EMPIRICAL BAYESIAN APPROACHES FOR ROBUST CONSTRAINT-BASED CAUSAL DISCOVERY UNDER INSUFFICIENT DATA

Zijun Cui (presenter) ¹
cui3@rpi.edu

Naiyu Yin ¹

Yuru Wang ²

Qiang Ji ¹
jiq@rpi.edu

¹ Rensselaer Polytechnic Institute
² Northeast Normal University
Causal Discovery

- Causal relations among variables are captured by a directed acyclic graph (DAG)
  - A direct link from node $X$ to node $Y$ indicates the cause-effect relation between cause variable $X$ and effect variable $Y$

- Causal discovery is to learn a DAG capturing cause-effect relationships among a set of random variables from observational data

- **Causal discovery under insufficient data is of great importance**
  - Existing methods are focused on learning a DAG with high confidence under sufficient data
  - However, in many domains, the availability of data is very limited
Constraint-based Causal Discovery

- Constraint-based causal discovery methods apply independence tests to determine a DAG from observational data.
- It can be performed globally or locally.

**Global approaches** aim at learning cause-effect relationships among all random variables.

**Local approaches** aim at identifying the direct causes and effects of a target variable, represented by a causal Markov blanket.

Example structure is revised based on ASIA dataset in Bnlearn Repository.
Bayesian Approaches for Independence Tests

- For both global and local approaches, the main challenge of the constraint-based causal discovery is that its performance highly depends on the accuracy of the independence test.

- We propose two Bayesian-augmented frequentist independence tests:
  - Bayesian approach is adopted to reliably estimate independence test statistics with limited data by considering the entire parameter space instead of using a point estimate one.
  - The Bayesian statistics are then used by frequentist independence tests.

- Specifically, we introduce Bayesian approach for two types of independence tests:
  - Mutual Information based independence test
  - Statistical testing based independence test
Independence Test

- **Mutual information based independence test**

  - The mutual information (MI) of two discrete random variables $X$ and $Y$ is defined as
  \[
  \text{MI}(X ; Y) = \sum_{i=1}^{K_X} \sum_{j=1}^{K_Y} \frac{\theta_{ij}}{\theta_i \theta_j} \ln \frac{\theta_{ij}}{\theta_i \theta_j}
  \]
  where $K_X$ and $K_Y$ denote the total number of possible states of $X$ and $Y$. $\theta_i = p(x_i)$, $\theta_j = p(y_j)$ and $\theta_{ij} = p(x_i, y_j)$ are probability distribution parameters.

  - If $\text{MI}(X ; Y) < \text{Threshold}$, $X$ and $Y$ are declared to be independent; Otherwise, $X$ and $Y$ are dependent.

- **Statistical testing based independence test**

  - G-test is a standard likelihood ratio test. Its statistics $g$ asymptotically follows the $\chi^2_{df=(K_X-1)(K_Y-1)}$ distribution and is defined as
  \[
  g = -2 \sum_{i=1}^{K_X} \sum_{j=1}^{K_Y} n_{ij} \ln \frac{\theta_i \theta_j}{\theta_{ij}}
  \]

  - If $p$-value is smaller than the significance level (default 5%), the null hypothesis is rejected and the alternative hypothesis is accepted. Thus, $X$ and $Y$ are declared to be dependent; Otherwise, $X$ and $Y$ are declared to be independent.

- **Independence Test Accuracy under insufficient data**

  - Existing methods perform a Maximum Likelihood estimation (MLE) of the parameters $\theta$ directly from data $D$, i.e.,
  \[
  \theta = \arg \max P(D | \theta)
  \]
  The MLE estimates are inaccurate when $D$ is insufficient. As a result, independence tests are subject to errors under limited data.
Bayesian Approach for Mutual Information based Independence Test

- Full Bayesian MI is based on estimating expected MI over data $D$:
  \[ MI^{FB}(X;Y|D) = \int \int MI(X;Y|\theta)p(\theta,\alpha|D)d\theta d\alpha = \int \int MI(X;Y|\theta)p(\theta|\alpha,D)p(\alpha|D)d\theta d\alpha \]

- The integration over $\alpha$ is approximated by maximizing it out as
  \[ MI^{eB}(X;Y|D) = \int \int MI(X;Y|\theta)p(\theta,\alpha|D)d\theta d\alpha = \int MI(X;Y|\theta)p(\theta|\alpha^*,D)d\theta \]

  with $\alpha^* = \text{argmax} \ p(\alpha|D) = \text{argmax} \ p(D|\alpha)p(\alpha)$. Assuming $p(\alpha)$ follows the uniform distribution, we have $\alpha^* = \text{argmax} \ p(D|\alpha)$ and can be solved through a fixed-point update.

- Given the $\alpha^*$, we in the end have
  \[ MI^{eB}(X;Y|D) = \psi(N + \alpha^*K + 1) - \sum_{ij} \frac{n_{ij} + \alpha^*}{N + \alpha^*} [\psi(n_i + \alpha^*K_y + 1) + \psi(n_j + \alpha^*K_x + 1) - \psi(n_{ij} + \alpha^* + 1)] \]

  where $\psi(x)$ is the digamma function. $n_i$ and $n_j$ are the number of samples for $X = i$ and $Y = j$ respectively, and $n_{ij}$ is the number of samples for $(X,Y) = (i,j)$.
Bayesian Approach for Statistical Testing based Independence Test

- A Bayesian estimate of hypothesis likelihood is considered as

\[
BF = \frac{p(D|H_0,a_0)}{p(D|H_1,a_1)} = \frac{\int p(D|\theta,H_0)p(\theta|H_0,a_0)d\theta}{\int p(D|\theta,H_1)p(\theta|H_1,a_1)d\theta}
\]

\(a_0\) and \(a_1\) are the respective hyper-parameters under null and alternative hypothesis.

- To apply BF for a statistical testing, like \(G\) test, we approximate it as

\[
BF_{\text{chi2}} = -2\ln BF = -2 \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} n_{ij} \ln \frac{\bar{\theta}_i \bar{\theta}_j}{\bar{\theta}_{ij}}
\]

\(BF_{\text{chi2}}\) asymptotically follows the distribution \(\chi^2_{df=(K_x-1)(K_y-1)}\). We set 5\% as the default significance level.
Local Causal Discovery

- We consider the causal Markov blanket (CMB) for comparison
- $c^{IeB}$ denotes the CMB with empirical Bayesian MI estimation; $cBF_{chi2}$ denotes the CMB with $BF_{chi2}$ independence test

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>$c^{IeB}$</th>
<th>$cBF_{chi2}$</th>
<th>CMB</th>
<th>$c^{IeB}$</th>
<th>$cBF_{chi2}$</th>
<th>CMB</th>
<th>#Independence Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHILD</td>
<td>100</td>
<td>2.90±0.28</td>
<td>2.65±0.40</td>
<td>5.94±0.65</td>
<td>1008</td>
<td>1154</td>
<td>16869</td>
<td>1709</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>2.61±0.26</td>
<td>2.64±0.59</td>
<td>6.95±0.63</td>
<td>709</td>
<td>1926</td>
<td>14578</td>
<td>2524</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>2.29±0.31</td>
<td>2.24±0.84</td>
<td>4.52±0.58</td>
<td>254</td>
<td>4751</td>
<td>13873</td>
<td>1747</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>2.60</td>
<td>2.51</td>
<td>5.80</td>
<td>1426</td>
<td>2763</td>
<td>18364</td>
<td>2807</td>
</tr>
<tr>
<td>INSURANCE</td>
<td>100</td>
<td>3.89±0.34</td>
<td>3.98±0.39</td>
<td>7.18±0.66</td>
<td>1261</td>
<td>1363</td>
<td>22168</td>
<td>1541</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>3.47±0.21</td>
<td>3.24±0.12</td>
<td>7.59±0.57</td>
<td>1477</td>
<td>3949</td>
<td>14881</td>
<td>1477</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>3.11±0.21</td>
<td>2.98±0.13</td>
<td>7.20±0.67</td>
<td>1426</td>
<td>2763</td>
<td>18364</td>
<td>1426</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>3.40</td>
<td>3.40</td>
<td>7.32</td>
<td>2210</td>
<td>3253</td>
<td>17907</td>
<td>2210</td>
</tr>
<tr>
<td>ALARM</td>
<td>100</td>
<td>2.09±0.07</td>
<td>2.39±0.19</td>
<td>5.20±0.71</td>
<td>1424</td>
<td>1109</td>
<td>27492</td>
<td>2398</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>2.50±0.19</td>
<td>2.27±0.15</td>
<td>4.36±0.83</td>
<td>2398</td>
<td>3885</td>
<td>14900</td>
<td>2398</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>2.40±0.11</td>
<td>2.26±0.19</td>
<td>3.53±0.62</td>
<td>2807</td>
<td>4766</td>
<td>11328</td>
<td>2807</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>2.53</td>
<td>2.31</td>
<td>4.36</td>
<td>2210</td>
<td>3253</td>
<td>17907</td>
<td>2210</td>
</tr>
<tr>
<td>HAILFINDER</td>
<td>500</td>
<td>3.33±0.02</td>
<td>4.22±0.04</td>
<td>7.90±0.11</td>
<td>676</td>
<td>1923</td>
<td>183350</td>
<td>1098</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>3.56±0.01</td>
<td>4.49±0.13</td>
<td>7.12±0.09</td>
<td>1924</td>
<td>2621</td>
<td>119815</td>
<td>1233</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>3.56±0.09</td>
<td>4.45±0.08</td>
<td>7.10±0.11</td>
<td>2233</td>
<td>2229</td>
<td>157620</td>
<td>2233</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>3.48</td>
<td>4.39</td>
<td>7.37</td>
<td>7168</td>
<td>7417</td>
<td>14789</td>
<td>7168</td>
</tr>
<tr>
<td>CHILD3</td>
<td>500</td>
<td>2.46±0.23</td>
<td>2.53±0.18</td>
<td>4.72±0.28</td>
<td>7168</td>
<td>7417</td>
<td>14789</td>
<td>7168</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>3.01±0.13</td>
<td>2.67±0.11</td>
<td>3.57±0.21</td>
<td>6720</td>
<td>7802</td>
<td>9765</td>
<td>6720</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.90±0.07</td>
<td>2.57±0.23</td>
<td>3.09±0.19</td>
<td>8424</td>
<td>8285</td>
<td>9516</td>
<td>8424</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>2.79</td>
<td>2.59</td>
<td>3.79</td>
<td>7437</td>
<td>7835</td>
<td>11357</td>
<td>7437</td>
</tr>
<tr>
<td>CHILD5</td>
<td>500</td>
<td>2.87±0.05</td>
<td>2.62±0.19</td>
<td>5.00±0.15</td>
<td>5234</td>
<td>11126</td>
<td>16819</td>
<td>8236</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>2.66±0.21</td>
<td>3.02±0.13</td>
<td>5.75±0.32</td>
<td>13384</td>
<td>9956</td>
<td>36888</td>
<td>13384</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.82±0.23</td>
<td>2.99±0.07</td>
<td>4.34±0.19</td>
<td>8951</td>
<td>10835</td>
<td>26322</td>
<td>8951</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>2.78</td>
<td>2.88</td>
<td>5.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Both $c^{IeB}$ and $cBF_{chi2}$ outperform CMB in terms of both accuracy (SHD) and efficiency (# Independence Test)

- Comparing the performance between the two proposed methods
  - $cBF_{chi2}$ achieves overall better accuracy
  - $c^{IeB}$ is more efficient with the fewest number of independence tests on all datasets
Global Causal Discovery

- We consider the RAI-BF and PC-Stable for comparison
- \( r_I^{EB} \) denotes the RAI with empirical Bayesian MI estimation; \( r_{BF_{chi2}} \) denotes the RAI with \( BF_{chi2} \) independence test
- \( r_I^{EB} \) and \( r_{BF_{chi2}} \) outperform RAI-BF and PC-Stable in terms of both accuracy (SHD) and efficiency (\# Independence Test)
- Comparing the performance between the two proposed methods
  - \( r_{BF_{chi2}} \) achieves overall better accuracy
  - \( r_I^{EB} \) achieves overall better efficiency
- We reach consistent conclusions
Conclusions

- We introduce Bayesian methods for robust constraint-based causal discovery under insufficient data.

- Two Bayesian-augmented frequentist independence tests are proposed for reliable statistic estimation under a frequentist independence test framework.

- Through extensive experiments, we show that, by introducing Bayesian approaches, the proposed methods not only outperform the competing methods in terms of accuracy, but also improve efficiency significantly.
Thank You!