Final Report : Dynamic Causal Bayesian Optimization and Optimal Intervention

Name: Lau, Ying Yee Ava

A. Q1

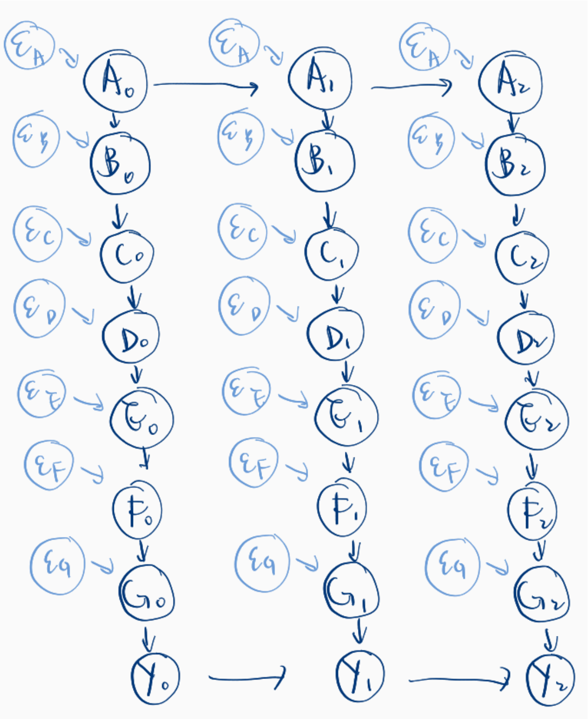
In the paper, the authors considered very small and simple graphs, which might not be the case in practice.

A1. Can you give an example of a causal graph with 15 nodes at each time step – 7 non-manipulable, 7 manipulable, and 1 target variable, how would you get the exploration set (a key input to the algorithm)?

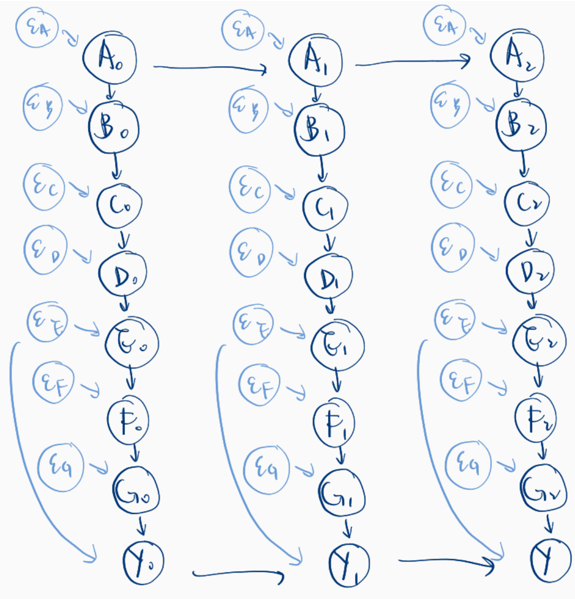
1. The exploration set is determined by the structure of the causal graph. As outlined in Appendix D, variables do not need to be included in the exploration set for intervention if they lack both back-door and front-door paths. Manually determining the exploration set requires the knowledge of structural causal graphs and do-calculus, and it will vary depending on the specific structure of the graph.

   - In the following context, node Y is the target variable. Dark blue nodes represent manipulable variables, while light blue nodes are non-manipulable. For instance, the common cause ε_E is non-manipulable.

   - In the example below, to investigate the causal effect of B or its descendants on Y, we might consider intervening on A. However, given that some variables cannot be manipulated, intervening on B could be advantageous for examining the causal effects of its descendants on the target variable, particularly if the non-manipulable variable has a substantial influence. Therefore, the exploration set is {{A},{B},{C},{D},{E},{F},{G}}.



   - In the next example, if we aim to study the causal effect of E on Y, we cannot intervene on ε_E as it is only observable. From do-calculus knowledge, we can perform a frontdoor adjustment by intervening on E, F, and G. For the causal effect of B,C or D on Y, we can intervene on {A, E, F, G}. Other non-manipulable variables might introduce noise to their children's nodes, so intervention could be necessary. One potential exploration set is {{E,F,G},{F,G}, {G},{A,E,F,G}, {B,E,F,G},{C,E,F,G},{D,E,F,G}}.



2. We can also rely on the variable of interest and domain knowledge. For example, in the DCBO paper there is a ODE experiment with 7 variables. They did not select the exploration set through the graph structure. Since they wanted to control the effect of mortality concentration in the chemostat, they chose to intervene on Nitrogen concentration, Predator juvenile concentration, and Predator adult concentration.
3. Consider conducting a literature review on methods for identifying optimal exploration sets in causal graphs with high dimensionality. It is possible that there are state-of-the-art algorithms available that can effectively address this challenge.

A2. Would you write a program for this purpose?

Yes, it is worthwhile to develop a program. Relying solely on domain knowledge or using small graphs is not always feasible. Manually finding the exploration set can be time-consuming and prone to errors, such as overlooking or over-identifying backdoor paths.

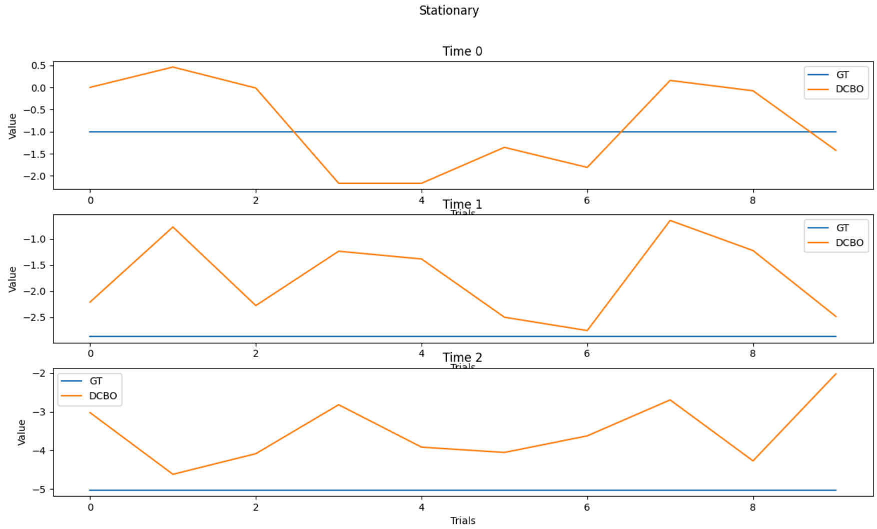
A3. Is it enough to have simply the causal diagram to get the exploration set? What additional specifications do you need ?

I believe if we have a clear domain of interest or a treatment for measurement and the model is small enough, it might be possible to determine the exploration set. However, considering all temporal and causal dependencies from an exploratory perspective in medium-sized graphs, there are numerous combinations of backdoor paths to account for. In such cases, it is essential to incorporate domain knowledge regarding the subject of the graph.

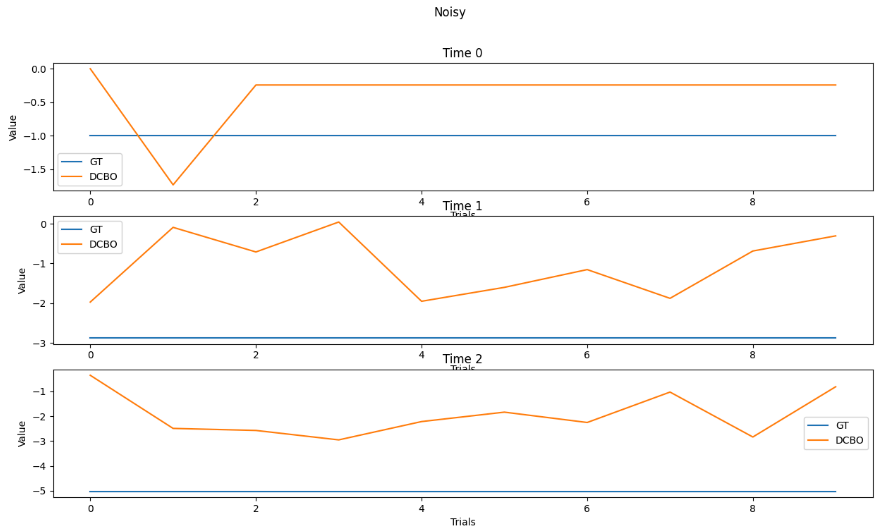
B. Replicate their synthetic experiment results in Tensorflow probability.

I encountered challenges in replicating the experiment results as reported in the paper. Firstly, the predicted optimal intervened values deviate significantly from the ground truth. Secondly, the plot of intervened values against the number of trials does not show a clear trend of convergence toward the ground truth.

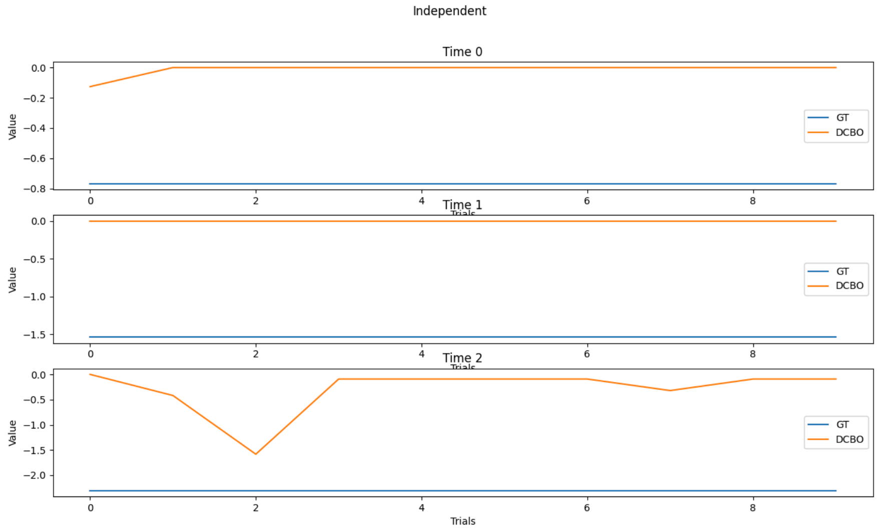
1. Experiment `Stat`
   - Ground truth y^*:
     - (t = 0) −1.0003635
     - (t = 1) −2.8754091
     - (t = 2) −5.027236
   - Predicted $\hat{y}^*$:
     - (t = 0) −2.1683995723724365
     - (t = 1) −2.757913112640381
     - (t = 2) −4.618849277496338
   - The graph below illustrates the comparison between the ground truth and predicted values across iterations.



2. Experiment `Noisy`
   - Ground truth y^*:
     - (t = 0) −1.0003635
     - (t = 1) −2.8754091
     - (t = 2) −5.027236
   - Predicted $\hat{y}^*$:
     - (t = 0) −1.7354867458343506
     - (t = 1) −1.969146966934204
     - (t = 2) −2.953901529312134



3. Experiment `Ind`
    - Ground truth y^*:
        - (t = 0) −0.7682934
        - (t = 1) −1.5365868
        - (t = 2) −2.3048801
    - Predicted $\hat{y}^*$:
        - (t = 0) −0.12610842287540436
        - (t = 1) −1.0921138803947614e-21
        - (t = 2) −1.5828551054000854



C. Q3 (Part 1)

C1. What are the key ideas of this paper? What are the ideas that excite you the most? Why do you find them interesting and critical?

The paper introduces several key ideas through its theorems and propositions, which are essential for understanding the DCBO algorithm. The central concept is Dynamic Causal Global Optimization (DCGO), an extension of Bayesian Optimization that incorporates temporal causal dynamics among variables. The goal of DCGO is to identify the optimal intervention set and levels that minimize the value of the target variable. Several critical details enable DCGO:

1. Time Operator Theorem: This theorem demonstrates that the transition function of the target variable can directly utilize the optimal intervention sets identified in previous timestamps. Specifically, at each time step t, interventions are built upon those from the preceding time step $t - 1$. Importantly, interventions are not applied to time steps earlier than $t - 2$.
2. Corollary 1: This emphasizes the necessity of updating the values of the children of intervened nodes in the causal graph during interventions.
3. Proposition 3.1: This proposition states that if the causal structure remains unchanged over time, the exploration set can remain consistent across time steps.

The most fascinating concept to me is the time operator. In temporal systems, the number of causal variable relationships grows over time. The time operator allows the intervention at time t to be directly conditioned on the optimal intervention set from $t - 1$. This is crucial because it prevents the exploration set from expanding over time, thereby simplifying the optimization problem and reducing computational complexity.

C2. Do you think the acquisition function is correct? Is there any typo? If yes, can you explain? If not, can you derive the correct one?

The definition presented in the paper for the acquisition function is

$$\mathrm{EI}_{s,t}(\mathbf{x}) = \mathbb{E}_{p_{(y_{s,t})}}\left[\max\left(y_{s,t} - y_t^*, 0\right)\right] / \operatorname{cost}\left(\mathbf{X}_{s,t}, \mathbf{x}_{x,t}\right)$$

. While the overall concept of the acquisition function is sound, there are two minor concerns:

1. The paper does not explicitly define $p\left(y_{s,t}\right)$, but I believe it refers to $p\left(y_{s,t} \mid f_{s,t}, \sigma^2\right)$, as discussed in the “Likelihood” subsection. This corresponds to the causal prior providing the mean function $f_{s,t}$ and the variance σ^2 to the bayesian model.
2. As noted in a last paragraph of the “Acquisition Function” subsection, the parameter ξ is employed to adjust the system’s level of exploration, which corresponds to the `jitter` parameter of the `class CausalExpectedImprovement` in the code.

Thus, a more precise formulation of the acquisition function would be:

$$\mathrm{EI}_{s,t}(\mathbf{x}; \xi) = \mathbb{E}_{p(y_{s,t} \mid f_{s,t}, \sigma^2)}\left[\max\left(y_{s,t} - (y_t^* + \xi), 0\right)\right] / \operatorname{cost}\left(\mathbf{X}_{s,t}, \mathbf{x}_{x,t}\right)$$

.

C3. Do you find any errors or questionable issues?

- I think there is a potential issue in the original codebase. In `dag_utils/adjacency_matrix_utils.py`, the function `get_emit_and_trans_adjacency_mats` generates emission and transition matrices from adjacency matrices. When repeated edges exist in the graph, the resulting matrix elements can exceed 1. However, this situation is not addressed in the `fit_arcs` function within `sem_utils/sem_estimate.py`; the program may mistakenly interpret repeated edges as fork nodes. Since causal graphs should not contain repeated edges, it is essential to enforce this constraint in the code.

- In `dcbo_base.py`, the likelihood variance is fixed using `model.likelihood.variance.fix()` during the Bayesian model initialization. This raises one question issue: why the authors decided to fix the variance instead of allowing it to adapt dynamically based on the kernel variance and noise in the likelihood, which could improve optimization accuracy? A likely reason is the instability of the `semhat` object (which represents graph nodes using kernel density estimation or Gaussian processes) at the start of the optimization process.

- I do not understand the process of the creation of the ground truth dataset: when constructing intervention sets and levels, we select them based on the smallest possible values that can be generated. This issue is also reflected in the codebase. As shown in `root.py/_post_optimisation_assignments`, the optimal intervention value is selected by minimizing all intervened values generated across all trials, as indicated by `best_objective_fnc_value_idx = self.outcome_values[t].index(eval(self.task)(self.outcome_values[t])) - 1` with `self.task = "min"`.

C4. Are there any parts that are not clear in the paper? Can you please include derivations that are not clear in their paper?

- In definition 4, the terms Y_t^{PT} and $\mathbf{X}_{s,t}^{\mathrm{PY}}$ appear, but their interpretations are ambiguous. I interpret "PT" as referring to the parent of the target variable, while "PY" seems to denote the parent of variable Y. Thereore, I think "PT" and "PY" carry the same meaning.
- The paper does not provide a definition for $\mathcal{D}^I$ iat the beginning, but it is introduced in the experiments section of CBO. There, it is defined as $\mathcal{D}^I = \{(\mathbf{x}_s^i, \mathbb{E}[Y|\,\mathrm{do}(\mathbf{X}_s^i = \mathbf{x}_s^i])\}_{i=1,s=1}^{P,\,|\mathbf{ES}|}$.
- In Corollary 1, the term "PW" is not defined; I assume it refers to the parents of W.
- Also in Corollary 1, although the function $C(\cdot)$ is mentioned in the main text, its definition as a function that returns child nodes is not explicitly stated. I inferred this meaning from Appendix B, where it is noted that $\{f_W(\cdot)\}_{W\in\mathbf{W}}$ is recursive.

C5. Is there any important part of their code that is not mentioned in the paper?

- The intervention grid for generating the intervened samples has a limit in size. In `utils/sequential_intervention_functions.py`, if the number of nodes in the `exploration_set` is less than or equal to 2, the value of `size_intervention_grid` is unrestricted. However, if the number of nodes exceeds 2, `size_intervention_grid` is capped at 10 when the given size is greater than or equal to 100.
- The concept of transition and emission matrices is introduced in Assumption 1, but its term is not explicitly introduced in the paper.
- Different BO models are utilized at different time steps. When initializing the BO model, there is an option to transfer variances from the previous time step to the current one. Alternatively, the user can choose to use a default initialization value of 1.0 for the variance.
- In the "Prior Surrogate Model" section, it is mentioned that kernels are created for source nodes, while "other nodes" are represented as GPs. In the code, "other node" is further divided into fork, chain and collider, in which there are differences in handling the GP regression model. For example, the GP function for colliders (n-to-1) is $f : \mathbb{R}^n \to \mathbb{R}$, while for forks (1-to-n), it is n initializations of $f : \mathbb{R} \to \mathbb{R}$.

C6. Can you explain the paper based on details you have learnt from their code and your result replication experience?

The following is the detailed explanation of the DCBO algorithm based on the code and the replication experience:

1. Observation Dataset and SEMs: Observation dataset samples are generated from the pre-defined SEMs.
2. Causal Priors Generation:
   - Analyze the SCM by separating its adjacency matrix into transition and emission matrices.
   - Generate two sets of priors for transition and emission, fitting KDEs for source nodes, and GP Regression for colliders and other nodes.
   - These priors form an estimated SEM, providing static and dynamic SEMs that encapsulate relationships in the SCMs.
   - Create two copies of SEMs for mean and variance, which serve as causal priors.
3. Bayesian Optimization Setup:
   - Before entering the optimization loop, pass the mean and variance functions of causal priors into the acquisition function to calculate Expected Improvement (EI).
   - For every exploration set, create an intervention input data utilizing the acquisition function. The output data is generated by sampling with the true SEMs. This output data can be sampled multiple times for the same input data, taking into account varying noise levels.
   - For each exploration set, create a BO model at each time step using the mean and variance functions of causal priors.
     - Define a GP Regression model with a mean function from the prior mean and a causal RBF kernel using the prior variance function, allowing hyperparameters to be inferred from the previous time step.
4. Bayesian Optimization Loop:
   - With at least one sample in the intervention dataset, repeat the BO loop for several iterations.
   - For each exploration set, calculate EI using the acquisition function, with mean and variance predicted by the BO model.
   - The acquisition function samples the next intervention input, and the true SEM provides the intervention value (output).
   - Update the BO model with the intervention dataset
   - Update the current best intervention level and value.
5. Outcome: Return the optimal intervention sets and levels for each time step.

C7. Can you think of a possible application of such techniques for a bank?

This method may be useful for banks that want to understand the impact of regulatory changes on outcomes like bank stability, risk-taking behavior, or market liquidity. For example, a bank can compare the changes in outcomes over time between a treatment group (e.g., banks affected by a new regulation) and a control group (e.g., banks not affected by the regulation). This method helps learn the causal effect of the regulation by finding optimal interventions on trends that would occur regardless of the regulatory change.

D. Q3 (Part 2)

Consider this paper: https://arxiv.org/abs/2406.10917 Bayesian Intervention Optimization for Causal Discovery.

D1. Suppose you were a serious reviewer, and a chair professor of a famous university, of this paper of a major ML conference. Do you find any error? Does the paper have any theoretical contribution? If you think they do, are the theoretical results proved?

Below are some identified issues:

- Typo in Heaviside function: Instead of

$$H_\beta(x) = \begin{cases} \exp\left(-\frac{x}{\beta}\right) \text{ if } x < 0 \\ 1 = \exp\left(-\frac{0}{\beta}\right) \text{ if } x \geq 0 \end{cases}$$

  , it should be

$$H_\beta(x) = \begin{cases} \exp\left(-\frac{x}{\beta}\right) \text{ if } x < 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) \text{ if } x \geq 0 \end{cases}$$

- Notation Ambiguity for Random Variables: In several instances, the same notation is used for both random variables and their values. For example, in Appendix A, the expression $\mathbb{E}_{y\sim m_0(y)}[...]$ should be written as $\mathbb{E}_{y\sim m_0(Y)}[...]$. This is because y represents values sampled from a mixture of normal distributions m_0, which is associated with the random variable Y.
- Typo in Section 3.1: the expression $P^i_{DC}(k_0, \mathbf{D}_{\text{int}}, \text{do}(X = x))$ should be corrected to $P^i_{DC}(k_i, \mathbf{D}_{\text{int}}, \text{do}(X = x))$.
- Abuse of Notation for m: The symbol m is overloaded. In Algorithm 1, it is used in the loop `for m = 1,..., M do`, where M represents the intervention size. Simultaneously, m_0 and m_1 are used to denote mixtures of normal distributions.
- Inconsistent notation for the link function in Section 3.2: Under the subsection "Estimation of the Interventional Distribution," the equation $Y = f(X, \gamma) + n_Y$ should be corrected to $Y = f(X, \theta) + n_Y$. Additionally, the symbol γ is not explained until Section 4.
- Optimization in Section 3.2: In the same subsection, the paper states that for $\mathbb{H}_1$, the goal is to estimate $\hat{m}_1 = \max_{\theta,\varphi} \sum_{i=1}^{N_{\text{obs}}} \log m_1(y_i \mid x_i)$. However, the optimization should be performed separately. First, optimize θ using mean squared error (MSE) since it pertains to generalized linear regression, i.e. $\hat{\theta} = \text{argmin}_\theta \|y_i - f(x_i)\|_2^2$. Then, optimize φ, i.e. $\hat{m}_1 = \max_\varphi \sum_{i=1}^{N_{\text{obs}}} \log m_1\left(y_i \mid x_i, \hat{\theta}\right)$. This separation is necessary because if θ does not accurately represent the relationship between X and Y, the residuals $y - f_\theta(x)$ may not follow a normal distribution, thereby affecting the estimation of the mixture of normals.

This paper did not provide significant theoretical contribution:

- The paper derives an approximation for computing the P_{DC} metric. It begins with the statistical definition of P_{DC}, transforms the probability of an event to its expectation form, and finally employs a smooth Heaviside function to approximate the expectation. The smooth Heaviside function is advantageous due to its differentiability, which facilitates optimization and numerical stability.
- However, the paper does not provide a theoretical justification for why this approach is superior to its alternative - information gain. This lack of comparative analysis leaves a gap in understanding the relative merits of the proposed method.

D2. Will you accept or reject the paper? Why? Can you please write a detail review of the paper, point out the pros and cons of the paper, including any possible errors?

Detailed review:

- Main motivation: The paper aims to test two hypotheses:
  - $\mathbb{H}_0$:X and Y have no causal relationship.
  - $\mathbb{H}_1$:X and Y have a causal relationship.

The decision to accept or reject $\mathbb{H}_0$ relies on the posterior probabilities $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}})$ and $P(\mathbb{H}_1 \mid \mathbf{D}_{\text{int}})$. This approach provides a statistically grounded framework for evaluating causal relationships, which is particularly useful in scientific research, such as genetic knockout experiments.

- Main Contributions

  1. Data-Driven Prior in Estimation: The paper introduces a data-driven approach to estimate priors, which is an improvement over traditional methods that rely on arbitrary assumptions.

  2. Use of P_{DC} as a Metric : The paper proposes using P_{DC} as a metric for optimizing intervention levels x_{opt}. This is particularly effective in scenarios involving confounders, where the prior probability of $\mathbb{H}_0/\mathbb{H}_1$ is initially around 50/50 but shifts to 0/100 after intervention. This shift occurs because the original relationship $X = f(U) + \varepsilon_x; Y = f(U) + \varepsilon_y$ is replaced under $\mathbb{H}_1$. When intervention is applied, i.e. $\text{do}(X = x')$), the dataset becomes $\left(x', y = f(u) + \varepsilon_y\right)$ for $u \in U$, no longer modeling the relationship between X and Y.

- Methodology

  The paper leverages Bayesian inference and hypothesis testing to evaluate causal relationships. Key steps include:

  1. The goal is to compute $P(\mathbb{H}_i \mid \mathbf{D}_{\text{int}})$, the probability that hypothesis $\mathbb{H}_i$ holds given the intervention data. This is achieved using Bayes' rule:

  $$P(\mathbb{H}_0 | \, \mathbf{D}_{\text{int}}) = P(\mathbf{D}_{\text{int}} \mid \mathbb{H}_0) \frac{P(\mathbb{H}_0)}{P(\mathbf{D}_{\text{int}})} = \frac{P(\mathbf{D}_{\text{int}} \mid \mathbb{H}_0) P(\mathbb{H}_0)}{\sum_{i \in [0,1]} P(\mathbf{D}_{\text{int}} \mid \mathbb{H}_i) P(\mathbb{H}_i)}$$

  .

  2. Prior Estimation
     - For $\mathbb{H}_0$, a Gaussian mixture model is used to model the distribution of $p(y)$.
     - For $\mathbb{H}_1$, a link function maps x to y, and the residual $y - f(x)$ is modeled using a Gaussian mixture model.

  3. Optimal Intervention Levels The optimal intervention level x_{opt} is determined by optimizing the P_{DC} metric, and its corresponding optimal intervention value y is sampled from the environment. The values of x_{opt} are then used to compute $P(\mathbf{D}_{\text{int}} \mid \mathbb{H}_0)$ and $P(\mathbf{D}_{\text{int}} \mid \mathbb{H}_1)$, which ultimately yield the posterior $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}})$.

- Strengths:

  1. Focus on Hypothesis Testing Unlike previous works that focus on developing models for accurate interventions, this paper emphasizes hypothesis testing. It provides statistical evidence for accepting or rejecting interventions, rather than relying solely on optimization algorithms to maximize prediction accuracy.

  2. Disadvantages of other hypothesis testing methods for causality compared to this method:
     - Granger Causality cannot account for confounding variables
     - Difference-in-Differences (DiD): Does not provide a method for generating intervened data
     - Instrumental Variables method: Requires identifying an instrumental variable and obtaining sufficient data without error terms.

  3. Data driven priors: The use of data-driven priors improves the robustness of the estimation process.

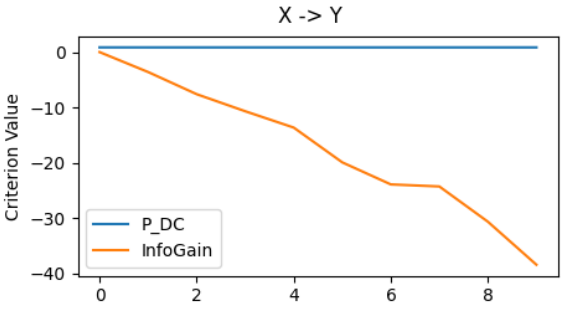
- Weaknesses:
  1. Limited Performance Advantage: The proposed P_{DC} metric does not show a significant performance advantage over its alternative Information Gain, and the computational costs are similar.
  2. Dependence on Link Function: The accuracy of the results depends on the link function used in the structural equation model (SEM). For example, if the true link function is a tanh function but a neural network is used for estimation, the approximation in $\hat{}\{m\}_1$ may be inaccurate, affecting the conclusions.
  3. Single Variable Case The paper focuses on a single-variable scenario, limiting its applicability to multivariable systems.

Decision: I tend to accept. This hypothesis testing methodology is highly suitable for scientific research, though it has a few limitations (such as the knowledge on the link function) that may restrict its practical applicability.

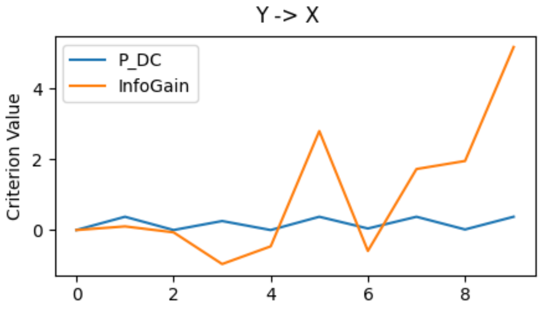
D3. Can you replicate their synthetic data generation results in Tensorflow probability? There is no need to replicate the actual experiment.

The reported prior and posterior values align with expectations. However, unlike what is claimed in the paper, P_{DC} did not demonstrate better convergence results.

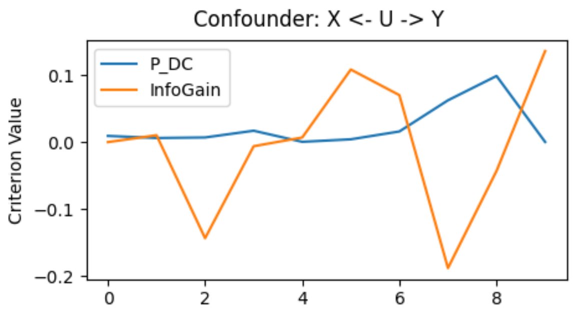
1. $X \rightarrow Y$
   - Computed prior: $P(\mathbb{H}_0) = 0.14866666; P(\mathbb{H}_1) = 0.8513333350419998$
   - Computed posterior: $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}}) = 0.0; P(\mathbb{H}_1 \mid \mathbf{D}_{\text{int}}) = 1.0$



2. $Y \rightarrow X$
   - Computed prior: $P(\mathbb{H}_0) = 0.625; P(\mathbb{H}_1) = 0.375$
   - Computed posterior: $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}}) = 1.0; P(\mathbb{H}_1 \mid \mathbf{D}_{\text{int}}) = 0.0$



3. $X \leftarrow U \rightarrow Y$
   - Computed prior: $P(\mathbb{H}_0) = 0.52633333; P(\mathbb{H}_1) = 0.4736666679382324$
   - Computed posterior: $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}}) = 1.0; P(\mathbb{H}_1 \mid \mathbf{D}_{\text{int}}) = 0.0$



D4. Is it possible to expand the paper to multiple variables? How would you advise the authors to do that? For this multivariable case, is it possible to apply Causal Bayesian Optimization to this Bayesian intervention optimization problem?

Yes, Causal Bayesian Optimization (CBO) is well-suited for identifying optimal intervention levels within structural causal models and can handle exploration sets involving multiple variables. For instance, consider three variables: A, B, and Y.

1. Define the hypotheses:
   - $\mathbb{H}_0 : p(y|\,\text{do}(A = a), \text{do}(B = b)) = p(y)$
   - $\mathbb{H}_1 : p(y\,|\,\text{do}(A = a), \text{do}(B = b)) = p(y|\; a, b)$
2. Specify the link function f as in CBO:
   - For source nodes, apply kernel density estimation.
   - For relationships from parent nodes to child nodes, use Gaussian Processes for modeling.
3. Parametrize m_0 and m_1 using a mixture of normals to approximate the true interventional distribution
   - $\hat{m}_0 = \max_\varphi \sum_{i=1}^{N_{\text{obs}}} \log m_0(y_i)$
   - $\hat{m}_1 = \max_\varphi \sum_{i=1}^{N_{\text{obs}}} \log m_1\left(y_i \mid a_i, b_i, \hat{\theta}\right)$, where $\hat{\theta} = \text{argmin}_\theta \|y_i - f(a_i, b_i)\|_2^2$.
4. With $\mathbb{H}_0$, $\mathbb{H}_1$, $\hat{m}_0$, $\hat{m}_1$ and f defined, Algorithm 1 from the paper can be used to compute $P(\mathbb{H}_0 \mid \mathbf{D}_{\text{int}})$ and $P(\mathbb{H}_1 \mid \mathbf{D}_{\text{int}})$.

D5. If you believe this paper is correct, can you think of a possible application of this paper for a bank?

While the paper is theoretically sound, its practical implementation faces several limitations, such as the need for prior knowledge of the link function and the requirement that the environment supports data sampling. Despite these constraints, the methodology can be applied to risk management. For example, consider a scenario where a bank observes a rise in loan defaults during a period of increasing unemployment. However, it also notices that interest rates were rising simultaneously. Using the hypothesis testing for causality proposed in the paper, the bank could determine whether:

- The increase in defaults was directly caused by higher unemployment.
- The increase in defaults was driven by higher interest rates.
- Both factors contributed independently or interactively.

By isolating the effect of unemployment, the bank can make more informed decisions, such as:

- Tightening credit standards in high-unemployment areas.
- Offering loan restructuring options to borrowers in affected regions.