

Tópicos de Inferência Causal*

Tiago C A Amorim (RA: 100675)^a

^aDoutorando no Departamento de Engenharia de Petróleo da Faculdade de Engenharia Mecânica, UNICAMP, Campinas, SP, Brasil

Keywords: Inferência Causal, Ignorabilidade, Estudos Randomizados, Estudos Observacionais, Controle Negativo

1. Introdução

Este relatório é um resumo de dois artigos na área de inferência causal. O primeiro artigo tem um enfoque mais geralista, e aborda diversos assuntos da área [1]. O segundo artigo foca em controles negativos¹ [2]. Na medida do possível este resumo buscará seguir a notação utilizada ao longo do curso de inferência causal, que segue o livro de Peng Ding [3].

2. Causal Inference in the Social Sciences

Este artigo discute diversos assuntos ligados à avaliação de inferência causal. Será dado maior enfoque às novas técnicas e discussões trazidas pelo autor no lugar de descrever em detalhes metodologias já vistas durante a disciplina MI628: Inferência Causal.

2.1. Estudos Experimentais Randomizados

O autor cita que a análise de estudos experimentais randomizados surge nos anos de 1920, com Fisher e Neyman. Apesar destas ferramentas ainda serem utilizadas atualmente, novas ferramentas foram desenvolvidas. O autor cita que em geral o objetivo de um estudo de inferência causal está mais ligado ao resultado médio do tratamento (visão *neymaniana*) e sua magnitude.

A técnica de **experimentos adaptativos**² foi desenvolvida para tratar de cenários em que existe uma resposta rápida ao tratamento, existem diferentes tratamentos disponíveis, a atribuição das unidades é sequencial e cujo objetivo é descobrir o *melhor* tratamento, e não a magnitude do seu efeito. O autor dá como exemplo a avaliação de qual é a melhor propaganda a ser veiculada online.

Em um estudo tradicional as unidades seriam alocadas de forma igual entre os diferentes tratamentos. São apresentadas duas técnicas de experimentos adaptativos que recalculam a probabilidade de atribuição das unidades aos tratamentos, em função dos resultados alcançados

até determinado momento. Uma técnica usa a probabilidade posterior de determinado tratamento ser o melhor como peso na sua atribuição. A segunda técnica usa o limite superior dos intervalos de confiança do sucesso de um tratamento: **UCB** (*upper confidence bounds*). A próxima unidade é atribuída ao tratamento com o maior UCB. Em ambas abordagens é preciso cuidar do balanço entre exploração e exploração.

Problemas citados pelo autor são viés de inferência ao usar média dos resultados como estimador da esperança dos resultados, efeito de covariáveis na probabilidade de atribuição e dificuldade com respostas que variam com o tempo (estocacidade ou sazonalidade).

Uma outra situação que precisa de tratamento especial é quando as condições de não-interferência e consistência (valor estável de tratamento da unidade ou SUTVA³) não são atendidas. Quando existe a possibilidade de transbordamento ou interferência entre as unidades, a estratégia é definida em função do tipo de transbordamento. Uma estratégia é particionar a população em subgrupos, de forma a limitar os efeitos de transbordamento ao subgrupo. Outra estratégia é com o uso de grafos bipartidos e clusterização aleatória. O último exemplo dado pelo autor é quando duas ou mais populações estão associadas a tratamentos, e as variações de atribuição dos tratamentos nos pares (ou triplas, ...) de indivíduos ajudam a estimar os efeitos de transbordamento.

2.2. Estudos Observacionais sob Ignorabilidade

Em estudos observacionais, assumir que a atribuição de tratamento é basicamente independente (condição de **ignorabilidade** forte: $Z_i \perp (Y_i(0), Y_i(1)) \mid X_i$) leva à possibilidade de analisar os dados como se fossem de um experimento randomizado. Além disto é comum assumir que a probabilidade de atribuição (**propensity score**) é limitada longe de zero e um: $e(x) = pr(Z_i = 1 \mid X_i = x) \in (\delta, 1 - \delta)$, com $\delta > 0$ (**sobreposição**).

A estimativa do efeito médio do tratamento sob ignorabilidade pode ser feita com diferentes abordagens (com comentários do autor):

*Relatório como parte dos requisitos da disciplina MI628: Inferência Causal.

¹Tradução livre de *negative controls*.

²Tradução livre de *adaptive experiments*.

³Do inglês: *Stable Unit Treatment Value Assumption*.

- **Matching:** Estimador intuitivo, mas necessita de significativo número de *matches* e sofre quando existe um número grande de covariáveis.
- **Regressão dos estimadores:** Pode ser difícil de construir bons estimadores quando a dimensão das covariáveis é muito grande.
- **Estimadores com o propensity score:** Pode ser utilizado como uma covariável escalar junto com *matching* ou regressão, ou usar seu inverso como peso na estimativa do efeito causal médio.
- **Estimadores duplamente robustos:** É a classe de estimadores atualmente sugerida na literatura, pois tem boas propriedades mesmo que os estimadores tenham baixa convergência.

Violações de sobreposição podem ser contornadas ignorando amostras que estão muito próximas de zero ou um, ou usando $h(x) = e(x)(1 - e(x))$ como peso no estimador do efeito causal médio.

A estimativa do efeito causal médio condicionado em covariáveis ($\tau(x) = E[Y_i(1) - Y_i(0) \mid X_i = x]$) é uma área de estudo mais recente. Segundo o autor nesta área existem bons resultados com métodos de aprendizado de máquina, como florestas de regressão e florestas aleatórias.

Por último o autor discorre sobre linhas de pesquisa que buscaram estimar o efeito de assumir ignorabilidade em aplicações onde esta não pode ser assumida exatamente. São citadas três abordagens:

- Assumir que não existe ignorabilidade e estimar limites para o efeito médio do tratamento. Esta abordagem pode levar a limites muito grandes.
- Assumir ignorabilidade, mas condicional a uma covariável não observada, que tem limitada associação com os resultados potenciais e a atribuição de tratamento. Esta suposição de limitação é o que leva a restringir os valores possíveis em uma análise de sensibilidade. Uma estratégia sugerida é assumir que o efeito das covariáveis não observadas é menor que o de variáveis observadas.
- Assumir ignorabilidade, mas não limitar a associação entre covariável não observada e resultados potenciais, e limitar as probabilidades de atribuição. A análise de sensibilidade é feita com base nos limites assumidos para as probabilidades de atribuição.

2.3. Estudos Observacionais sem Ignorabilidade

Sem ignorabilidade não é possível descobrir o efeito causal médio, apenas limites em alguns casos. Diferentes ferramentas foram desenvolvidas assumindo algum tipo de informação adicional. Esta informação adicional pode ser uma variável adicional com uma estrutura causal conhecida, ou hipóteses adicionais quanto ao mecanismo de atribuição ou a distribuição dos resultados potenciais.

As **variáveis instrumentais** surgiram nos anos de 1930, como forma de atacar problemas em que pode existir uma covariável não medida que se correlaciona tanto com a atribuição como os resultados. Neste cenário não é possível estimar o efeito causal diretamente entre tratamento e resultados.

Uma variável instrumental é uma variável que se assume ter efeito direto no tratamento, mas não no resultado (Exemplo na Figura 1). Se não existir covariável não medida que correlacione a variável instrumental com o tratamento ou com o resultado, é possível estimar o efeito causal médio do instrumento no resultado.

Para estimar o efeito do tratamento no resultado é preciso assumir monotonicidade, i.e., que o efeito causal do instrumento no tratamento é monotônico. Com estas suposições é possível estimar o efeito causal para a subpopulação em que uma mudança no instrumento de 0 para 1 leva a mudar de controle para tratamento, chamado de LATE⁴.

O autor argumenta que o LATE normalmente não é o resultado de interesse, mas é a única subpopulação na qual é possível estimar o efeito causal médio sem assumir hipóteses fortes. Ele também menciona que a monotonicidade é possível de ser encontrada em diferentes aplicações. Uma dificuldade é quando o efeito da variável instrumental no tratamento é limitado, levando a estimadores ruins.

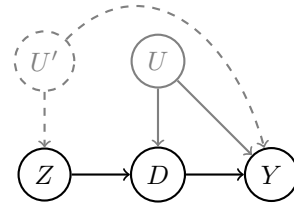


Figura 1: A atribuição de tratamento **Z** funciona como uma variável instrumental do tratamento recebido **D**, pois não tem efeito direto no resultado **Y**. A existência de variável não medida que correlaciona **Z** com **Y** (elementos tracejados) invalida o uso de **Z** como variável instrumental. (Adaptado de [1])

No método das **diferenças em diferenças** original a análise parte dos resultados de unidades de tratamento e controle em dois períodos diferentes: antes do tratamento (ambos grupos não receberam tratamento) e depois do tratamento (apenas um grupo recebe o tratamento). É assumido um efeito aditivo fixo entre os instantes de tempo. Esta ideia foi estendida para trabalhar com múltiplos períodos ou com estudos em que as unidades passam ao grupo de tratamento em períodos distintos (*stepped wedge design*).

No cenário de uma única unidade no grupo de tratamento, a ideia de **controle sintético** é de aproximar o contrafactual da unidade de tratamento por uma combinação convexa das unidades de controle. Os pesos são defini-

⁴De *local average treatment effect*.

dos de forma a melhor aproximar a unidade de tratamento nos tempos anteriores ao tratamento.

Este método já foi estendido para aceitar o controle sintético fora da envoltória convexa das unidades de controle: alguns pesos negativos ou aceitar um fator constante na função objetivo dos pesos. Uma abordagem mais robusta associa diferenças em diferenças com controle sintético, adicionando termos que põe mais ênfase em tempos mais similares ao em que ocorre o tratamento.

O último método citado para avaliar efeito causal onde não é possível assumir ignorabilidade é a **regressão descontinuada**, que começou nas ciências sociais nos anos 1960, e ganhou maior notoriedade e desenvolvimento a partir dos anos 2000. Neste cenário a atribuição de tratamento é definida por um limite (c) de alguma covariável. A descontinuidade *fuzzy* assume que a probabilidade de tratamento não salta de zero para um no limite, mas que existe uma descontinuidade neste ponto.

O estimador mais comum é baseado em regressão linear nas vizinhanças de c . Segundo o autor, estudos mais recentes focam em encontrar funções de peso que melhorem a estimativa do efeito causal como a diferença entre médias ponderadas à esquerda e à direita de c .

Uma dificuldade associada a este método é sua limitada validade externa, pois toda a estimativa é feita com as unidades próximas do limite. No caso *fuzzy* apenas as unidades que estão próximas do limite e que obedecem à separação tratamento-controle são utilizadas⁵.

2.4. Combinação de Estudos Experimentais e Observacionais

O autor cita dois exemplos de técnicas que combinam dados experimentais e observacionais.

A primeira é a de **substituição**⁶, em que se deseja tomar decisões associadas a efeitos de longo-prazo, a partir de medidas de curto-prazo. A ideia é combinar os resultados de curto-prazo (substitutos) em um preditor de resultados de longo-prazo. A relação entre tratamento e os substitutos viria de um estudo experimental, enquanto que a relação entre os substitutos e o resultado de interesse viria de um estudo observacional.

A principal hipótese é que todas as possíveis correlações causais entre o tratamento e o resultado de interesse passam por ao menos um dos substitutos. Desta forma não há efeito causal direto entre o tratamento e a resposta de interesse.

A segunda técnica parte do pressuposto de que no estudo observacional são observados tratamento, resultado secundário e resultado principal, e que no estudo experimental apenas tratamento e resultado secundário são observados. É assumido que existe uma variável de confusão não medida, mas que é a mesma para os resultados secundário e principal.

Com estas hipóteses é possível extrair a covariável não medida dos dados observacionais e experimentais entre tratamento e resposta secundária. E com este resultado é feito o ajuste da estimativa do efeito causal médio do tratamento na resposta principal no estudo observacional.

3. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies

É apresentado o conceito de **controle negativo**. Este mecanismo é utilizado em estudos de biologia experimental, e os autores argumentam que pode ser adaptado para experimentos observacionais e de outras áreas. Associações não-causais podem ser causadas por erro de medição, confusão e viés de seleção de indivíduos para a análise. O uso de controles negativos ajuda a verificar o risco de interpretação não-causal como causal.

A ideia principal do controle negativo é repetir o experimento que atestou efeito causal, mas de forma que o resultado esperado seja a hipótese nula, i.e., sem efeito causal. Vários controles negativos podem ser necessários, em função dos possíveis efeitos não-causais (variáveis de confusão não medidas). A observação de efeito causal no controle negativo é indicativo de que variáveis não medidas estão influenciando os resultados de interesse. Os mecanismos de construção de experimentos de controle negativo citados pelos autores são:

- Retirar um dos elementos essenciais ao mecanismo em avaliação: Testar hipótese de viés de disponibilidade com perguntas *placebo*, que já se sabe não ter relação com o mecanismo avaliado, em um questionário sobre alguma doença.
- Neutralizar o efeito do mecanismo: Avaliar saúde de pacientes que recebem doses ineficazes de algum medicamento.
- Testar com condições em que o mecanismo avaliado não funciona: Índice de hospitalizações por quebra de ossos em função de tomar ou não a vacina da gripe.

Em resumo, o propósito do controle negativo é de testar condições em que o mecanismo avaliado não está ativo, mas que podem sofrer dos mesmos vieses (medidos e não medidos) do experimento original. Seguindo a Figura 2, uma resposta de controle negativo é uma resposta (**N**) que é influenciada pelos mesmos parâmetros que influenciam o tratamento (**Z**) e a resposta de interesse (**Y**). Se as variáveis de confusão não medidas (**U**) forem as mesmas para os pares **Z+Y** e **Z+N**, então **Y** e **N** são ditos *U-comparáveis*.

Assumindo que esta nova resposta (**N**) não é influenciada pelo tratamento (**Z**), e que a resposta de interesse (**Y**) e a nova resposta (**N**) são *U-comparáveis*, se não é observado efeito causal de **Z** em **N**, então podemos afirmar que **provavelmente** um efeito causal observado de **Z** em **Y** não é resultado de um viés desconhecido. Segundo

⁵ *Complier subpopulation* no original em inglês.

⁶ Tradução livre de *surrogacy*.

os autores, na prática não é possível tecer uma afirmação estrita sobre a causalidade de \mathbf{Z} em \mathbf{Y} porque em geral \mathbf{Y} e \mathbf{N} são apenas aproximadamente *U-comparáveis*.

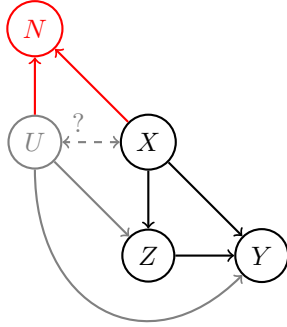


Figura 2: Efeito do tratamento \mathbf{Z} no resultado \mathbf{Y} , com as variáveis de confusão \mathbf{X} (medida) e \mathbf{U} (não medida). A linha tracejada entre \mathbf{X} e \mathbf{U} indica que um pode influenciar o outro. Neste exemplo os resultados \mathbf{N} e \mathbf{Y} são ditos *U-comparáveis*. (Adaptado de [2])

A mesma lógica pode ser aplicada para encontrar um tratamento de controle negativo (\mathbf{B}). Se as variáveis de confusão não medidas (\mathbf{U}) forem as mesmas para os pares $\mathbf{Z}+\mathbf{Y}$ e $\mathbf{B}+\mathbf{Y}$, então \mathbf{Z} e \mathbf{B} são ditos *U-comparáveis*. Novamente, se não existe causalidade de \mathbf{B} em \mathbf{Y} , então uma causalidade observada de \mathbf{Z} em \mathbf{Y} provavelmente não é efeito de um viés desconhecido.

Os autores argumentam que é uma boa prática usar vários controles negativos para ajudar na validação de um estudo que apontou causalidade, e que conhecimento do assunto objeto do estudo é crucial para fazer boas escolhas de controles negativos.

Referências

- [1] G. W. Imbens, Causal inference in the social sciences, Annual Review of Statistics and Its Application 11 (2024).
- [2] M. Lipsitch, E. T. Tchetgen, T. Cohen, Negative controls: a tool for detecting confounding and bias in observational studies, Epidemiology 21 (3) (2010) 383–388.
- [3] P. Ding, A first course in causal inference, arXiv preprint arXiv:2305.18793 (2023).