Overview
We propose an alternative to the estimand based paradigm for answering causal queries. The idea is to learn the full causal model from the observational data and causal diagram, and then answer the query by applying Probabilistic Graphical Models (PGM) algorithms. We show that this model completion learning approach can be far more effective than estimated approaches, particularly in large models when the estimand computation is complex and the induced width of the diagram is small.

Contributions:
1. Provide a first of its kind, extensive empirical evaluation on causal effect algorithms on varied and large synthetic and real networks.
2. Show empirically that our approach has more accurate estimates than estimated based schemes.

Problem
Given a causal diagram, an identifiable query $P(Y|do(X=x))$ and samples from the observed distribution, the task is to output the distribution of $P(Y|do(X=x))$.

Current Practice
1. Apply state of the art algorithms for identifiability. These are polynomial algorithms involving the graph and the query only.
2. More sophisticated statistical estimation techniques don't scale to be practical [J. Zhang et al, 2022].
3. There exists theoretical bounds on sufficient domain sizes.

Motivating Example
Figure 1: Chain Model with 7 observable variables and 3 latent variables

Learning for Causal Inference

Identifiability
- Any two models that agree on the observational distribution and causal diagram will also agree on $P(Y|do(X=x))$.

EM for Causal Inference (EM4CI)

Algorithm 1: EM4CI

input : $A$ causal diagram $G = (U, V, E)$, $U$ latent and $V$ observable variables; samples from $P(V)$;
output : $\hat{P}(V|do(X=x))$

1. Initialize: $BIC_{CP} = \infty$.
2. If id-identifiable ($G, Q$), terminate.
3. For $k = 2, \ldots$, do
   a. $(B', LCP) \leftarrow \max \{EM(G, D, k) | k \leq \text{10}\}$
   b. Calculate $BIC_{CP}$ from $LCP$.
   c. If $BIC_{CP} \leq BIC_{CP}$, $B \rightarrow B'$.
   d. else, break.
4. endfor
5. $B(x) \leftarrow$ generate truncated CBn from $B$.
6. return $\hat{P}(X=x, V)$.

Theorem: $BIC_{CP}$ is a probability distribution over the exogenous variables

Complexity
- Time and memory are exponential in the induced width.

Benefits & Challenges

Challenges
1. In order to learn the full model we need to learn a domain size for the latent variables.
2. There exists theoretical bounds on sufficient domain sizes. However the bounds are very conservative & can be very large to be practical [J. Zhang et al, 2022].
3. EM algorithm can be slow and converge to incorrect local optima in high dimensional space.

Benefits
1. Learning phase only needs to be performed once to answer any identifiable of form $P(Y|do(X=x))$; traditionally a new estimand would need to be derived for each query.
2. EM4CI consistently yields extremely accurate results.

Experimental Setup

Benchmarks
- Each benchmark includes a causal diagram, a query, and observational data synthetically generated from the full model.
- Used a range of domain sizes of for the variables.
- Test on bayesian networks from real world domains, and created single-value queries over all instantiations.
- Return $P(Y|do(X=x))$.

Performance Measures
- To evaluate the accuracy of $P(Y|do(X=x))$, we use the mean absolute deviation (mad) over all single-value queries over all instantiations of the intervened and queried variables.
- BIC score is used to evaluate fitness of the learned model and impose some regularization over the domain sizes.

Notation
- Capital letters (X) represent variables, & small letters (x) represent their values. Boldfaced capital letters (X) denote a collection of variables.
- $n = |V|$, $d = |DV(X)|$, $k = |DV(U)|$ in the true model, and $k_{\text{est}} = |DV(U)|$ the latent domain of the learned model

Conclusion
- EM4CI was extremely accurate on all benchmarks we tried.
- Learns causal effects by weighted empirical risk minimization.
- State of the art method that focuses on estimating the quantities in the estimand using statistical methods.

Synthetic Network Results

UAI Benchmark Results

Table 1: Plug-In & EM4CI results on the A Network $|V| = 46$, |E| = 80; $d = 2$, $k = 2$ treewidth $\approx 16$

<table>
<thead>
<tr>
<th>Query</th>
<th>Plug-In (% Correct)</th>
<th>EM4CI (% Correct)</th>
<th>Plug-In (BIC)</th>
<th>EM4CI (BIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(I</td>
<td>do(X=x))$</td>
<td>0.8984</td>
<td>0.9016</td>
<td>0.2457</td>
</tr>
<tr>
<td>$P(O</td>
<td>do(X=x))$</td>
<td>0.8022</td>
<td>0.8082</td>
<td>0.0646</td>
</tr>
<tr>
<td>$P(L</td>
<td>do(X=x))$</td>
<td>0.6213</td>
<td>0.6437</td>
<td>0.3857</td>
</tr>
<tr>
<td>$P(R</td>
<td>do(X=x))$</td>
<td>0.6542</td>
<td>0.6742</td>
<td>0.2457</td>
</tr>
</tbody>
</table>

EM4CI Learning: BIC(\% correct): $\text{BIC}(\text{EM4CI}) = 0.1(\text{EM4CI})$, $\text{BIC}(\text{Plug-in}) = 0.1(\text{Plug-in})$.