Before collecting data, nothing is known that will provide positive or negative evidence about the influence of any of the variables on any of the others. There are several ways to obtain data and to make inferences:

1. Conduct a study in which all variables are passively observed, and use the inferred associations or correlations among the variables to learn as much as possible about the causal relations among the variables.

2. Conduct an experiment in which one variable is assigned values randomly (randomized) and use the inferred associations or correlations among the variables to learn as much as possible about the causal relations..

3. Do (2) while intervening to hold some other variable or variables constant.

Procedure 1. is characteristic of non-experimental social science, and it has also been proposed and pursued for discovering the structure of gene regulation networks (Spirtes, et. al, 2001). Consistent algorithms for causal inferences from such data have been developed in computer science over the last 15 years, under weak assumptions about the data generating process, specifically the **Causal Markov Assumption**, which says that the direct causes of a variable screen it off from variables that are not its effects, and the **Faithfulness Assumption**, which says that all of the conditional independence relations are consequences of the **Causal Markov Assumption** applied to the directed graph representing the causal relations. Consistent search algorithms are available based on conditional independence facts — the PC-Algorithm, for example (Spirtes, et al., 2000) — and other consistent procedures are available based on assignments of prior probabilities and computation of posterior probabilities from the data (Meek, 1995; Chickering, 2002). We will appeal to facts about such procedures in what follows, but the details of the algorithms need not concern us.

There are, however, strong limitations on what can be learned from data that satisfy these assumptions, even supplemented with other, ideal simplifications. Thus suppose we have available the true joint probability distribution on the variables, and there are no unrecorded common causes of the variables (we say the variable set is **causally sufficient**), and there are no feedback relations among the variables. Under these assumptions, the algorithms can determine from the observed associations whether it is true that **X** and **Y** are adjacent, i.e., **whether X directly causes Y or Y directly causes X**, for all variables **X; Y**, but only in certain cases can the direction of causation be determined. For example, if the true structure is

$$\begin{tikzcd} X \arrow[dr] & & \\ & Z \arrow[r] & W \\ Y \arrow[ur] & & \end{tikzcd}$$

rious search lgorith s

only M values, in the worst case[2] we require at least

$$\frac{N}{2} M^{(N-2)}$$

different experiments to determine the entire structure. Suppose we have measured the messenger RNA (mRNA) expression levels of 10 genes and divide the expression levels into high, medium and low values. We would require in the worst case at least [illegible] experiments.

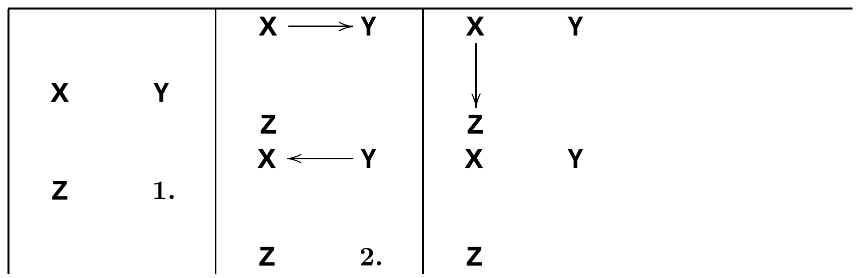
Various modifications of the control procedure might improve these worst case results, and for many probability distributions over the possible causal structures the expected case number of experiments would presumably be much better. But we propose a principled result. By combining procedure 1 with procedure 2, under the assumptions so far listed, for $N > 2$, in the worst case, the complete causal structure on N variables can be determined with $N - 1$ experiments, counting the null experiment of passive observation (procedure 1) as one experiment, if conducted. Further, this is the best possible result when at most one variable is randomized in each experiment.

2 The Idea

Consider the case of $N = 3$ variables. There are 25 directed acyclic graphs on 3 vertices. In Figure X we show the graphs sorted into sub-classes that are indistinguishable without experimental intervention.

---

[2] [illegible]


X Y
Z 1.
X Y
Z
X Y
Z 2.
X Y
Z
X Y
Z

imply that when **Y** and **Z** are independent conditional on **X**, there is no direct causal relation between **Y** and **Z**. The top graph in Box 6 must therefore be the true graph. By combining search procedures (in this case used informally) with experimentation, we have determined the truth with a single experiment. (We were lucky; if we had begun by randomizing **Y** or **Z**, two experiments would have been required.) When we randomize **X** and follow up with a consistent search procedure, which requires no additional experimentation, all of the direct connections between the remaining variables can be estimated. Only the directions of some of the edges remain unknown. Those directions can clearly be determined by randomizing each of the remaining variables.

In some cases, we lose something when we experiment. If, when **X** is randomized, **X** and **Y** do not covary, we know that **X** does not cause **Y**, but we do not know whether **Y** causes **X** or neither causes the other, because our manipulation of **X** has destroyed any possible influence of **Y** on **X**. Thus in the single structure in Box 6, if we randomize **X**, and **Y** and **Z** do not covary with **X**, every structure in which **X** is not a direct or indirect cause of **Y** or **Z**, and effect. Suppose instead, we begin by randomizing **X**. If we find **X**;**Y** are not associated, a second experiment is required to determine whether **Y** causes **X**.

The proof of the bound has three perhaps surprising corollaries. (1) Any procedure that includes passive observation in which no variables are randomized exceeds the lower bound for some cases, when the passive observation is counted as one experiment. (2) Controlling for variables by experimentally fixing their values is never an advantage. (3) Adaptive search procedures (Murphy, 1998; Tong and Koller, 2001) choose the most informative next experiment given the results of previous experiments. That is, they choose the next experiment that maximizes the expected information to be obtained. We also show tiple si ultaneous rando ization

The data is such that we can identify the conditional independencies if there are any.

Interventions: Interventions are possible on every variable.

e n Cons

An **experiment** randomizes at most one variable and returns the joint distribution of all variables.
A **procedure** is a sequence of experiments and a structure learning algorithm applied to the results of these experiments.
A procedure is **reliable** for an N vertex problem iff for all DAGs on N vertices the procedure determines the correct graph uniquely.
A procedure is **order reliable** for an N vertex problem iff it is reliable for all non-redundant orderings of experiments.
A procedure is **adaptive** iff it chooses at each step one from among the possible subsequent experiments as a non-trivial function of the results of the previous experiments.

C s

Proposition 1 **For $N > 2$, there is an order reliable procedure that in the worst case requires no more than $N-1$ experiments, allowing only single interventions.**

Proof: Consider a graph with N vertices where $N > 2$ and let $X_1, \ldots, X_N$ specify an arbitrary ordering of these vertices. Let each experiment consist of an intervention on one variable. Perform $N-1$ experiments, one intervention on each X_i where $1 \le i \le N-1$. By Lemma , applying the PC algorithm to the first experiment determines the adjacencies among at least $X_2 \ldots X_N$. The kth experiment determines the directions of all edges adjacent to X_k: if X_j is adjacent to X_k, then X_k is a direct cause of X_j if and only if X_j covaries with X_k when X_k is randomized (since if X_k were only an indirect cause of X_j, and since X_j and X_k are adjacent, X_j would have to be a direct cause of X_k, and there would be a cycle), otherwise, X_j is a direct cause of X_k. X_N has not been randomized, but its adjacencies with every other variable have been determined by the $N-1$ experiments. Suppose X_N and X_k are adjacent. Since X_k has been randomized, X_k is a cause of X_N if and only if X_N covaries with X_k when X_k is randomized. In that case, if X_k were an indirect but not a direct cause of X_N, then X_N would be a direct cause of X_k, because X_N and X_k are adjacent, and hence there would be a cycle. If X_N and X_k do not covary when X_k is randomized, then, since they are adjacent, X_N is a direct cause of X_k. If X_k and X_N are not adjacent, then this missing edge would have been identified in one of the interventions on X_j where $j \neq k$. These are all of the cases. Q.E.D.

Lemma 1 **If G is a causal graph over a set of variables V, and G' the manipulated graph resulting from an ideal intervention on variable X in G, then for all**

The fact that the sequence of e peri ental inter entions is ar itrary in the pre ious proof suggests that this result is still true for the orst case e en hen the choice of the ne t e peri ent is adapti e, that is, e en if at each point during the sequence of e peri ents the est e peri ent gi en the e idence fro the pre ious e peri ent is chosen. Although Proposition 5 follo s fro the pre ious t o proofs as a corollary, the proof e e phasizes the aspect that no **adaptive**

C Other types of experiments

In the previous two proofs an experiment was always assumed to consist of an intervention on one particular variable. However, it might be thought that other types of experiments, such as passive observations or interventions on more than one variable might improve the worst case result of $N-1$ experiments. While it is true that multiple interventions (randomizing more than one variable at a time) can shorten the experimental sequence, this is not the case for passive observational studies. We call a passive observational experiment a null-experiment.

The above proofs indicate that the worst case always occurs for particular complete graphs. If one were to run a null-experiment at any point in the experiment sequence when the underlying graph is complete - the most likely time would probably be at the beginning - then one would realize that one is confronted with a complete graph. However, this information (and more) is obtained anyway from two sequential experiments, each consisting of an intervention on a particular variable. The null-experiment paired with any other experiment cannot generate more information about the graph than two single intervention experiments, since a single intervention experiment also identifies all adjacencies except for those into the intervened variable. But a second intervention on a different variable would identify these interventions, too. So the only advantage of the null-experiment is in the case where only one experiment is run. The above proofs only apply to graphs of three or more variables, which certainly cannot be identified by one experiment alone. In fact, even for two variables, two experiments are needed in the worst case (see discussion in main body of the paper).

6 References

D.M. Chic ering (2002) Learning Equi alence Classes of Bayesian Net or Structures. **Journal of Machine Learning Research**, 2 5 8.

C. Mee , ( ) Ph.D hesis, Depart ent of Philosophy, Carnegie Mellon ni ersity

.P. Murphy, (200) Acti e Learning of Causal Bayes Net Structure, echnical Report, Depart ent of Co puter Science, .C. Ber eley.

J. Pearl, (88) **Probabilistic Reasoning in Intelligent Systems**, San Mateo, CA. Morgan auf ann.

J. Pearl, (2000) **Causality**, O ford ni ersity Press.

P. Spirtes, C.Gly our and R. Scheines, ( ) **Causation, Prediction and Search**, Springer Lecture Notes in Statistics, , 2nd edition, M Press, 2000.

P. Spirtes, C. Gly our, R. Scheines, S. an, . Ai alie, and F. i erly, (200) Constructing Bayesian Net or Models of Gene E pression Net or s fro Microarray Data, in **Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems and Technology**, Du e ni ersity, March.

S. ong and D. oller, (200) Acti e Learning for Structure in Bayesian Net or s, **Proceedings of the International Joint Conference on Arti cial Intelligence**.