterior is considered.

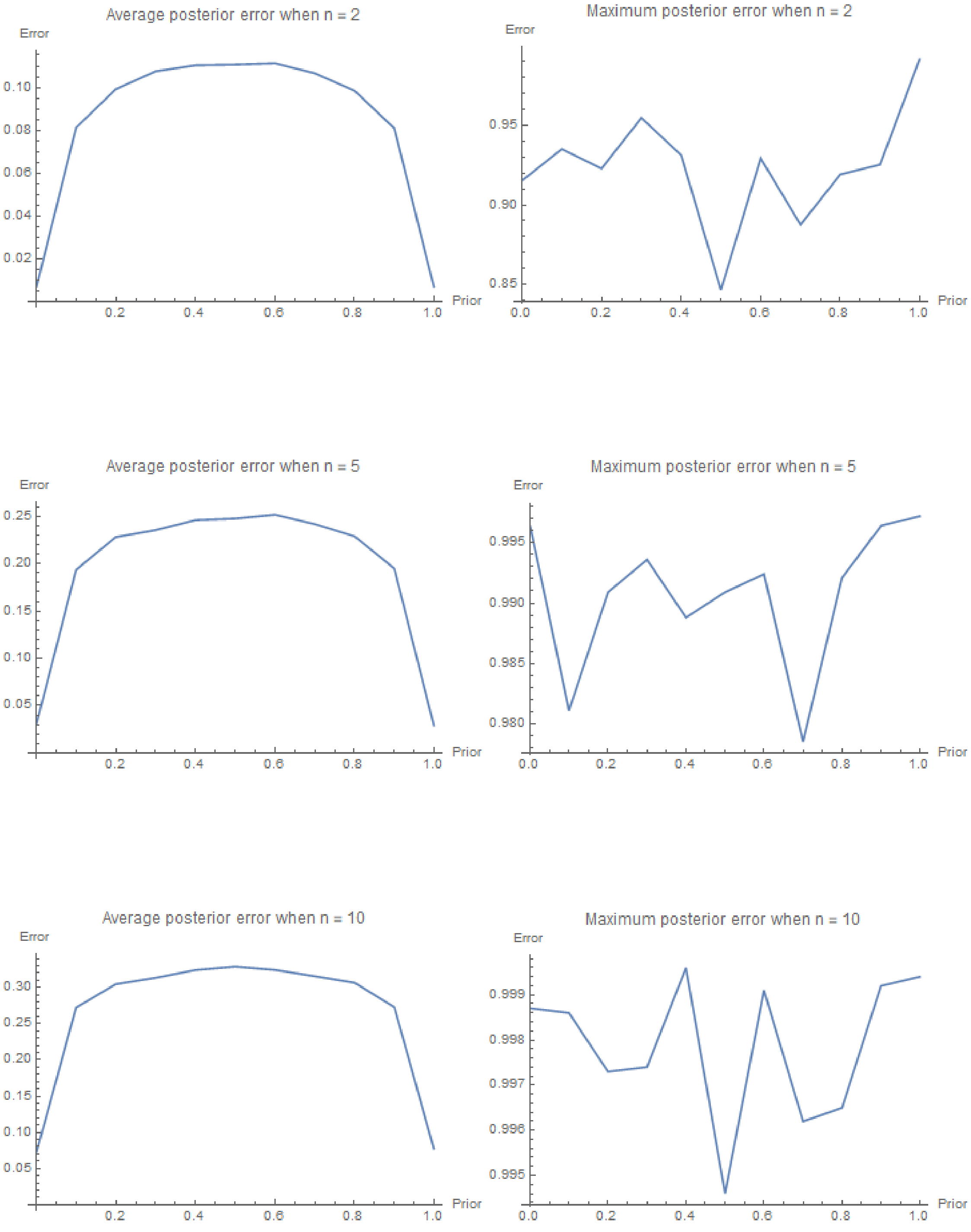


Figure 7: Average and Maximum errors found for each prior, for n = {2, 5, 10}

6.2.2. Correlation

In this section all correlations between errors and all correlations between errors and dependency measures are shown. Again only the results found for $n = 5$ are shown, since this is the highest amount of evidence variables for which the average errors of entire networks and the multivariate versions of mutual information are calculated, and since the other values of n have similar results. Results for all values of n can be found in the Appendix in Section C.

Tables 11 and 12 show the correlation between different errors and average errors.

Table 11: Correlation between errors for completely random networks with $n = 5$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.03033	0.01656	0.01232
Posterior Error	0.03033	1	0.001215	0.1219
Random Likelihood Error	0.01656	0.001215	1	0.0008565
Random Posterior Error	0.01232	0.1219	0.0008565	1

Table 12: Correlation between errors and average errors for completely random networks with $n = 5$

All Random	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.1473	0.05236	0.1358	0.05789
Posterior Error	0.09378	0.3233	0.09204	0.304
Random Likelihood Error	0.1236	0.02482	0.1154	0.02974
Random Posterior Error	0.08342	0.3329	0.08076	0.3206
Avg full Likelihood Error	1	0.2808	0.9136	0.304
Avg full Posterior Error	0.2808	1	0.2755	0.9463
Avg all Likelihood Error	0.9136	0.2755	1	0.3323
Avg all Posterior Error	0.304	0.9463	0.3323	1

Table 13 shows correlations between all errors and the pairwise dependency measures between evidence variables.

Table 13: Correlation between errors and pairwise dependency measures between evidence variables for completely random networks with $n = 5$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1028	0.04382	0.06337	0.06895
Posterior Error	0.09787	0.008088	0.002898	0.005412
Random Likelihood Error	0.03464	0.02366	0.04384	0.05138
Random Posterior Error	0.0006429	0.002902	0.006099	-0.0004764
Avg full Likelihood Error	0.261	0.1921	0.3376	0.4116
Avg full Posterior Error	-0.002488	0.004491	0.01063	0.0005879
Avg all Likelihood Error	0.3763	0.3254	0.4567	0.5326
Avg all Posterior Error	0.01023	0.02262	0.02314	0.01764

Table 14 shows the correlation between average errors in entire networks and pairwise dependency measures between the class and each evidence variable. The individual errors have been omitted, since the dependencies as measured by these dependency measures are properties of the entire network, which means it would correlate better with average error than with individual errors. This is the case because the average error is a property of the entire network as well, while an individual error is only a property of a specific part of the network.

Table 14: Correlation between errors and pairwise dependency measures between the class variable and each evidence variable for completely random networks with $n = 5$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Avg full Likelihood Error	-0.2379	-0.3183	-0.2978	-0.1771
Avg full Posterior Error	0.04771	-0.1384	0.04578	0.2783
Avg all Likelihood Error	-0.2268	-0.2981	-0.2803	-0.168
Avg all Posterior Error	0.1139	-0.08626	0.1385	0.3755

Table 15 shows the correlation between average errors and the various variants of multivariate mutual information.

Table 15: Correlation between errors and multivariate MI for completely random networks with $n = 5$

All Random	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8184	0.0275	0.3859	0.05481
Avg full Posterior Error	0.2903	-0.003732	0.2624	-0.03479
Avg all Likelihood Error	0.8206	-0.05733	0.4021	0.08558
Avg all Posterior Error	0.3172	-0.02669	0.36	0.003336

The following tables show the impact of several degrees of dependency on the correlations between average errors and multivariate mutual information. This shows whether or not dependency affects the correlations, as it did in the pilot experiment.

Table 16: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8041	0.2564	0.4062	0.0003541
Avg full Posterior Error	0.2635	0.05396	0.2783	-0.02181
Avg all Likelihood Error	0.8311	0.2563	0.4456	-0.02457
Avg all Posterior Error	0.2714	0.03941	0.3817	0.0335

Table 17: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8633	0.8231	0.4282	-0.4661
Avg full Posterior Error	-0.1682	-0.1406	0.06821	0.2727
Avg all Likelihood Error	0.8639	0.8246	0.4292	-0.4712
Avg all Posterior Error	-0.2566	-0.2145	0.1371	0.4406

Table 18: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8636	0.8635	0.4886	-0.5267
Avg full Posterior Error	-0.4551	-0.4545	-0.207	0.3373
Avg all Likelihood Error	0.8614	0.8614	0.4896	-0.5228
Avg all Posterior Error	-0.564	-0.5633	-0.1936	0.4923

Table 19: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8633	0.8633	0.5109	-0.5
Avg full Posterior Error	0.05957	0.05957	-0.01131	-0.09187
Avg all Likelihood Error	0.8624	0.8624	0.5123	-0.4971
Avg all Posterior Error	-0.17	-0.17	-0.05549	0.1541

The following tables show the impact of several degrees of determinism on the correlations between average errors and multivariate mutual information.

Table 20: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8229	0.03062	0.398	0.02612
Avg full Posterior Error	0.284	0.00985	0.2475	-0.05964
Avg all Likelihood Error	0.8247	-0.0571	0.4119	0.06839
Avg all Posterior Error	0.314	-0.03294	0.3464	-0.01306

Table 21: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8063	0.05408	0.2857	-0.0085
Avg full Posterior Error	0.1934	0.01148	0.1833	-0.08068
Avg all Likelihood Error	0.8109	-0.07125	0.3608	0.07105
Avg all Posterior Error	0.2316	-0.06049	0.3084	-0.002749

Table 22: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8226	0.09617	0.264	-0.05735
Avg full Posterior Error	0.3305	0.08858	0.1123	-0.1379
Avg all Likelihood Error	0.8091	-0.05229	0.3461	0.0553
Avg all Posterior Error	0.2891	-0.08382	0.3159	0.01183

Table 23: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Cond TC	Cond II	TC	II
Avg full Likelihood Error	0.8155	0.09013	0.2745	-0.06194
Avg full Posterior Error	0.3298	0.08726	0.01795	-0.0957
Avg all Likelihood Error	0.8051	-0.05112	0.3593	0.04109
Avg all Posterior Error	0.2764	-0.1169	0.2862	0.0412

6.3. Analysis

Again recall that the goals of this experiment were to find out how large errors can get and how large errors typically get. We also want to find out the circumstances under which the error in both likelihood and posterior is largest. For this reason we want to find out the relation between the error and various properties of networks. The properties researched in this experiment are the degree of dependency among evidence variables, the degree of determinism in the distributions, and the prior probability of the class variable W. Finally this research tries to find the relation between the errors in likelihood and posterior itself.

Figures 5 and 6 show that average errors in likelihood tend to be quite low, and even the maximum error found starts to decrease at $n = 6$, which corresponds to the results found in the pilot experiment. The maximum errors for likelihood found in the entire experiment shown in Table 10 correspond perfectly with the theoretical maximum from Table 1.

The errors in posterior, however, do seem to get higher when the number of evidence variables increases. The average error in posterior even reaches 0.292 at $n = 10$. The maximum errors found for posterior are all close to 1, and the maximum error found in Table 10 correspond to the theoretical results from Section 4 (note that an error of 1 in the experimental results means an error between 0.99995 and 1, since the values are rounded to four significant digits).

Table 8 shows a clear link between the average likelihood error and forced dependency. However, these networks are specifically tailored for the specific error measured, since we measure the error where all evidence is x_i and constrain only the paths in the network where all evidence is x_i and where all evidence is $\neg x_i$. When a random error is measured the effect of dependency is much lower, but still present, with an error of 0.02403 when $\rho = 0$ and an error of 0.04195 when $\rho = 1$. The maximum error found still increases in both cases, which makes sense, since in those cases the random error measured was the error when all evidence is x_i. Dependency seems to have a lesser effect on posterior, where the specific error measured in the main path of the tree actually has smaller average errors when dependency is forced, which already suggests that a high error in likelihood does not necessarily correspond to a high error in posterior. Random posterior errors in this network, however, do seem to get worse when dependency is forced, with an average error of 0.4806 and a maximum error of 1 when $\rho = 1$.

Table 9 shows the effect determinism has on the errors in likelihood and posterior. These networks are not biased towards any single error, so random errors are equal to the main path errors. Determinism has only a minor effect on likelihood error, which would be even smaller when the network has more variables, since determinism only affects a small part of the equation for the error in likelihood found in Equation 53. Posterior is affected by determinism as expected, with both higher average errors and maximum errors approaching 1. When $\rho = 1$ the average error even comes close to 0.5. This is the highest average error that can be expected, since outputting completely random values as a posterior would have an average error of 0.5 (since the exact posterior would always be 0 or 1 when $\rho = 1$).

Figure 7 shows how fixing the prior to specific values influences the average and maximum posterior errors found. As we have seen in section 4, one of the worst case scenarios is when the prior approaches either 0 or 1. However, the average error seems to get extremely small for extreme values. The maximum errors found are close to 1 for all values of Ω, although larger errors seem to be slightly more consistent for extreme priors. This would imply that even though the chance of getting a large error with an extreme prior is smaller, the resulting error is likely to be larger.

Another way this experiment tries to examine the relation between dependency and error is through correlations found between errors and various dependency measures. Table 13 shows correlations between the various errors and pairwise dependency measures between evidence variables. For individual errors, rather than average errors in entire networks, these values are all rather low, since dependency of an entire network does not correlate well with just a single error. The simple dependency measure seems to have the highest values for both likelihood and posterior, at -0.1028 and 0.09787, but this is not true for errors where a random set of evidence is used. This is because when the simple dependency measure calculates the dependency between a variable A and B it only considers evidence for a and b, and ignores all cases where $\neg a$ or $\neg b$ are true. This means simple dependency is biased towards the case where all evidence is x_i. The same holds for pairwise likelihood error. Average errors seem to work better for measuring the correlation between error and dependency measures. Even though the simple dependency measure is biased it still seems to have a higher correlation with the error in likelihood than Q-Score. Mutual information has an even better correlation though, and as can be expected, pairwise likelihood error has the highest correlation with the average likelihood error for the entire network. Errors in posterior, however, do not seem to correlate with these pairwise dependencies at all. This can be explained by the fact that posterior is dependent on more than just the relations between the evidence variables, but also depends on the class variable. This is why I also measured pairwise dependency between the class variable and each evidence variable, which can be seen in Table 14. This table shows a small, negative correlation between likelihoods and dependency between class and evidence, even though likelihood should not be affected by the class at all. This could be explained by considering the cases in which we get high dependency. In these cases the probabilities when $W = w$ and $W = \neg w$ differ greatly, and this seems to lead to a situation in which large errors in likelihood are unlikely. The correlations between average posterior errors and class-evidence dependencies do seem to be stronger than those from Table 13, with the pairwise likelihood error even reaching a correlation of 0.3755 with the average posterior error for all possible combinations of evidence. The highest overall results where achieved by using the total correlation measure, which can be seen in table 15. The conditional total correlation correlates very well with likelihood errors, which is to be expected, since part of Equation 71 is very similar to the equation for the error in likelihood found in Equation 50. Both the conditional and regular total correlation seem to correlate quite well with the error in posterior as well. The slightly more complicated interaction information does not seem to correlate with the error at all. This may be because the equation for interaction information from Equation 70 includes every possible combination of evidence, while the errors are calculated by only using a few of these.

Tables 16 through 19 show how forcing dependencies influences these correlations. The results for the two conditional measures become more similar when dependency is forced. This is because all the additional terms interaction information has over total correlation become equal to 1 under complete dependency. The normal interaction information, however, seems to correlate better with posterior when dependency is enforced. Tables 20 through 23 show similar tables for various degrees of determinism. Since the average error in posterior for full evidence sets gets larger in deterministic networks, while determinism does not influence the multivariate mutual information measures much, the correlation between this error and the Total Correlation measure is weakened. The conditional case, however, does seem to work better. Interaction information is even worse when determinism is involved, since only a small part of the equation uses the full set of evidence.

We have already seen that having a large error in likelihood does not necessarily mean we have a high error in posterior as well, or vice versa. Tables 11 and 12 show this in more detail. Table 11 shows that the correlation between likelihood and posterior for the same combination of evidence is extremely small at only 0.03033. The correlations between average values in Table 12 show that there is some correlation between the average error in likelihood and posterior. This corresponds with earlier data, since some networks seem to have higher errors in both likelihood and posterior, while other networks only have high errors in one of the two (deterministic networks).

7. Conclusion

As expected, the degree of dependency has a large impact on the error in both likelihood and posterior when ignoring dependencies among evidence variables. For likelihood, this thesis has shown in Section 3 that for 2 and 3 evidence variables complete dependency as defined in Section 2 results in the largest errors. In Section 3 I also introduced Equation 35, which is an equation that shows how big the error can get with n evidence variables under complete dependency. The experimental results from both Section 5 and 6 for the error in likelihood never exceed the values obtained from Equation 35, which makes it likely that complete dependency is indeed the worst case scenario for the error in likelihood for any number of evidence variables. The experiment from Section 5 shows that a high level of dependency is required for large errors in likelihood, but is no guarantee for large errors. Even though the error in likelihood can get quite large when the evidence variables are highly dependent, the average error in likelihood tends to be low.

The error in posterior seems more volatile, where the only requirement for an error approaching 1 appears to be determinism. As shown in Section 4, the error in posterior can get arbitrarily close to 1 under various circumstances when a value of the class variable is deterministically determined by a set of evidence variables, as defined in Section 2. These circumstances can be an extreme value on the prior $\Pr(w)$, or any conditional probability $\Pr(x_i|w)$ only conditioned on the class variable being extremely close to 0. The experimental results from Section 6 confirm this, finding maximum error values higher than 0.99995 for every number of evidence nodes in deterministic networks. The average error in posterior also seems to be more heavily influenced by determinism than by dependency. The error in posterior is less heavily influenced by determinism than the error in likelihood is influenced by dependency, since the average error in posterior is already quite large.

In the experiments from Section 5 and 6 I tested several measures of dependency in search of one that strongly correlates with the errors, which could be used to predict how large errors could be in a certain network. The pairwise dependency measures used in the experiments do not seem to be decent indicators for the errors found in any network, since they do not seem to correlate well with the errors. The total correlation measure seems better at predicting errors, however this measure is computationally not any less complex than directly computing the average error found in a network. Unfortunately, I therefore have not found any suitable way of predicting the error of using a naive approach in a network during this research.

The results found in the experiments are based on binary values for all variables. The results for the maximum errors can be generalized to any amount of values for each variable by grouping the values that are not true when this maximum error is found into a single value. The structures used in this thesis are generic as well, and thus the results can be generalized to more complex structures like Bayesian networks.

Further research into this subject could include real networks in the experiments, which would give some insights into how large errors typically get in cases where the naive approach is actually used, since the networks used in the experiments from Sections 5 and 6 are no realistic representations of real networks. Another idea is to separate the cases where the naive Bayesian classifier is correct from those that are classified incorrectly and repeat my experiments on these two separated groups. Since previous research has shown that the naive Bayesian classifier works quite well, it would be interesting to see how large the error would get in the cases where the naive approach works correctly as a classifier.

References

[1] I. Ben-Gal. Bayesian Networks. In F. Ruggeri, R.S. Kennett, F.W. Faltin. Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons. 2007.

[2] A. Stuart, K. Ord. Kendall's Advanced Theory of Statistics: Volume I - Distribution Theory. Edward Arnold. 1994.

[3] D. Hand and K. Yu. Idiots Bayes - not so stupid after all. International Statistical Review, 69: 385-398, 2001.

[4] I. Rish, J. Hellerstein, and J. Thathachar. An analysis of data characteristics that affect Naive Bayes performance. Technical Report RC21993, IBM TJ Watson Research Center, 2001.

[5] P. Domingos and M. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29: 103-130, 1997.

[6] L. Kuncheva and Z. Hoare. Error-dependency relationships for the naive Bayes classifier with binary features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30: 735-740, 2008.

[7] S. Renooij. Evidence Evaluation: a Study of Likelihoods and Independence. Proceedings of the Eighth International Conference on Probabilistic Graphical Models, PMLR 52: 426-437, 2016.

[8] D. Schum. The Evidential Foundations of Probabilistic Reasoning. Northwestern University Press. 1994

[9] G.U. Yule. On the methods of measuring association between two attributes. Journal of the Royal Statistical Society, 49 (6): 579-652. 1912.

[10] T.M. Cover and J.A. Thomas. Elements of Information Theory. John Wiley & Sons. 1991.

[11] T. van de Cruys. Two Multivariate Generalizations of Pointwise Mutual Information. Proceedings of the Workshop on Distributional Semantics and Compositionality: 16-20. 2011.

A. Probability functions

This section shows how all possible probabilities involving values x and y conditioned on any set of values $\mathbf{z}$ can be calculated from just $\Pr(x|\mathbf{z})$, $\Pr(y|\mathbf{z})$ and $\Pr(y|x,\mathbf{z})$. Note that $\mathbf{z}$ can be empty as well. All of these equations are derived from either Bayes' rule, the definition of conditional probability or from the fact that the sum of all probabilities for a single variable equals 1.

$$\Pr(x|y,\mathbf{z}) = \frac{\Pr(y|x,\mathbf{z}) \cdot \Pr(x|\mathbf{z})}{\Pr(y|\mathbf{z})} \tag{77}$$

$$\Pr(\neg x|\mathbf{z}) = 1 - \Pr(x|\mathbf{z}) \tag{78}$$

$$\Pr(\neg y|\mathbf{z}) = 1 - \Pr(y|\mathbf{z}) \tag{79}$$

$$\Pr(\neg x|y,\mathbf{z}) = 1 - \Pr(x|y,\mathbf{z}) \tag{80}$$

$$\Pr(\neg y|x,\mathbf{z}) = 1 - \Pr(y|x,\mathbf{z}) \tag{81}$$

$$\Pr(x|\neg y,\mathbf{z}) = \frac{\Pr(\neg y|x,\mathbf{z}) \cdot \Pr(x|\mathbf{z})}{\Pr(\neg y|\mathbf{z})} \tag{82}$$

$$\Pr(y|\neg x,\mathbf{z}) = \frac{\Pr(\neg x|y,\mathbf{z}) \cdot \Pr(y|\mathbf{z})}{\Pr(\neg x|\mathbf{z})} \tag{83}$$

$$\Pr(\neg x|\neg y,\mathbf{z}) = 1 - \Pr(x|\neg y,\mathbf{z}) \tag{84}$$

$$\Pr(\neg y|\neg x,\mathbf{z}) = 1 - \Pr(y|\neg x,\mathbf{z}) \tag{85}$$

$$\Pr(x,y|\mathbf{z}) = \Pr(x|y,\mathbf{z}) \cdot \Pr(y|\mathbf{z}) \tag{86}$$

$$\Pr(\neg x,y|\mathbf{z}) = \Pr(\neg x|y,\mathbf{z}) \cdot \Pr(y|\mathbf{z}) \tag{87}$$

$$\Pr(x,\neg y|\mathbf{z}) = \Pr(x|\neg y,\mathbf{z}) \cdot \Pr(\neg y|\mathbf{z}) \tag{88}$$

$$\Pr(\neg x,\neg y|\mathbf{z}) = \Pr(\neg x|\neg y,\mathbf{z}) \cdot \Pr(\neg y|\mathbf{z}) \tag{89}$$

B. Errors

All average and maximum errors found in the experiment are displayed in this section.

B.1. Completely random

Table 24: Average and Maximum errors found in completely random networks for $n \leq 5$

All Random	2	3	4	5
Average Likelihood error	0.05525	0.05341	0.03791	0.02415
Maximum Likelihood error	0.2473	0.2935	0.3456	0.3194
Average Posterior error	0.09248	0.1497	0.1883	0.2191
Maximum Posterior error	0.9335	0.9747	0.9905	0.9883
Average Random Likelihood error	0.05561	0.05395	0.038	0.02403
Maximum Random Likelihood error	0.2473	0.343	0.314	0.3294
Average Random Posterior error	0.09123	0.1518	0.1904	0.2182
Maximum Random Posterior error	0.9239	0.9546	0.982	0.9972
Average avg Full Likelihood error	0.05539	0.05302	0.03813	0.0241
Average avg Full Posterior error	0.0931	0.1499	0.1893	0.2186
Average avg All Likelihood error	0.02769	0.04023	0.04145	0.03655
Average avg All Posterior error	0.04655	0.08345	0.1106	0.1304

Table 25: Average and Maximum errors found in completely random networks for $n > 5$

All Random	6	7	8	9	10
Avg Likelihood error	0.01416	0.008152	0.004303	0.002365	0.001273
Max Likelihood error	0.3446	0.2348	0.2614	0.1292	0.07886
Avg Posterior error	0.2445	0.256	0.2691	0.2849	0.292
Max Posterior error	0.9917	0.995	0.9984	0.9977	0.9949
Avg Random Likelihood error	0.01402	0.008019	0.004279	0.002345	0.001272
Max Random Likelihood error	0.3019	0.2246	0.1643	0.1696	0.08479
Avg Random Posterior error	0.2403	0.2605	0.2714	0.2836	0.292
Max Random Posterior error	0.995	0.9893	0.9956	0.9999	0.9969

B.2. Dependency $\rho = 0.5$

Table 26: Average and Maximum errors found in networks with forced dependency, $\rho = 0.5$ for $n \leq 5$

Dependency 0.5	2	3	4	5
Average Likelihood error	0.08354	0.1108	0.1126	0.102
Maximum Likelihood error	0.2463	0.3468	0.3746	0.3898
Average Posterior error	0.05209	0.08405	0.1051	0.121
Maximum Posterior error	0.5933	0.7161	0.8782	0.8356
Average Random Likelihood error	0.08416	0.06426	0.04279	0.02569
Maximum Random Likelihood error	0.2444	0.3252	0.3297	0.356
Average Random Posterior error	0.08372	0.1358	0.1759	0.2054
Maximum Random Posterior error	0.91	0.9591	0.9744	0.978
Average avg Full Likelihood error	0.08367	0.06429	0.04225	0.0258
Average avg Full Posterior error	0.08287	0.1348	0.1756	0.2097
Average avg All Likelihood error	0.04184	0.05431	0.05151	0.04325
Average avg All Posterior error	0.04143	0.07317	0.09873	0.1207

Table 27: Average and Maximum errors found in networks with forced dependency, $\rho = 0.5$ for $n > 5$

Dependency 0.5	6	7	8	9	10
Avg Likelihood error	0.08774	0.07261	0.05791	0.04505	0.03425
Max Likelihood error	0.4119	0.4078	0.4193	0.371	0.2639
Avg Posterior error	0.1323	0.138	0.147	0.152	0.1531
Max Posterior error	0.9013	0.8632	0.8965	0.9265	0.8817
Avg Random Likelihood error	0.01478	0.008594	0.00439	0.00247	0.001265
Max Random Likelihood error	0.3669	0.3514	0.253	0.209	0.1265
Avg Random Posterior error	0.2385	0.2525	0.2702	0.2817	0.2957
Max Random Posterior error	0.9881	0.9875	0.9951	0.9957	0.9962

B.3. Dependency $\rho = 0.9$

Table 28: Average and Maximum errors found in networks with forced dependency, $\rho = 0.9$ for $n \leq 5$

Dependency 0.9	2	3	4	5
Average Likelihood error	0.1487	0.2198	0.259	0.2805
Maximum Likelihood error	0.2487	0.3734	0.4605	0.4992
Average Posterior error	0.07818	0.1261	0.1608	0.188
Maximum Posterior error	0.6818	0.8097	0.8933	0.9379
Average Random Likelihood error	0.1487	0.1125	0.06494	0.03599
Maximum Random Likelihood error	0.2487	0.3723	0.4524	0.4845
Average Random Posterior error	0.1254	0.181	0.2196	0.26
Maximum Random Posterior error	0.8708	0.9695	0.9979	0.9955
Average avg Full Likelihood error	0.1495	0.1111	0.06598	0.03655
Average avg Full Posterior error	0.1238	0.1791	0.2221	0.2539
Average avg All Likelihood error	0.07477	0.102	0.09988	0.08471
Average avg All Posterior error	0.06188	0.1056	0.1392	0.1651

Table 29: Average and Maximum errors found in networks with forced dependency, $\rho = 0.9$ for $n > 5$

Dependency 0.9	6	7	8	9	10
Avg Likelihood error	0.2888	0.2935	0.2953	0.2875	0.2862
Max Likelihood error	0.5394	0.5507	0.5582	0.581	0.589
Avg Posterior error	0.2022	0.217	0.2275	0.2403	0.2442
Max Posterior error	0.9623	0.9858	0.9898	0.9868	0.9907
Avg Random Likelihood error	0.01965	0.01038	0.005399	0.003113	0.001514
Max Random Likelihood error	0.4931	0.5143	0.5267	0.5	0.4794
Avg Random Posterior error	0.2795	0.293	0.3114	0.3242	0.3389
Max Random Posterior error	0.999	0.9993	0.9996	0.9999	0.9996

B.4. Dependency $\rho = 0.999$

Table 30: Average and Maximum errors found in networks with forced dependency, $\rho = 0.999$ for $n \leq 5$

Dependency 0.999	2	3	4	5
Average Likelihood error	0.1657	0.2503	0.2991	0.3326
Maximum Likelihood error	0.25	0.3848	0.4721	0.5344
Average Posterior error	0.08757	0.1419	0.1816	0.2037
Maximum Posterior error	0.7584	0.9297	0.9836	0.9712
Average Random Likelihood error	0.1658	0.1236	0.07581	0.04255
Maximum Random Likelihood error	0.25	0.3846	0.472	0.5344
Average Random Posterior error	0.1446	0.2061	0.2513	0.286
Maximum Random Posterior error	0.9771	0.9974	0.9985	0.9998
Average avg Full Likelihood error	0.1658	0.1249	0.07503	0.04148
Average avg Full Posterior error	0.1457	0.207	0.2525	0.2858
Average avg All Likelihood error	0.08289	0.1153	0.1151	0.09871
Average avg All Posterior error	0.07283	0.126	0.168	0.2001

Table 31: Average and Maximum errors found in networks with forced dependency, $\rho = 0.999$ for $n > 5$

Dependency 0.999	6	7	8	9	10
Avg Likelihood error	0.3595	0.3724	0.3903	0.3983	0.4048
Max Likelihood error	0.5817	0.6187	0.6489	0.674	0.6952
Avg Posterior error	0.2219	0.2399	0.251	0.2621	0.2708
Max Posterior error	0.9649	0.9838	0.9965	0.9917	0.9995
Avg Random Likelihood error	0.02211	0.01193	0.00638	0.003273	0.002057
Max Random Likelihood error	0.5812	0.6184	0.6481	0.6731	0.6943
Avg Random Posterior error	0.3141	0.3274	0.3502	0.3639	0.3711
Max Random Posterior error	1	1	1	1	1

B.5. Dependency $\rho = 1$

Table 32: Average and Maximum errors found in networks with forced dependency, $\rho = 1$ for $n \leq 5$

Dependency 1	2	3	4	5
Average Likelihood error	0.1664	0.2495	0.3022	0.3322
Maximum Likelihood error	0.25	0.3849	0.4725	0.535
Average Posterior error	0.08875	0.1432	0.1822	0.2061
Maximum Posterior error	0.646	0.9156	0.9655	0.9893
Average Random Likelihood error	0.1659	0.1251	0.07382	0.04195
Maximum Random Likelihood error	0.25	0.3849	0.4725	0.535
Average Random Posterior error	0.2918	0.4166	0.4554	0.4806
Maximum Random Posterior error	0.9999	1	1	1
Average avg Full Likelihood error	0.1666	0.1252	0.07545	0.04162
Average avg Full Posterior error	0.2945	0.4109	0.46	0.4818
Average avg All Likelihood error	0.08331	0.1155	0.1157	0.09902
Average avg All Posterior error	0.1472	0.2623	0.3445	0.4005

Table 33: Average and Maximum errors found in networks with forced dependency, $\rho = 1$ for $n > 5$

Dependency 1	6	7	8	9	10
Avg Likelihood error	0.3558	0.3752	0.391	0.3991	0.4115
Max Likelihood error	0.5824	0.6197	0.6501	0.6754	0.6968
Avg Posterior error	0.2233	0.2379	0.2528	0.2634	0.2704
Max Posterior error	0.9841	0.9989	0.9872	0.997	0.9986
Avg Random Likelihood error	0.02211	0.01138	0.006134	0.003367	0.001651
Max Random Likelihood error	0.5824	0.6197	0.6501	0.6754	0.6843
Avg Random Posterior error	0.4951	0.4957	0.4991	0.5003	0.506
Max Random Posterior error	1	1	1	1	1

B.6. Determinism $\rho = 0.5$

Table 34: Average and Maximum errors found in networks with forced determinism, $\rho = 0.5$ for $n \leq 5$

Determinism 0.5	2	3	4	5
Average Likelihood error	0.08302	0.05486	0.03809	0.02437
Maximum Likelihood error	0.248	0.3426	0.2858	0.3023
Average Posterior error	0.1699	0.1715	0.2067	0.2331
Maximum Posterior error	0.9365	0.977	0.9796	0.9816
Average Random Likelihood error	0.08277	0.05494	0.03786	0.02433
Maximum Random Likelihood error	0.2466	0.3155	0.296	0.2827
Average Random Posterior error	0.1696	0.174	0.2065	0.2307
Maximum Random Posterior error	0.954	0.984	0.9906	0.9983
Average avg Full Likelihood error	0.08298	0.05523	0.03807	0.02404
Average avg Full Posterior error	0.1692	0.1752	0.2057	0.2301
Average avg All Likelihood error	0.04149	0.04205	0.04135	0.03642
Average avg All Posterior error	0.0846	0.09692	0.1187	0.1359

Table 35: Average and Maximum errors found in networks with forced determinism, $\rho = 0.5$ for $n > 5$

Determinism 0.5	6	7	8	9	10
Average Likelihood error	0.01419	0.007989	0.004342	0.002397	0.001245
Maximum Likelihood error	0.2943	0.241	0.2426	0.1615	0.1083
Average Posterior error	0.2483	0.2675	0.28	0.2926	0.2973
Maximum Posterior error	0.9979	0.9921	0.9979	0.9987	0.997
Average Random Likelihood error	0.01427	0.008023	0.004413	0.002327	0.00128
Maximum Random Likelihood error	0.3163	0.2443	0.2346	0.2827	0.2021
Average Random Posterior error	0.2508	0.2668	0.2817	0.2896	0.2974
Maximum Random Posterior error	0.9945	0.9951	0.9975	0.9989	0.9985

B.7. Determinism $\rho = 0.9$

Table 36: Average and Maximum errors found in networks with forced determinism, $\rho = 0.9$ for $n \leq 5$

Determinism 0.9	2	3	4	5
Average Likelihood error	0.1501	0.07839	0.05153	0.03075
Maximum Likelihood error	0.2484	0.3449	0.4071	0.2975
Average Posterior error	0.3497	0.2866	0.3015	0.3138
Maximum Posterior error	0.9829	0.9952	0.9954	0.9947
Average Random Likelihood error	0.1493	0.08116	0.05131	0.03046
Maximum Random Likelihood error	0.2484	0.3612	0.3399	0.317
Average Random Posterior error	0.3446	0.2972	0.3055	0.3182
Maximum Random Posterior error	0.9829	0.995	0.9988	0.9951
Average avg Full Likelihood error	0.1498	0.08111	0.0514	0.03091
Average avg Full Posterior error	0.3464	0.2966	0.3033	0.3169
Average avg All Likelihood error	0.07491	0.05974	0.05382	0.04507
Average avg All Posterior error	0.1732	0.16	0.1702	0.1819

Table 37: Average and Maximum errors found in networks with forced determinism, $\rho = 0.9$ for $n > 5$

Determinism 0.9	6	7	8	9	10
Average Likelihood error	0.01727	0.009397	0.005138	0.00263	0.00142
Maximum Likelihood error	0.3134	0.2632	0.2707	0.1972	0.1471
Average Posterior error	0.3244	0.327	0.3331	0.3377	0.3365
Maximum Posterior error	0.9955	0.996	0.9978	0.999	0.9982
Average Random Likelihood error	0.01739	0.009406	0.005027	0.002532	0.001491
Maximum Random Likelihood error	0.3066	0.3601	0.2105	0.1319	0.3301
Average Random Posterior error	0.3212	0.331	0.3297	0.3357	0.3394
Maximum Random Posterior error	0.9994	0.9977	0.999	0.9993	0.9996

B.8. Determinism $\rho = 0.999$

Table 38: Average and Maximum errors found in networks with forced determinism, $\rho = 0.999$ for $n \leq 5$

Determinism 0.999	2	3	4	5
Average Likelihood error	0.1663	0.08586	0.05574	0.03236
Maximum Likelihood error	0.25	0.3668	0.3963	0.3736
Average Posterior error	0.4709	0.4097	0.4426	0.4524
Maximum Posterior error	0.9995	0.9999	0.9997	0.9998
Average Random Likelihood error	0.1652	0.0886	0.0563	0.0338
Maximum Random Likelihood error	0.25	0.3781	0.3779	0.3835
Average Random Posterior error	0.4723	0.4242	0.4411	0.4598
Maximum Random Posterior error	0.9995	0.9994	0.9996	0.9998
Average avg Full Likelihood error	0.1659	0.08878	0.05615	0.03339
Average avg Full Posterior error	0.4699	0.4233	0.4385	0.4567
Average avg All Likelihood error	0.08293	0.06497	0.05781	0.04787
Average avg All Posterior error	0.235	0.2171	0.2244	0.2315

Table 39: Average and Maximum errors found in networks with forced determinism, $\rho = 0.999$ for $n > 5$

Determinism 0.999	6	7	8	9	10
Average Likelihood error	0.01905	0.01014	0.005428	0.002796	0.001505
Maximum Likelihood error	0.3578	0.3565	0.2766	0.1666	0.1857
Average Posterior error	0.4623	0.4647	0.4656	0.4543	0.4543
Maximum Posterior error	0.9996	0.9996	0.9998	0.9998	0.9995
Average Random Likelihood error	0.01916	0.009934	0.005294	0.002848	0.001497
Maximum Random Likelihood error	0.3564	0.3084	0.2623	0.1922	0.2386
Average Random Posterior error	0.4627	0.4636	0.4575	0.4607	0.4515
Maximum Random Posterior error	0.9996	0.9993	0.9999	0.9997	1

B.9. Determinism $\rho = 1$

Table 40: Average and Maximum errors found in networks with forced determinism, $\rho = 1$ for $n \leq 5$

Determinism 1	2	3	4	5
Average Likelihood error	0.1662	0.08695	0.05587	0.03352
Maximum Likelihood error	0.25	0.381	0.3859	0.3683
Average Posterior error	0.4739	0.4218	0.4515	0.4799
Maximum Posterior error	1	1	1	1
Average Random Likelihood error	0.1677	0.08975	0.05667	0.03338
Maximum Random Likelihood error	0.25	0.3683	0.3785	0.3622
Average Random Posterior error	0.4778	0.434	0.4525	0.4801
Maximum Random Posterior error	1	1	1	1
Average avg Full Likelihood error	0.1667	0.08964	0.05621	0.03344
Average avg Full Posterior error	0.4781	0.4336	0.4551	0.4799
Average avg All Likelihood error	0.08337	0.0656	0.05771	0.0479
Average avg All Posterior error	0.2391	0.2208	0.2289	0.2361

Table 41: Average and Maximum errors found in networks with forced determinism, $\rho = 1$ for $n > 5$

Determinism 1	6	7	8	9	10
Average Likelihood error	0.01851	0.0103	0.005207	0.002819	0.001409
Maximum Likelihood error	0.3097	0.3133	0.3099	0.1665	0.1407
Average Posterior error	0.4991	0.4946	0.495	0.4962	0.4962
Maximum Posterior error	1	1	1	1	1
Average Random Likelihood error	0.01914	0.009989	0.005417	0.00282	0.001498
Maximum Random Likelihood error	0.3692	0.3511	0.2579	0.2053	0.1278
Average Random Posterior error	0.4916	0.4949	0.4943	0.5004	0.4996
Maximum Random Posterior error	1	1	1	1	1

B.10. Fixed prior $\Omega = 0.001$

Table 42: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.001$ for $n \leq 5$

Prior 0.001	2	3	4	5
Average Likelihood error	0.05672	0.05239	0.03842	0.02381
Maximum Likelihood error	0.2435	0.3034	0.3044	0.3327
Average Posterior error	0.007084	0.01307	0.02251	0.03115
Maximum Posterior error	0.9155	0.9665	0.9858	0.9963
Average Random Likelihood error	0.05577	0.05303	0.03817	0.02397
Maximum Random Likelihood error	0.2415	0.3379	0.3623	0.298
Average Random Posterior error	0.007101	0.01441	0.02273	0.02995
Maximum Random Posterior error	0.9696	0.9264	0.9919	0.9948
Average avg Full Likelihood error	0.05605	0.05312	0.03801	0.02404
Average avg Full Posterior error	0.007007	0.01371	0.02163	0.03025
Average avg All Likelihood error	0.02802	0.04023	0.04132	0.03645
Average avg All Posterior error	0.003504	0.006039	0.008299	0.01049

Table 43: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.001$ for $n > 5$

Prior 0.001	6	7	8	9	10
Avg Likelihood error	0.01415	0.00805	0.004467	0.002308	0.0013
Max Likelihood error	0.3082	0.2219	0.225	0.1577	0.1634
Avg Posterior error	0.03713	0.04916	0.05248	0.06449	0.07319
Max Posterior error	0.9987	0.9957	0.9989	0.9984	0.9987
Avg Random Likelihood error	0.01386	0.007895	0.004398	0.002303	0.00123
Max Random Likelihood error	0.365	0.2363	0.1983	0.1306	0.121
Avg Random Posterior error	0.04	0.04886	0.05759	0.06511	0.07445
Max Random Posterior error	0.9969	0.995	0.9978	0.9995	0.9992

B.11. Fixed prior $\Omega = 0.1$

Table 44: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.1$ for $n \leq 5$

Prior 0.1	2	3	4	5
Average Likelihood error	0.0551	0.05394	0.03822	0.02388
Maximum Likelihood error	0.2435	0.3158	0.313	0.3366
Average Posterior error	0.08117	0.1292	0.1693	0.1936
Maximum Posterior error	0.9351	0.9564	0.9866	0.9811
Average Random Likelihood error	0.05496	0.05332	0.03785	0.02348
Maximum Random Likelihood error	0.2445	0.347	0.3113	0.2786
Average Random Posterior error	0.08222	0.13	0.168	0.195
Maximum Random Posterior error	0.938	0.9845	0.9752	0.9964
Average avg Full Likelihood error	0.05526	0.05313	0.03799	0.02401
Average avg Full Posterior error	0.08078	0.13	0.1671	0.1949
Average avg All Likelihood error	0.02763	0.04032	0.04123	0.03638
Average avg All Posterior error	0.04039	0.07031	0.09261	0.1093

Table 45: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.1$ for $n > 5$

Prior 0.1	6	7	8	9	10
Avg Likelihood error	0.01425	0.007922	0.004206	0.002387	0.001219
Max Likelihood error	0.3548	0.2195	0.1484	0.1606	0.09189
Avg Posterior error	0.2188	0.2371	0.2498	0.2681	0.2723
Max Posterior error	0.9929	0.9921	0.9993	0.999	0.9986
Avg Random Likelihood error	0.01378	0.00798	0.004378	0.002356	0.001249
Max Random Likelihood error	0.2407	0.2414	0.2089	0.1372	0.1186
Avg Random Posterior error	0.2149	0.2302	0.2481	0.2628	0.2699
Max Random Posterior error	0.9963	0.9925	0.9931	0.9999	0.9962

B.12. Fixed prior $\Omega = 0.2$

Table 46: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.2$ for $n \leq 5$

Prior 0.2	2	3	4	5
Average Likelihood error	0.05602	0.05241	0.03808	0.02475
Maximum Likelihood error	0.239	0.3094	0.2999	0.3087
Average Posterior error	0.09921	0.1583	0.2003	0.2282
Maximum Posterior error	0.9229	0.97	0.9766	0.9909
Average Random Likelihood error	0.05538	0.05344	0.03875	0.02432
Maximum Random Likelihood error	0.2358	0.3182	0.3036	0.3806
Average Random Posterior error	0.09847	0.1575	0.1969	0.2234
Maximum Random Posterior error	0.8742	0.9601	0.9916	0.9951
Average avg Full Likelihood error	0.05584	0.05326	0.03799	0.02406
Average avg Full Posterior error	0.09893	0.157	0.1966	0.2276
Average avg All Likelihood error	0.02792	0.04043	0.04127	0.0365
Average avg All Posterior error	0.04947	0.08736	0.1146	0.1362

Table 47: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.2$ for $n > 5$

Prior 0.2	6	7	8	9	10
Avg Likelihood error	0.01422	0.007831	0.004502	0.002314	0.001196
Max Likelihood error	0.3	0.3117	0.2163	0.19	0.168
Avg Posterior error	0.2504	0.267	0.285	0.2947	0.3047
Max Posterior error	0.9913	0.9988	0.992	0.995	0.9973
Avg Random Likelihood error	0.01471	0.007897	0.004363	0.002389	0.001249
Max Random Likelihood error	0.3592	0.2899	0.2233	0.1852	0.1129
Avg Random Posterior error	0.2491	0.2661	0.2803	0.293	0.3007
Max Random Posterior error	0.9912	0.998	0.9957	0.9968	0.9939

B.13. Fixed prior $\Omega = 0.3$

Table 48: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.3$ for $n \leq 5$

Prior 0.3	2	3	4	5
Average Likelihood error	0.05549	0.05255	0.03808	0.02352
Maximum Likelihood error	0.2422	0.2868	0.3263	0.32
Average Posterior error	0.1075	0.1684	0.2099	0.2355
Maximum Posterior error	0.9547	0.9827	0.986	0.9936
Average Random Likelihood error	0.05541	0.05259	0.03849	0.02393
Maximum Random Likelihood error	0.2439	0.3027	0.3744	0.3125
Average Random Posterior error	0.106	0.171	0.2099	0.2391
Maximum Random Posterior error	0.9547	0.9827	0.9659	0.9779
Average avg Full Likelihood error	0.0557	0.05297	0.03792	0.02395
Average avg Full Posterior error	0.1073	0.1676	0.2082	0.2394
Average avg All Likelihood error	0.02785	0.04014	0.04122	0.03631
Average avg All Posterior error	0.05366	0.09436	0.1243	0.1473

Table 49: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.3$ for $n > 5$

Prior 0.3	6	7	8	9	10
Avg Likelihood error	0.01431	0.007969	0.004409	0.002453	0.001285
Max Likelihood error	0.2683	0.2358	0.2559	0.1071	0.1102
Avg Posterior error	0.2646	0.2806	0.2914	0.3098	0.3134
Max Posterior error	0.99	0.9948	0.9997	0.9909	0.9974
Avg Random Likelihood error	0.01395	0.007996	0.004275	0.002314	0.001206
Max Random Likelihood error	0.2995	0.3036	0.2508	0.1255	0.09228
Avg Random Posterior error	0.2634	0.2844	0.2925	0.3037	0.3154
Max Random Posterior error	0.986	0.9959	0.9982	0.9986	0.9966

B.14. Fixed prior $\Omega = 0.4$

Table 50: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.4$ for $n \leq 5$

Prior 0.4	2	3	4	5
Average Likelihood error	0.05626	0.05293	0.03838	0.02436
Maximum Likelihood error	0.241	0.2771	0.3158	0.273
Average Posterior error	0.1103	0.1757	0.2165	0.2461
Maximum Posterior error	0.9313	0.965	0.9866	0.9888
Average Random Likelihood error	0.0556	0.05323	0.03894	0.02419
Maximum Random Likelihood error	0.241	0.3207	0.3224	0.3178
Average Random Posterior error	0.1112	0.172	0.2156	0.247
Maximum Random Posterior error	0.979	0.9597	0.9744	0.9899
Average avg Full Likelihood error	0.05589	0.05277	0.03821	0.02402
Average avg Full Posterior error	0.1112	0.1734	0.2168	0.247
Average avg All Likelihood error	0.02795	0.03995	0.04154	0.03643
Average avg All Posterior error	0.05558	0.09824	0.1306	0.154

Table 51: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.4$ for $n > 5$

Prior 0.4	6	7	8	9	10
Avg Likelihood error	0.01412	0.007937	0.004325	0.00236	0.001257
Max Likelihood error	0.3012	0.296	0.1795	0.2251	0.1355
Avg Posterior error	0.2715	0.2896	0.3031	0.3179	0.3241
Max Posterior error	0.9934	0.9975	0.997	0.9964	0.9996
Avg Random Likelihood error	0.01425	0.008049	0.004276	0.002226	0.001226
Max Random Likelihood error	0.347	0.2648	0.1653	0.1527	0.07185
Avg Random Posterior error	0.275	0.2906	0.303	0.3156	0.3221
Max Random Posterior error	0.987	0.9953	0.9953	0.9956	0.9971

B.15. Fixed prior $\Omega = 0.5$

Table 52: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.5$ for $n \leq 5$

Prior 0.5	2	3	4	5
Average Likelihood error	0.05489	0.05281	0.03834	0.02408
Maximum Likelihood error	0.2473	0.3268	0.3455	0.2808
Average Posterior error	0.1106	0.1751	0.216	0.2477
Maximum Posterior error	0.8467	0.9624	0.987	0.9909
Average Random Likelihood error	0.05468	0.05329	0.03784	0.0237
Maximum Random Likelihood error	0.2425	0.2951	0.3646	0.3008
Average Random Posterior error	0.1094	0.1728	0.2176	0.2491
Maximum Random Posterior error	0.9255	0.9646	0.9787	0.9845
Average avg Full Likelihood error	0.05477	0.05317	0.03795	0.02395
Average avg Full Posterior error	0.1109	0.1749	0.2173	0.2483
Average avg All Likelihood error	0.02739	0.04024	0.04121	0.03636
Average avg All Posterior error	0.05546	0.09936	0.1313	0.1556

Table 53: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.5$ for $n > 5$

Prior 0.5	6	7	8	9	10
Avg Likelihood error	0.01385	0.008119	0.004421	0.00231	0.001234
Max Likelihood error	0.2783	0.2762	0.1887	0.2042	0.08357
Avg Posterior error	0.2686	0.2908	0.3051	0.3155	0.3288
Max Posterior error	0.9971	0.9982	0.9936	0.9984	0.9946
Avg Random Likelihood error	0.01374	0.008004	0.004409	0.002354	0.001231
Max Random Likelihood error	0.28	0.2068	0.1994	0.1123	0.09007
Avg Random Posterior error	0.2726	0.2856	0.3072	0.3162	0.3286
Max Random Posterior error	0.9986	0.998	0.995	0.995	0.9938

B.16. Fixed prior $\Omega = 0.6$

Table 54: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.6$ for $n \leq 5$

Prior 0.6	2	3	4	5
Average Likelihood error	0.05488	0.05358	0.03855	0.02426
Maximum Likelihood error	0.2441	0.3112	0.3267	0.293
Average Posterior error	0.1112	0.1724	0.217	0.2517
Maximum Posterior error	0.9293	0.9385	0.9867	0.9924
Average Random Likelihood error	0.05578	0.0523	0.03909	0.02397
Maximum Random Likelihood error	0.2441	0.2969	0.3792	0.2746
Average Random Posterior error	0.1103	0.1737	0.2169	0.246
Maximum Random Posterior error	0.9233	0.9699	0.9859	0.9892
Average avg Full Likelihood error	0.05514	0.05297	0.03821	0.02403
Average avg Full Posterior error	0.1108	0.1726	0.217	0.2471
Average avg All Likelihood error	0.02757	0.0401	0.04157	0.03637
Average avg All Posterior error	0.05542	0.09808	0.1307	0.154

Table 55: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.6$ for $n > 5$

Prior 0.6	6	7	8	9	10
Avg Likelihood error	0.01422	0.007935	0.004435	0.002305	0.001283
Max Likelihood error	0.2982	0.3202	0.2242	0.1543	0.1225
Avg Posterior error	0.2704	0.2851	0.3022	0.3187	0.3243
Max Posterior error	0.9827	0.9947	0.9976	0.9876	0.9991
Avg Random Likelihood error	0.01411	0.008068	0.004321	0.002337	0.001254
Max Random Likelihood error	0.2858	0.2874	0.1786	0.1253	0.1563
Avg Random Posterior error	0.2749	0.2896	0.3026	0.3126	0.3259
Max Random Posterior error	0.9942	0.9902	0.9944	0.9983	0.9992

B.17. Fixed prior $\Omega = 0.7$

Table 56: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.7$ for $n \leq 5$

Prior 0.7	2	3	4	5
Average Likelihood error	0.05552	0.05375	0.03861	0.02409
Maximum Likelihood error	0.24	0.3159	0.3242	0.3216
Average Posterior error	0.1066	0.1661	0.2105	0.2419
Maximum Posterior error	0.8876	0.9797	0.9692	0.9785
Average Random Likelihood error	0.05512	0.05268	0.03811	0.02385
Maximum Random Likelihood error	0.2392	0.3336	0.373	0.2956
Average Random Posterior error	0.1081	0.1707	0.2113	0.2409
Maximum Random Posterior error	0.8846	0.965	0.998	0.9893
Average avg Full Likelihood error	0.05551	0.05296	0.03803	0.02402
Average avg Full Posterior error	0.1065	0.1678	0.21	0.2403
Average avg All Likelihood error	0.02775	0.04006	0.04129	0.03641
Average avg All Posterior error	0.05325	0.0945	0.125	0.1477

Table 57: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.7$ for $n > 5$

Prior 0.7	6	7	8	9	10
Avg Likelihood error	0.0138	0.007834	0.004355	0.002299	0.001274
Max Likelihood error	0.2812	0.2126	0.2288	0.1045	0.1286
Avg Posterior error	0.267	0.2801	0.2973	0.3085	0.3156
Max Posterior error	0.9956	0.9951	0.996	0.9935	0.9962
Avg Random Likelihood error	0.01427	0.008075	0.004253	0.002325	0.001265
Max Random Likelihood error	0.2636	0.2461	0.1876	0.1981	0.1067
Avg Random Posterior error	0.2623	0.2829	0.2958	0.3081	0.3196
Max Random Posterior error	0.988	0.9894	0.9922	0.9954	0.9933

B.18. Fixed prior $\Omega = 0.8$

Table 58: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.8$ for $n \leq 5$

Prior 0.8	2	3	4	5
Average Likelihood error	0.05641	0.05347	0.03763	0.024
Maximum Likelihood error	0.2422	0.3116	0.3043	0.2966
Average Posterior error	0.09846	0.1527	0.1952	0.229
Maximum Posterior error	0.9192	0.9321	0.9959	0.9921
Average Random Likelihood error	0.05493	0.0526	0.0373	0.02432
Maximum Random Likelihood error	0.2379	0.2983	0.3324	0.2888
Average Random Posterior error	0.1009	0.1551	0.198	0.2268
Maximum Random Posterior error	0.867	0.95	0.9911	0.9912
Average avg Full Likelihood error	0.05573	0.05303	0.03802	0.02405
Average avg Full Posterior error	0.0997	0.1551	0.1968	0.2256
Average avg All Likelihood error	0.02786	0.04017	0.04137	0.03643
Average avg All Posterior error	0.04985	0.08635	0.1149	0.1348

Table 59: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.8$ for $n > 5$

Prior 0.8	6	7	8	9	10
Avg Likelihood error	0.01443	0.007994	0.004478	0.002409	0.001258
Max Likelihood error	0.288	0.2731	0.2302	0.1439	0.1334
Avg Posterior error	0.2455	0.2643	0.2797	0.2966	0.3066
Max Posterior error	0.9887	0.9955	0.9993	0.9989	0.9965
Avg Random Likelihood error	0.01376	0.007955	0.004349	0.002421	0.001256
Max Random Likelihood error	0.3185	0.2523	0.1972	0.2155	0.09583
Avg Random Posterior error	0.2483	0.2654	0.2828	0.2935	0.3041
Max Random Posterior error	0.9903	0.9952	0.9926	0.9981	0.998

B.19. Fixed prior $\Omega = 0.9$

Table 60: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.9$ for $n \leq 5$

Prior 0.9	2	3	4	5
Average Likelihood error	0.05593	0.05319	0.03791	0.02348
Maximum Likelihood error	0.2429	0.3184	0.3567	0.2715
Average Posterior error	0.08082	0.1335	0.164	0.1949
Maximum Posterior error	0.9254	0.9479	0.9736	0.9964
Average Random Likelihood error	0.05587	0.05265	0.03843	0.02387
Maximum Random Likelihood error	0.2362	0.2994	0.3076	0.2961
Average Random Posterior error	0.08129	0.1294	0.1674	0.1946
Maximum Random Posterior error	0.8754	0.966	0.9806	0.9904
Average avg Full Likelihood error	0.05574	0.05273	0.03801	0.02395
Average avg Full Posterior error	0.0817	0.13	0.1677	0.1944
Average avg All Likelihood error	0.02787	0.03998	0.0413	0.0363
Average avg All Posterior error	0.04085	0.0703	0.09306	0.1093

Table 61: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.9$ for $n > 5$

Prior 0.9	6	7	8	9	10
Avg Likelihood error	0.0139	0.007838	0.00412	0.002423	0.001308
Max Likelihood error	0.275	0.2266	0.1612	0.1169	0.1024
Avg Posterior error	0.2162	0.2325	0.2522	0.2615	0.2729
Max Posterior error	0.9919	0.9958	0.9957	0.9974	0.9992
Avg Random Likelihood error	0.01386	0.008081	0.004498	0.002374	0.001266
Max Random Likelihood error	0.2749	0.2576	0.1719	0.2058	0.1195
Avg Random Posterior error	0.2158	0.2379	0.2523	0.2598	0.2711
Max Random Posterior error	0.9903	0.9964	0.9962	0.996	0.9974

B.20. Fixed prior $\Omega = 0.999$

Table 62: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.999$ for $n \leq 5$

Prior 0.9	2	3	4	5
Average Likelihood error	0.05512	0.05342	0.03822	0.0237
Maximum Likelihood error	0.2446	0.2921	0.2908	0.3024
Average Posterior error	0.006937	0.01524	0.02185	0.02923
Maximum Posterior error	0.9914	0.9738	0.9928	0.9972
Average Random Likelihood error	0.05588	0.05349	0.03764	0.02391
Maximum Random Likelihood error	0.2438	0.3177	0.3032	0.3245
Average Random Posterior error	0.006891	0.01415	0.02254	0.02802
Maximum Random Posterior error	0.9914	0.9793	0.9956	0.992
Average avg Full Likelihood error	0.05537	0.05293	0.03809	0.02402
Average avg Full Posterior error	0.007309	0.01398	0.02138	0.02992
Average avg All Likelihood error	0.02769	0.04015	0.0414	0.03646
Average avg All Posterior error	0.003655	0.006088	0.008244	0.01031

Table 63: Average and Maximum errors found in networks with fixed prior, $\Omega = 0.999$ for $n > 5$

Prior 0.9	6	7	8	9	10
Avg Likelihood error	0.01428	0.008097	0.004472	0.002479	0.001253
Max Likelihood error	0.2617	0.2437	0.2771	0.1844	0.138
Avg Posterior error	0.03833	0.04956	0.05481	0.06391	0.07757
Max Posterior error	0.9965	0.9978	0.9992	0.9998	0.9994
Avg Random Likelihood error	0.01431	0.008172	0.004287	0.002301	0.001267
Max Random Likelihood error	0.3013	0.2874	0.1475	0.138	0.07232
Avg Random Posterior error	0.04025	0.04806	0.0549	0.06679	0.07667
Max Random Posterior error	0.9949	0.9975	0.9998	0.9989	0.9993

C. Correlation

In this section all correlations found for completely random networks will be displayed, and all correlations found for forced dependency and forced determinism networks for $n = 5$.

C.1. $n = 2$

Table 64: Correlation between errors for completely random networks with $n = 2$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.2473	0.4907	0.2355
Posterior Error	0.2473	1	0.2543	0.3896
Random Likelihood Error	0.4907	0.2543	1	0.2329
Random Posterior Error	0.2355	0.3896	0.2329	1
Avg full Likelihood Error	0.7047	0.3539	0.7116	0.3415
Avg full Posterior Error	0.3875	0.6249	0.3969	0.6303
Avg all Likelihood Error	0.7047	0.3539	0.7116	0.3415
Avg all Posterior Error	0.3875	0.6249	0.3969	0.6303

Table 65: Correlation between errors and average errors for completely random networks with $n = 2$

All Random	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.7047	0.3875	0.7047	0.3875
Posterior Error	0.3539	0.6249	0.3539	0.6249
Random Likelihood Error	0.7116	0.3969	0.7116	0.3969
Random Posterior Error	0.3415	0.6303	0.3415	0.6303
Avg full Likelihood Error	1	0.5568	1	0.5568
Avg full Posterior Error	0.5568	1	0.5568	1
Avg all Likelihood Error	1	0.5568	1	0.5568
Avg all Posterior Error	0.5568	1	0.5568	1

Table 66: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 2$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	0.3098	0.2739	0.3627	0.3953
Posterior Error	0.2855	0.1319	0.1439	0.1355
Random Likelihood Error	0.3217	0.2802	0.383	0.4098
Random Posterior Error	0.1118	0.116	0.1207	0.1105
Avg full Likelihood Error	0.4465	0.3977	0.5235	0.5662
Avg full Posterior Error	0.1828	0.1931	0.2071	0.19
Avg all Likelihood Error	0.4465	0.3977	0.5235	0.5662
Avg all Posterior Error	0.1828	0.1931	0.2071	0.19

Table 67: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 2$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.1708	-0.2414	-0.2103	-0.1437
Posterior Error	0.07347	-0.07622	-0.03351	0.02544
Random Likelihood Error	-0.1743	-0.2286	-0.2052	-0.1346
Random Posterior Error	-0.01634	-0.0789	-0.03427	0.02242
Avg full Likelihood Error	-0.2356	-0.3269	-0.2899	-0.1973
Avg full Posterior Error	-0.01589	-0.1251	-0.05504	0.03795
Avg all Likelihood Error	-0.2356	-0.3269	-0.2899	-0.1973
Avg all Posterior Error	-0.01589	-0.1251	-0.05504	0.03795

Table 68: Correlation between errors and multivariate MI for completely random networks with $n = 2$

All Random	Cond TC	Cond II	TC	II
Likelihood Error	0.5607	0.5607	0.1341	-0.332
Posterior Error	0.3899	0.3899	0.1962	-0.3809
Random Likelihood Error	0.585	0.585	0.1529	-0.3401
Random Posterior Error	0.3716	0.3716	0.1849	-0.3859
Avg full Likelihood Error	0.809	0.809	0.2058	-0.4788
Avg full Posterior Error	0.5999	0.5999	0.2987	-0.606
Avg all Likelihood Error	0.809	0.809	0.2058	-0.4788
Avg all Posterior Error	0.5999	0.5999	0.2987	-0.606

C.2. $n = 3$

Table 69: Correlation between errors for completely random networks with $n = 3$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.1329	0.1521	0.07326
Posterior Error	0.1329	1	0.06808	0.1864
Random Likelihood Error	0.1521	0.06808	1	0.06941
Random Posterior Error	0.07326	0.1864	0.06941	1
Avg full Likelihood Error	0.3932	0.193	0.3897	0.1846
Avg full Posterior Error	0.1488	0.4463	0.1577	0.4544
Avg all Likelihood Error	0.3717	0.1948	0.3792	0.1894
Avg all Posterior Error	0.1647	0.4313	0.1663	0.4367

Table 70: Correlation between errors and average errors for completely random networks with $n = 3$

All Random	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.3932	0.1488	0.3717	0.1647
Posterior Error	0.193	0.4463	0.1948	0.4313
Random Likelihood Error	0.3897	0.1577	0.3792	0.1663
Random Posterior Error	0.1846	0.4544	0.1894	0.4367
Avg full Likelihood Error	1	0.4106	0.9555	0.4441
Avg full Posterior Error	0.4106	1	0.4135	0.9642
Avg all Likelihood Error	0.9555	0.4135	1	0.468
Avg all Posterior Error	0.4441	0.9642	0.468	1

Table 71: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 3$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	0.04158	0.113	0.174	0.195
Posterior Error	0.2137	0.07687	0.09017	0.07175
Random Likelihood Error	0.1316	0.1114	0.166	0.1913
Random Posterior Error	0.06672	0.07275	0.0741	0.06442
Avg full Likelihood Error	0.3474	0.292	0.4399	0.4937
Avg full Posterior Error	0.1452	0.1471	0.1616	0.1328
Avg all Likelihood Error	0.3961	0.3442	0.486	0.5486
Avg all Posterior Error	0.1584	0.1619	0.1716	0.1484

Table 72: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 3$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.13	-0.1388	-0.1103	-0.06829
Posterior Error	0.08225	-0.04028	0.02033	0.08341
Random Likelihood Error	-0.1083	-0.1402	-0.13	-0.08797
Random Posterior Error	0.01109	-0.04197	0.005093	0.07228
Avg full Likelihood Error	-0.2386	-0.3471	-0.294	-0.1914
Avg full Posterior Error	0.02148	-0.1195	0.007192	0.1579
Avg all Likelihood Error	-0.2473	-0.3575	-0.3059	-0.1984
Avg all Posterior Error	0.04352	-0.1048	0.03993	0.1932

Table 73: Correlation between errors and multivariate MI for completely random networks with $n = 3$

All Random	Cond TC	Cond II	TC	II
Likelihood Error	0.3117	-0.001986	0.1226	-0.06715
Posterior Error	0.217	0.0508	0.1684	-0.01559
Random Likelihood Error	0.3074	0.007307	0.1027	-0.06691
Random Posterior Error	0.2121	0.03709	0.1519	-0.006962
Avg full Likelihood Error	0.8057	0.01087	0.3093	-0.1855
Avg full Posterior Error	0.4697	0.07898	0.3331	-0.01163
Avg all Likelihood Error	0.8071	0.1933	0.3002	-0.2546
Avg all Posterior Error	0.5011	0.1904	0.3829	-0.1862

C.3. $n = 4$

Table 74: Correlation between errors for completely random networks with $n = 4$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.053	0.03812	0.02328
Posterior Error	0.053	1	0.033	0.13
Random Likelihood Error	0.03812	0.033	1	0.01024
Random Posterior Error	0.02328	0.13	0.01024	1
Avg full Likelihood Error	0.2246	0.1087	0.213	0.1241
Avg full Posterior Error	0.06273	0.3625	0.06768	0.3703
Avg all Likelihood Error	0.2109	0.1072	0.2001	0.1265
Avg all Posterior Error	0.07343	0.3446	0.07719	0.3523

Table 75: Correlation between errors and average errors for completely random networks with $n = 4$

All Random	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.2246	0.06273	0.2109	0.07343
Posterior Error	0.1087	0.3625	0.1072	0.3446
Random Likelihood Error	0.213	0.06768	0.2001	0.07719
Random Posterior Error	0.1241	0.3703	0.1265	0.3523
Avg full Likelihood Error	1	0.3221	0.9303	0.3537
Avg full Posterior Error	0.3221	1	0.3249	0.9498
Avg all Likelihood Error	0.9303	0.3249	1	0.3877
Avg all Posterior Error	0.3537	0.9498	0.3877	1

Table 76: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 4$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.0702	0.0539	0.08694	0.1083
Posterior Error	0.1374	0.02671	0.02279	0.01552
Random Likelihood Error	0.06236	0.04406	0.07539	0.09152
Random Posterior Error	0.04878	0.05286	0.0507	0.04566
Avg full Likelihood Error	0.2884	0.2214	0.3676	0.4316
Avg full Posterior Error	0.07208	0.07731	0.07679	0.06067
Avg all Likelihood Error	0.3711	0.3163	0.4506	0.5248
Avg all Posterior Error	0.09343	0.1011	0.09255	0.08371

Table 77: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 4$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.09031	-0.07828	-0.05429	-0.03231
Posterior Error	0.06048	-0.0445	0.005511	0.07357
Random Likelihood Error	-0.03647	-0.06868	-0.05907	-0.03654
Random Posterior Error	0.03101	-0.03928	0.01192	0.07436
Avg full Likelihood Error	-0.2328	-0.3411	-0.2952	-0.1929
Avg full Posterior Error	0.04906	-0.1216	0.01523	0.2034
Avg all Likelihood Error	-0.2324	-0.3418	-0.2935	-0.1881
Avg all Posterior Error	0.1053	-0.08074	0.09257	0.2829

Table 78: Correlation between errors and multivariate MI for completely random networks with $n = 4$

All Random	Cond TC	Cond II	TC	II
Likelihood Error	0.1866	-0.03978	0.09187	0.02312
Posterior Error	0.1182	-0.0432	0.0919	0.0602
Random Likelihood Error	0.1726	-0.04881	0.07752	0.03276
Random Posterior Error	0.1441	-0.02281	0.1164	0.02478
Avg full Likelihood Error	0.815	-0.1865	0.3522	0.1089
Avg full Posterior Error	0.362	-0.09201	0.2801	0.1111
Avg all Likelihood Error	0.8126	-0.1338	0.3519	0.03113
Avg all Posterior Error	0.3937	-0.09604	0.3688	0.1173

C.4. $n = 5$

Table 79: Correlation between errors for completely random networks with $n = 5$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.03033	0.01656	0.01232
Posterior Error	0.03033	1	0.001215	0.1219
Random Likelihood Error	0.01656	0.001215	1	0.0008565
Random Posterior Error	0.01232	0.1219	0.0008565	1
Avg full Likelihood Error	0.1473	0.09378	0.1236	0.08342
Avg full Posterior Error	0.05236	0.3233	0.02482	0.3329
Avg all Likelihood Error	0.1358	0.09204	0.1154	0.08076
Avg all Posterior Error	0.05789	0.304	0.02974	0.3206

Table 80: Correlation between errors and average errors for completely random networks with $n = 5$

All Random	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.1473	0.05236	0.1358	0.05789
Posterior Error	0.09378	0.3233	0.09204	0.304
Random Likelihood Error	0.1236	0.02482	0.1154	0.02974
Random Posterior Error	0.08342	0.3329	0.08076	0.3206
Avg full Likelihood Error	1	0.2808	0.9136	0.304
Avg full Posterior Error	0.2808	1	0.2755	0.9463
Avg all Likelihood Error	0.9136	0.2755	1	0.3323
Avg all Posterior Error	0.304	0.9463	0.3323	1

Table 81: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 5$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1028	0.04382	0.06337	0.06895
Posterior Error	0.09787	0.008088	0.002898	0.005412
Random Likelihood Error	0.03464	0.02366	0.04384	0.05138
Random Posterior Error	0.0006429	0.002902	0.006099	-0.0004764
Avg full Likelihood Error	0.261	0.1921	0.3376	0.4116
Avg full Posterior Error	-0.002488	0.004491	0.01063	0.0005879
Avg all Likelihood Error	0.3763	0.3254	0.4567	0.5326
Avg all Posterior Error	0.01023	0.02262	0.02314	0.01764

Table 82: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 5$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.08637	-0.05373	-0.05485	-0.03289
Posterior Error	0.03957	-0.05209	0.009678	0.0917
Random Likelihood Error	-0.02313	-0.03687	-0.03414	-0.01673
Random Posterior Error	0.04383	-0.03591	0.02506	0.1029
Avg full Likelihood Error	-0.2379	-0.3183	-0.2978	-0.1771
Avg full Posterior Error	0.04771	-0.1384	0.04578	0.2783
Avg all Likelihood Error	-0.2268	-0.2981	-0.2803	-0.168
Avg all Posterior Error	0.1139	-0.08626	0.1385	0.3755

Table 83: Correlation between errors and multivariate MI for completely random networks with $n = 5$

All Random	Cond TC	Cond II	TC	II
Likelihood Error	0.1344	0.007505	0.05846	0.0006535
Posterior Error	0.1022	-0.00391	0.08703	0.01337
Random Likelihood Error	0.106	-0.02372	0.05362	0.01005
Random Posterior Error	0.09486	-0.00196	0.0941	-0.02859
Avg full Likelihood Error	0.8184	0.0275	0.3859	0.05481
Avg full Posterior Error	0.2903	-0.003732	0.2624	-0.03479
Avg all Likelihood Error	0.8206	-0.05733	0.4021	0.08558
Avg all Posterior Error	0.3172	-0.02669	0.36	0.003336

C.5. $n = 6$

Table 84: Correlation between errors for completely random networks with $n = 6$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.02398	0.0136	-0.002092
Posterior Error	0.02398	1	-0.007682	0.08102
Random Likelihood Error	0.0136	-0.007682	1	0.01349
Random Posterior Error	-0.002092	0.08102	0.01349	1

Table 85: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 6$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1152	0.007337	0.002497	0.008635
Posterior Error	0.06548	-0.01242	-0.006669	-0.004738
Random Likelihood Error	0.03515	0.02496	0.04136	0.05695
Random Posterior Error	-0.004552	-0.01364	-0.00356	-0.002643

Table 86: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 6$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.03195	-0.02833	-0.01792	-0.01518
Posterior Error	0.03251	-0.04844	0.01976	0.09661
Random Likelihood Error	-0.01372	-0.0251	-0.01837	-0.009957
Random Posterior Error	0.01645	-0.03475	0.01537	0.08781

C.6. $n = 7$

Table 87: Correlation between errors for completely random networks with $n = 7$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.007662	-0.01038	0.006109
Posterior Error	-0.007662	1	0.001825	0.064
Random Likelihood Error	-0.01038	0.001825	1	0.01122
Random Posterior Error	0.006109	0.064	0.01122	1

Table 88: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 7$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1042	0.02064	0.02227	0.03431
Posterior Error	0.05101	-0.01471	-0.005407	-0.0133
Random Likelihood Error	-0.0003622	-0.001491	0.001792	0.008522
Random Posterior Error	-0.005083	0.000871	-0.001017	-0.002431

Table 89: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 7$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.04188	0.002682	0.002082	0.008558
Posterior Error	0.05533	-0.03376	0.02324	0.08624
Random Likelihood Error	-0.007611	-0.02051	0.002719	0.001541
Random Posterior Error	0.01757	-0.03641	0.01047	0.08039

C.7. $n = 8$

Table 90: Correlation between errors for completely random networks with $n = 8$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.01457	0.009846	0.01526
Posterior Error	-0.01457	1	-0.009211	0.08173
Random Likelihood Error	0.009846	-0.009211	1	-0.003683
Random Posterior Error	0.01526	0.08173	-0.003683	1

Table 91: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 8$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.09021	0.01328	0.007357	0.01919
Posterior Error	0.05415	-0.01735	-0.01301	-0.01124
Random Likelihood Error	-0.01525	-0.0179	-0.008017	-0.004394
Random Posterior Error	-0.01291	-0.01076	-0.007233	-0.01392

Table 92: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 8$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.03118	-0.0006961	-0.0008601	0.007117
Posterior Error	0.06496	-0.02864	0.03441	0.1017
Random Likelihood Error	-0.016	-0.0169	-0.01293	-0.003775
Random Posterior Error	0.01758	-0.03523	0.005461	0.07424

C.8. $n = 9$

Table 93: Correlation between errors for completely random networks with $n = 9$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.001089	0.01057	0.003644
Posterior Error	0.001089	1	-0.004748	0.04878
Random Likelihood Error	0.01057	-0.004748	1	-0.001938
Random Posterior Error	0.003644	0.04878	-0.001938	1

Table 94: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 9$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.08259	-0.002605	0.00112	0.005097
Posterior Error	0.06124	0.005131	0.01306	0.007114
Random Likelihood Error	-0.01815	-0.01728	-0.01321	-0.01042
Random Posterior Error	-0.01684	-0.02316	-0.006055	-0.009694

Table 95: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 9$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.01191	-0.005495	-0.006012	-0.006181
Posterior Error	0.03638	-0.03543	0.01513	0.07503
Random Likelihood Error	-0.007745	-0.01326	-0.008398	-0.01022
Random Posterior Error	-0.000283	-0.04877	0.01468	0.07722

C.9. $n = 10$

Table 96: Correlation between errors for completely random networks with $n = 10$

All Random	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.004858	0.003604	0.0007509
Posterior Error	-0.004858	1	-0.008497	0.05601
Random Likelihood Error	0.003604	-0.008497	1	-0.008635
Random Posterior Error	0.0007509	0.05601	-0.008635	1

Table 97: Correlation between errors and pairwise dependency classifiers between evidence variables for completely random networks with $n = 10$

All Random	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.08333	0.0003579	0.003633	0.006816
Posterior Error	0.02091	-0.03401	-0.03496	-0.03978
Random Likelihood Error	0.008098	-0.002561	-0.002596	0.0007068
Random Posterior Error	-0.01561	-0.01953	-0.01707	-0.02444

Table 98: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for completely random networks with $n = 10$

All Random	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.02743	-0.007777	-0.008313	0.001066
Posterior Error	0.04887	-0.04705	0.01166	0.07379
Random Likelihood Error	-0.01132	-0.0109	-0.00464	-0.002832
Random Posterior Error	0.02937	-0.02222	0.009502	0.0773

C.10. Dependency $\rho = 0.5$, $n = 5$

Table 99: Correlation between errors for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.0877	0.0453	0.03431
Posterior Error	-0.0877	1	-0.02944	0.07092
Random Likelihood Error	0.0453	-0.02944	1	0.009822
Random Posterior Error	0.03431	0.07092	0.009822	1
Avg full Likelihood Error	0.3539	-0.04695	0.1178	0.07196
Avg full Posterior Error	0.09849	0.2322	0.01625	0.3165
Avg all Likelihood Error	0.4161	-0.008528	0.1121	0.06724
Avg all Posterior Error	0.09968	0.27	0.01312	0.2976

Table 100: Correlation between errors and average errors for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.3539	0.09849	0.4161	0.09968
Posterior Error	-0.04695	0.2322	-0.008528	0.27
Random Likelihood Error	0.1178	0.01625	0.1121	0.01312
Random Posterior Error	0.07196	0.3165	0.06724	0.2976
Avg full Likelihood Error	1	0.232	0.9098	0.2221
Avg full Posterior Error	0.232	1	0.2219	0.9417
Avg all Likelihood Error	0.9098	0.2219	1	0.2252
Avg all Posterior Error	0.2221	0.9417	0.2252	1

Table 101: Correlation between errors and pairwise dependency classifiers between evidence variables for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.02547	0.3077	0.3141	0.3389
Posterior Error	0.3704	0.1173	0.07284	0.0546
Random Likelihood Error	0.03458	0.03689	0.04841	0.0556
Random Posterior Error	0.03661	0.07262	0.05598	0.05148
Avg full Likelihood Error	0.301	0.3328	0.4697	0.5208
Avg full Posterior Error	0.1514	0.2257	0.1997	0.1897
Avg all Likelihood Error	0.4383	0.533	0.6322	0.6861
Avg all Posterior Error	0.1872	0.2783	0.2456	0.2327

Table 102: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.1494	-0.0995	-0.1137	-0.06712
Posterior Error	0.2316	0.2099	0.08836	0.158
Random Likelihood Error	-0.03282	-0.04027	-0.0406	-0.03066
Random Posterior Error	0.01641	-0.03306	0.01006	0.07852
Avg full Likelihood Error	-0.2828	-0.3808	-0.331	-0.2371
Avg full Posterior Error	0.1078	-0.04342	0.075	0.2794
Avg all Likelihood Error	-0.2532	-0.34	-0.3096	-0.2101
Avg all Posterior Error	0.2064	0.0478	0.1973	0.405

Table 103: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.5$ and $n = 5$

Dependency 0.5	Cond TC	Cond II	TC	II
Likelihood Error	0.3618	0.18	0.2106	0.002373
Posterior Error	0.01514	0.02197	0.0825	-0.03895
Random Likelihood Error	0.08194	0.02576	0.03596	0.01533
Random Posterior Error	0.07696	0.005957	0.07191	-0.01474
Avg full Likelihood Error	0.8041	0.2564	0.4062	0.0003541
Avg full Posterior Error	0.2635	0.05396	0.2783	-0.02181
Avg all Likelihood Error	0.8311	0.2563	0.4456	-0.02457
Avg all Posterior Error	0.2714	0.03941	0.3817	0.0335

C.11. Dependency $\rho = 0.9$, $n = 5$

Table 104: Correlation between errors for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.02699	0.05289	-0.01382
Posterior Error	-0.02699	1	-0.001765	0.06635
Random Likelihood Error	0.05289	-0.001765	1	0.0004196
Random Posterior Error	-0.01382	0.06635	0.0004196	1
Avg full Likelihood Error	0.5255	0.04908	0.1101	-0.02051
Avg full Posterior Error	-0.02571	0.1801	-0.004214	0.3301
Avg all Likelihood Error	0.5244	0.04513	0.1098	-0.02665
Avg all Posterior Error	-0.08682	0.2001	-0.01894	0.3124

Table 105: Correlation between errors and average errors for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.5255	-0.02571	0.5244	-0.08682
Posterior Error	0.04908	0.1801	0.04513	0.2001
Random Likelihood Error	0.1101	-0.004214	0.1098	-0.01894
Random Posterior Error	-0.02051	0.3301	-0.02665	0.3124
Avg full Likelihood Error	1	-0.03109	0.9928	-0.1523
Avg full Posterior Error	-0.03109	1	-0.04551	0.937
Avg all Likelihood Error	0.9928	-0.04551	1	-0.1667
Avg all Posterior Error	-0.1523	0.937	-0.1667	1

Table 106: Correlation between errors and pairwise dependency classifiers between evidence variables for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.3102	0.1944	0.306	0.3153
Posterior Error	0.4078	0.03267	0.01398	0.01167
Random Likelihood Error	0.02597	0.04175	0.0664	0.06682
Random Posterior Error	0.009048	0.04495	-0.009181	-0.02769
Avg full Likelihood Error	0.1292	0.3644	0.578	0.596
Avg full Posterior Error	0.01566	0.1398	0.01146	-0.03488
Avg all Likelihood Error	0.1295	0.365	0.5834	0.6021
Avg all Posterior Error	0.02135	0.1609	0.03761	-0.009732

Table 107: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.2851	-0.2779	-0.2786	-0.197
Posterior Error	0.2026	0.2576	-0.01424	0.07922
Random Likelihood Error	-0.0517	-0.06756	-0.068	-0.05151
Random Posterior Error	0.1098	0.09011	0.09533	0.1342
Avg full Likelihood Error	-0.4295	-0.5254	-0.5138	-0.3715
Avg full Posterior Error	0.2994	0.2274	0.2726	0.4057
Avg all Likelihood Error	-0.4307	-0.5248	-0.5133	-0.3755
Avg all Posterior Error	0.467	0.4236	0.4534	0.5922

Table 108: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.9$ and $n = 5$

Dependency 0.9	Cond TC	Cond II	TC	II
Likelihood Error	0.4587	0.4398	0.2225	-0.253
Posterior Error	0.02006	0.01703	0.007924	-0.004893
Random Likelihood Error	0.1039	0.0999	0.04607	-0.06044
Random Posterior Error	-0.07141	-0.0594	0.01103	0.09486
Avg full Likelihood Error	0.8633	0.8231	0.4282	-0.4661
Avg full Posterior Error	-0.1682	-0.1406	0.06821	0.2727
Avg all Likelihood Error	0.8639	0.8246	0.4292	-0.4712
Avg all Posterior Error	-0.2566	-0.2145	0.1371	0.4406

C.12. Dependency $\rho = 0.999$, $n = 5$

Table 109: Correlation between errors for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.02313	0.07241	-0.03563
Posterior Error	-0.02313	1	0.003925	0.06102
Random Likelihood Error	0.07241	0.003925	1	-0.01431
Random Posterior Error	-0.03563	0.06102	-0.01431	1
Avg full Likelihood Error	0.5562	0.08035	0.1398	-0.08163
Avg full Posterior Error	-0.1023	0.1265	-0.03891	0.353
Avg all Likelihood Error	0.5545	0.0733	0.1398	-0.08244
Avg all Posterior Error	-0.1844	0.1133	-0.0537	0.3337

Table 110: Correlation between errors and average errors for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.5562	-0.1023	0.5545	-0.1844
Posterior Error	0.08035	0.1265	0.0733	0.1133
Random Likelihood Error	0.1398	-0.03891	0.1398	-0.0537
Random Posterior Error	-0.08163	0.353	-0.08244	0.3337
Avg full Likelihood Error	1	-0.2172	0.9971	-0.3606
Avg full Posterior Error	-0.2172	1	-0.2183	0.9405
Avg all Likelihood Error	0.9971	-0.2183	1	-0.3562
Avg all Posterior Error	-0.3606	0.9405	-0.3562	1

Table 111: Correlation between errors and pairwise dependency classifiers between evidence variables for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.2971	0.1544	0.3359	0.3336
Posterior Error	0.4105	-0.01167	-0.001758	-0.001662
Random Likelihood Error	0.004741	0.03997	0.09379	0.09339
Random Posterior Error	0.004374	-0.04239	-0.09824	-0.09688
Avg full Likelihood Error	-0.001764	0.2744	0.6028	0.5997
Avg full Posterior Error	-0.01115	-0.1151	-0.2754	-0.2745
Avg all Likelihood Error	-0.001286	0.264	0.6031	0.6021
Avg all Posterior Error	-0.01243	-0.125	-0.2891	-0.2878

Table 112: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.282	-0.3125	-0.2796	-0.1903
Posterior Error	0.1439	0.1895	-0.07051	0.01952
Random Likelihood Error	-0.05148	-0.08223	-0.06453	-0.04114
Random Posterior Error	0.144	0.1287	0.1232	0.1517
Avg full Likelihood Error	-0.4459	-0.5668	-0.5277	-0.3649
Avg full Posterior Error	0.3606	0.3232	0.3378	0.4091
Avg all Likelihood Error	-0.4423	-0.5615	-0.5237	-0.3663
Avg all Posterior Error	0.5093	0.5111	0.4932	0.556

Table 113: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 0.999$ and $n = 5$

Dependency 0.999	Cond TC	Cond II	TC	II
Likelihood Error	0.4723	0.4723	0.2747	-0.2791
Posterior Error	0.04249	0.0421	-0.01359	-0.07012
Random Likelihood Error	0.1234	0.1234	0.07905	-0.06442
Random Posterior Error	-0.164	-0.1639	-0.07336	0.1231
Avg full Likelihood Error	0.8636	0.8635	0.4886	-0.5267
Avg full Posterior Error	-0.4551	-0.4545	-0.207	0.3373
Avg all Likelihood Error	0.8614	0.8614	0.4896	-0.5228
Avg all Posterior Error	-0.564	-0.5633	-0.1936	0.4923

C.13. Dependency $\rho = 1$, $n = 5$

Table 114: Correlation between errors for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.03643	0.07262	-0.01553
Posterior Error	-0.03643	1	0.002774	-0.002
Random Likelihood Error	0.07262	0.002774	1	-0.003082
Random Posterior Error	-0.01553	-0.002	-0.003082	1
Avg full Likelihood Error	0.574	0.06504	0.1242	0.003199
Avg full Posterior Error	0.0756	0.5298	0.02669	0.01581
Avg all Likelihood Error	0.5724	0.05836	0.1238	0.002969
Avg all Posterior Error	-0.02869	0.475	0.004076	0.008718

Table 115: Correlation between errors and average errors for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.574	0.0756	0.5724	-0.02869
Posterior Error	0.06504	0.5298	0.05836	0.475
Random Likelihood Error	0.1242	0.02669	0.1238	0.004076
Random Posterior Error	0.003199	0.01581	0.002969	0.008718
Avg full Likelihood Error	1	0.1287	0.9971	-0.05344
Avg full Posterior Error	0.1287	1	0.1166	0.8931
Avg all Likelihood Error	0.9971	0.1166	1	-0.06865
Avg all Posterior Error	-0.05344	0.8931	-0.06865	1

Table 116: Correlation between errors and pairwise dependency classifiers between evidence variables for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.3044	-	0.3498	0.3469
Posterior Error	0.427	-	-0.005613	-0.004869
Random Likelihood Error	-0.006781	-	0.07175	0.07059
Random Posterior Error	0.005606	-	0.007984	0.008278
Avg full Likelihood Error	-0.01023	-	0.6172	0.6129
Avg full Posterior Error	-0.0063	-	0.003759	0.003746
Avg all Likelihood Error	-0.009728	-	0.6182	0.616
Avg all Posterior Error	-0.008312	-	-0.08393	-0.08298

Table 117: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.2916	-0.3186	-0.2902	-0.1954
Posterior Error	0.1628	0.2014	-0.04999	0.04081
Random Likelihood Error	-0.049	-0.05962	-0.05784	-0.03648
Random Posterior Error	-0.003807	0.003929	-0.002354	0.005588
Avg full Likelihood Error	-0.4131	-0.5525	-0.5	-0.3322
Avg full Posterior Error	0.1504	0.3895	-0.09187	0.07836
Avg all Likelihood Error	-0.4115	-0.5475	-0.4971	-0.3346
Avg all Posterior Error	0.3691	0.5661	0.1541	0.314

Table 118: Correlation between errors and multivariate mutual information for forced dependency networks with $\rho = 1$ and $n = 5$

Dependency 1	Cond TC	Cond II	TC	II
Likelihood Error	0.4935	0.4935	0.2885	-0.2902
Posterior Error	0.02549	0.02549	-0.0135	-0.04999
Random Likelihood Error	0.1002	0.1002	0.05944	-0.05784
Random Posterior Error	0.008652	0.008652	0.007277	-0.002354
Avg full Likelihood Error	0.8633	0.8633	0.5109	-0.5
Avg full Posterior Error	0.05957	0.05957	-0.01131	-0.09187
Avg all Likelihood Error	0.8624	0.8624	0.5123	-0.4971
Avg all Posterior Error	-0.17	-0.17	-0.05549	0.1541

C.14. Determinism $\rho = 0.5$, $n = 5$

Table 119: Correlation between errors for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.07038	0.002554	-0.00112
Posterior Error	0.07038	1	0.02514	0.07204
Random Likelihood Error	0.002554	0.02514	1	0.001858
Random Posterior Error	-0.00112	0.07204	0.001858	1
Avg full Likelihood Error	0.1292	0.09381	0.14	0.08226
Avg full Posterior Error	0.03389	0.2963	0.04805	0.2939
Avg all Likelihood Error	0.1213	0.09119	0.1315	0.08209
Avg all Posterior Error	0.03831	0.2802	0.04747	0.2778

Table 120: Correlation between errors and average errors for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.1292	0.03389	0.1213	0.03831
Posterior Error	0.09381	0.2963	0.09119	0.2802
Random Likelihood Error	0.14	0.04805	0.1315	0.04747
Random Posterior Error	0.08226	0.2939	0.08209	0.2778
Avg full Likelihood Error	1	0.284	0.9156	0.3076
Avg full Posterior Error	0.284	1	0.2729	0.9424
Avg all Likelihood Error	0.9156	0.2729	1	0.3377
Avg all Posterior Error	0.3076	0.9424	0.3377	1

Table 121: Correlation between errors and pairwise dependency classifiers between evidence variables for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1132	0.0266	0.04264	0.05383
Posterior Error	0.08482	0.001876	-0.007387	-0.004769
Random Likelihood Error	0.02696	0.01808	0.03997	0.04652
Random Posterior Error	0.005059	0.005175	0.00796	-0.001424
Avg full Likelihood Error	0.2771	0.2113	0.3571	0.4247
Avg full Posterior Error	-0.00967	-0.02133	-0.00997	-0.02222
Avg all Likelihood Error	0.3784	0.3275	0.4647	0.5324
Avg all Posterior Error	0.008458	0.0006146	0.004314	-0.004169

Table 122: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.06345	-0.03579	-0.0415	-0.02301
Posterior Error	0.03537	-0.06027	-0.007175	0.06121
Random Likelihood Error	-0.03589	-0.06164	-0.05667	-0.0342
Random Posterior Error	0.02148	-0.03613	0.01219	0.07528
Avg full Likelihood Error	-0.2259	-0.3257	-0.2868	-0.1802
Avg full Posterior Error	0.03893	-0.1558	0.03432	0.2663
Avg all Likelihood Error	-0.2133	-0.2988	-0.2718	-0.1685
Avg all Posterior Error	0.1059	-0.1066	0.1261	0.3592

Table 123: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.5$ and $n = 5$

Determinism 0.5	Cond TC	Cond II	TC	II
Likelihood Error	0.1094	0.0142	0.05015	-0.01175
Posterior Error	0.09656	-0.00324	0.06858	-0.02066
Random Likelihood Error	0.1129	-0.0007085	0.04028	0.01354
Random Posterior Error	0.08673	0.00245	0.07699	-0.02585
Avg full Likelihood Error	0.8229	0.03062	0.398	0.02612
Avg full Posterior Error	0.284	0.00985	0.2475	-0.05964
Avg all Likelihood Error	0.8247	-0.0571	0.4119	0.06839
Avg all Posterior Error	0.314	-0.03294	0.3464	-0.01306

C.15. Determinism $\rho = 0.9$, $n = 5$

Table 124: Correlation between errors for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	0.05074	0.008003	0.01601
Posterior Error	0.05074	1	0.01374	0.06342
Random Likelihood Error	0.008003	0.01374	1	0.008652
Random Posterior Error	0.01601	0.06342	0.008652	1
Avg full Likelihood Error	0.09461	0.07472	0.07915	0.05635
Avg full Posterior Error	0.007694	0.2496	0.01225	0.2449
Avg all Likelihood Error	0.07562	0.05604	0.06052	0.0505
Avg all Posterior Error	0.009087	0.2247	0.01128	0.2189

Table 125: Correlation between errors and average errors for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.09461	0.007694	0.07562	0.009087
Posterior Error	0.07472	0.2496	0.05604	0.2247
Random Likelihood Error	0.07915	0.01225	0.06052	0.01128
Random Posterior Error	0.05635	0.2449	0.0505	0.2189
Avg full Likelihood Error	1	0.2503	0.8313	0.248
Avg full Posterior Error	0.2503	1	0.2096	0.9003
Avg all Likelihood Error	0.8313	0.2096	1	0.3146
Avg all Posterior Error	0.248	0.9003	0.3146	1

Table 126: Correlation between errors and pairwise dependency classifiers between evidence variables for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1434	-0.005121	0.01051	0.02191
Posterior Error	0.03805	-0.03515	-0.03241	-0.021
Random Likelihood Error	0.01532	0.003737	0.01408	0.0243
Random Posterior Error	-0.01628	-0.02996	-0.0241	-0.02043
Avg full Likelihood Error	0.07181	-0.02357	0.1396	0.2381
Avg full Posterior Error	-0.1484	-0.173	-0.1775	-0.1348
Avg all Likelihood Error	0.2784	0.2097	0.3602	0.4476
Avg all Posterior Error	-0.1039	-0.116	-0.1339	-0.09754

Table 127: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.06566	-0.05443	-0.04136	-0.03071
Posterior Error	0.0279	-0.06189	0.005824	0.06993
Random Likelihood Error	-0.03547	-0.03175	-0.04499	-0.026
Random Posterior Error	0.01533	-0.04606	0.02394	0.07579
Avg full Likelihood Error	-0.3626	-0.5312	-0.4402	-0.298
Avg full Posterior Error	0.008578	-0.2608	0.02013	0.2601
Avg all Likelihood Error	-0.289	-0.4262	-0.3623	-0.2415
Avg all Posterior Error	0.08727	-0.188	0.1214	0.3591

Table 128: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.9$ and $n = 5$

Determinism 0.9	Cond TC	Cond II	TC	II
Likelihood Error	0.0687	0.001226	0.02083	-0.003226
Posterior Error	0.06108	-0.004142	0.05741	-0.01214
Random Likelihood Error	0.07104	0.02917	0.01951	-0.02491
Random Posterior Error	0.04462	-0.002887	0.05992	-0.01151
Avg full Likelihood Error	0.8063	0.05408	0.2857	-0.0085
Avg full Posterior Error	0.1934	0.01148	0.1833	-0.08068
Avg all Likelihood Error	0.8109	-0.07125	0.3608	0.07105
Avg all Posterior Error	0.2316	-0.06049	0.3084	-0.002749

C.16. Determinism $\rho = 0.999$, $n = 5$

Table 129: Correlation between errors for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.003943	0.02258	0.007083
Posterior Error	-0.003943	1	0.0007798	0.01845
Random Likelihood Error	0.02258	0.0007798	1	-0.008897
Random Posterior Error	0.007083	0.01845	-0.008897	1
Avg full Likelihood Error	0.1088	0.04177	0.09779	0.04908
Avg full Posterior Error	0.0525	0.1552	0.03956	0.1326
Avg all Likelihood Error	0.08448	0.02724	0.07724	0.03064
Avg all Posterior Error	0.03576	0.09495	0.02757	0.06374

Table 130: Correlation between errors and average errors for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.1088	0.0525	0.08448	0.03576
Posterior Error	0.04177	0.1552	0.02724	0.09495
Random Likelihood Error	0.09779	0.03956	0.07724	0.02757
Random Posterior Error	0.04908	0.1326	0.03064	0.06374
Avg full Likelihood Error	1	0.4177	0.8204	0.2948
Avg full Posterior Error	0.4177	1	0.2769	0.5934
Avg all Likelihood Error	0.8204	0.2769	1	0.3925
Avg all Posterior Error	0.2948	0.5934	0.3925	1

Table 131: Correlation between errors and pairwise dependency classifiers between evidence variables for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1625	-0.01792	-0.003332	0.01126
Posterior Error	0.04866	0.009484	-0.005986	0.00324
Random Likelihood Error	-0.004241	-0.004055	0.01649	0.02151
Random Posterior Error	-0.001433	-0.004134	-0.009171	-0.002675
Avg full Likelihood Error	0.04291	-0.03628	0.09354	0.21
Avg full Posterior Error	-0.02997	-0.0527	-0.0397	-0.002633
Avg all Likelihood Error	0.2552	0.2004	0.3255	0.4279
Avg all Posterior Error	0.004214	0.008639	-0.0161	0.002909

Table 132: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.09502	-0.06639	-0.06246	-0.03803
Posterior Error	-0.01632	-0.02883	-0.02169	-0.001972
Random Likelihood Error	-0.04272	-0.05337	-0.06298	-0.04657
Random Posterior Error	-0.02134	-0.02625	-0.02603	-0.01408
Avg full Likelihood Error	-0.3744	-0.5229	-0.4829	-0.3185
Avg full Posterior Error	-0.1475	-0.2626	-0.1874	-0.03652
Avg all Likelihood Error	-0.2985	-0.4191	-0.3836	-0.2516
Avg all Posterior Error	0.04755	-0.155	0.06619	0.2687

Table 133: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 0.999$ and $n = 5$

Determinism 0.999	Cond TC	Cond II	TC	II
Likelihood Error	0.08996	0.01051	0.01974	-0.01464
Posterior Error	0.03331	0.008367	0.008671	-0.02493
Random Likelihood Error	0.08849	-0.002399	0.01797	0.001786
Random Posterior Error	0.04107	0.01325	0.01136	-0.01567
Avg full Likelihood Error	0.8226	0.09617	0.264	-0.05735
Avg full Posterior Error	0.3305	0.08858	0.1123	-0.1379
Avg all Likelihood Error	0.8091	-0.05229	0.3461	0.0553
Avg all Posterior Error	0.2891	-0.08382	0.3159	0.01183

C.17. Determinism $\rho = 1$, $n = 5$

Table 134: Correlation between errors for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Lh Error	Post Error	Rnd Lh Error	Rnd Post Error
Likelihood Error	1	-0.02798	0.01718	-0.01122
Posterior Error	-0.02798	1	0.005035	0.002201
Random Likelihood Error	0.01718	0.005035	1	0.01567
Random Posterior Error	-0.01122	0.002201	0.01567	1
Avg full Likelihood Error	0.1046	0.04434	0.1097	0.05603
Avg full Posterior Error	0.03082	0.1106	0.05689	0.1143
Avg all Likelihood Error	0.0982	0.02576	0.08203	0.03981
Avg all Posterior Error	0.03985	0.03766	0.03142	0.04265

Table 135: Correlation between errors and average errors for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Avg full Lh	Avg full Post	Avg all Lh	Avg all Post
Likelihood Error	0.1046	0.03082	0.0982	0.03985
Posterior Error	0.04434	0.1106	0.02576	0.03766
Random Likelihood Error	0.1097	0.05689	0.08203	0.03142
Random Posterior Error	0.05603	0.1143	0.03981	0.04265
Avg full Likelihood Error	1	0.4051	0.8199	0.2853
Avg full Posterior Error	0.4051	1	0.2599	0.3088
Avg all Likelihood Error	0.8199	0.2599	1	0.4173
Avg all Posterior Error	0.2853	0.3088	0.4173	1

Table 136: Correlation between errors and pairwise dependency classifiers between evidence variables for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Simple Dep	Q-Score	MI	Pairwise Lh Error
Likelihood Error	-0.1389	0.0007432	0.01544	0.03166
Posterior Error	0.03714	-0.003093	-0.006198	-0.007946
Random Likelihood Error	-0.003611	-0.01738	-0.001517	0.01493
Random Posterior Error	0.02309	0.01586	0.01681	0.01883
Avg full Likelihood Error	0.05484	-0.04037	0.1061	0.2225
Avg full Posterior Error	0.085	0.08862	0.1003	0.09051
Avg all Likelihood Error	0.269	0.2069	0.3384	0.4433
Avg all Posterior Error	0.0526	0.06709	0.0358	0.03165

Table 137: Correlation between errors and pairwise dependency classifiers between the class variable and each evidence variable for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Class Simple Dep	Class Q	Class MI	Class Lh Error
Likelihood Error	-0.07928	-0.05853	-0.03942	-0.02243
Posterior Error	-0.02422	-0.01728	-0.03498	-0.02625
Random Likelihood Error	-0.02932	-0.05408	-0.04592	-0.02317
Random Posterior Error	-0.01664	-0.02594	-0.0394	-0.02963
Avg full Likelihood Error	-0.3827	-0.549	-0.4754	-0.3165
Avg full Posterior Error	-0.2036	-0.1983	-0.2903	-0.2488
Avg all Likelihood Error	-0.2973	-0.4275	-0.3762	-0.2438
Avg all Posterior Error	0.04593	-0.1338	0.04238	0.2318

Table 138: Correlation between errors and multivariate mutual information for forced determinism networks with $\rho = 1$ and $n = 5$

Determinism 1	Cond TC	Cond II	TC	II
Likelihood Error	0.09293	0.005874	0.04528	-0.003709
Posterior Error	0.02786	0.004194	-0.008404	-0.005519
Random Likelihood Error	0.08675	0.03079	0.03361	-0.01153
Random Posterior Error	0.04948	0.02769	0.006639	-0.02235
Avg full Likelihood Error	0.8155	0.09013	0.2745	-0.06194
Avg full Posterior Error	0.3298	0.08726	0.01795	-0.0957
Avg all Likelihood Error	0.8051	-0.05112	0.3593	0.04109
Avg all Posterior Error	0.2764	-0.1169	0.2862	0.0412