

Ghi chú bài giảng của khóa học

Giới thiệu về suy luận nhân quả  
từ góc độ học máy

Brady Neal

Ngày 17 tháng 12 năm 2020

## Lời nói đầu

Điều kiện tiên quyết Có một điều kiện tiên quyết chính: xác suất cơ bản. Khóa học này giả định rằng bạn đã tham gia khóa học giới thiệu về xác suất hoặc đã có kinh nghiệm tương đương.

Các chủ đề về thống kê và học máy thịnh thoảng sẽ xuất hiện trong khóa học , vì vậy việc làm quen với những chủ đề đó sẽ hữu ích nhưng không cần thiết. Ví dụ: nếu xác thực chéo là một khái niệm mới đối với bạn, bạn có thể tìm hiểu nó tương đối nhanh chóng tại thời điểm nó xuất hiện trong cuốn sách. Và chúng tôi sẽ giới thiệu sơ lược về một số thuật ngữ thống kê mà chúng tôi sẽ sử dụng trong Phần 2.4.

Bài tập Đọc tích cực Nghiên cứu cho thấy rằng một trong những kỹ thuật tốt nhất để ghi nhớ tài liệu là cố gắng tích cực nhớ lại thông tin mà bạn đã học gần đây. Bạn sẽ thấy “các bài tập đọc tích cực” xuyên suốt cuốn sách để giúp bạn thực hiện điều này. Chúng sẽ được đánh dấu bằng bài [tập Đọc chủ động: tiêu đề](#).

Nhiều nhân vật trong cuốn sách này Như bạn sẽ thấy, có rất nhiều nhân vật trong cuốn sách này. Đây là mục đích. Điều này nhằm giúp bạn có được nhiều trực giác trực quan nhất có thể.

Đôi khi, chúng tôi sẽ sao chép các số liệu, phương trình, v.v. giống nhau mà bạn có thể đã thấy trong các chương trước để có thể đảm bảo rằng các số liệu này luôn ở ngay bên cạnh văn bản tham chiếu đến chúng.

Gửi phản hồi cho tôi Đây là bản nháp của cuốn sách, vì vậy tôi đánh giá rất cao bất kỳ phản hồi nào bạn sẵn sàng gửi cho tôi. Nếu bạn không chắc liệu tôi có chấp nhận điều đó hay không thì đừng làm vậy. Vui lòng gửi bất kỳ phản hồi nào cho tôi theo địa chỉ bradyneal11@gmail.com với “[Sách nhân quả]” ở đầu chủ đề email của bạn. Phản hồi có thể ở cấp độ từ, cấp độ câu, cấp độ phần, cấp độ chương, v.v. Dưới đây là danh sách không đầy đủ các loại phản hồi hữu ích:

Typeoz.

Một số phần là khó hiểu.

Bạn nhận thấy tâm trí mình bắt đầu lang thang hoặc bạn không cảm thấy có động lực để đọc một phần nào đó.

Một số phần có vẻ như có thể cắt được.

Bạn cảm thấy mạnh mẽ rằng tuyệt đối không nên cắt bỏ một phần nào đó.

Một số bộ phận không được kết nối tốt. Chuyển từ phần này sang phần khác, bạn nhận thấy rằng không có một dòng chảy tự nhiên nào cả.

Một bài tập đọc tích cực mới mà bạn nghĩ ra.

Ghi chú thư mục Mặc dù chúng tôi có gắng hết sức để trích dẫn các kết quả có liên quan nhưng chúng tôi không muốn làm gián đoạn dòng chảy của tài liệu bằng cách đào sâu vào chính xác từng khái niệm đến từ đâu. Sẽ có những phần ghi chú thư mục đầy đủ trong phiên bản cuối cùng của cuốn sách này, nhưng chúng sẽ không xuất hiện cho đến khi khóa học kết thúc.

# Nội dung

Lời nói đầu	ii
Nội dung	iii
1 Động lực: Tại sao bạn có thể quan tâm	1
1.1 Nghịch lý của Simpson . . . . .	1
1.2 Ứng dụng của suy luận nhân quả . . . . .	2
1.3 Mối tương quan không ngụ ý quan hệ nhân quả . . . . .	3
1.3.1 Nicolas Lòng và Chết đuối ở bể bơi . . . . .	3
1.3.2 Tại sao Hiệp hội không phải là Nhân quả? . . . . .	4
1.4 Chủ đề chính . . . . .	5
2 Kết quả tiềm năng	6
2.1 Kết quả tiềm năng và tác dụng điều trị riêng lẻ . . . . .	6
2.2 Vấn đề cơ bản của suy luận nhân quả . . . . .	7
2.3 Giải quyết vấn đề cơ bản . . . . .	số 8
2.3.1 Hiệu quả xử lý trung bình và diễn giải dữ liệu bị thiếu . . . . .	số 8
2.3.2 Tính không thể hiểu được và khả năng trao đổi . . . . .	9
2.3.3 Khả năng trao đổi có điều kiện và tính vô căn cứ . . . . .	10
2.3.4 Tích cực/Chồng chéo và ngoại suy . . . . .	12
2.3.5 Không can thiệp, nhất quán và SUTVA . . . . .	13
2.3.6 Liên kết tất cả lại với nhau . . . . .	14
2.4 Thuật ngữ thống kê ưa thích đã được xác định lại . . . . .	15
2.5 Một ví dụ hoàn chỉnh với ước tính . . . . .	16
3 Dòng liên kết và quan hệ nhân quả trong đồ thị	19
3.1 Thuật ngữ đồ thị . . . . .	19
3.2 Mạng Bayes . . . . .	20
3.3 Đồ thị nhân quả . . . . .	22
3.4 Đồ thị hai nút và Khối xây dựng đồ họa . . . . .	23
3.5 Xích và Nĩa . . . . .	24
3.6 Máy va chạm và con cháu của chúng . . . . .	26
3.7 d-tách . . . . .	28
3.8 Dòng Hiệp hội và Nhân quả . . . . .	30
4 mô hình nhân quả	32
4.1 Toán tử do và Phân phối can thiệp . . . . .	32
4.2 Giả định chính: Tính mô đun . . . . .	34
4.3 Hệ số rút gọn . . . . .	35
4.3.1 Ví dụ áp dụng và xem lại "Kết hợp không phải là nhân quả" .	36
4.4 Điều chỉnh cửa sau . . . . .	37
4.4.1 Mối liên hệ với các kết quả tiềm năng . . . . .	39
4.5 Mô hình nhân quả cấu trúc (SCM) . . . . .	40
4.5.1 Phương trình cấu trúc . . . . .	40
4.5.2 Can thiệp . . . . .	42
4.5.3 Thành kiến va chạm và tại sao không có điều kiện đổi với con cháu của việc điều trị .	43
4.6 Các ứng dụng ví dụ về Điều chỉnh cửa sau . . . . .	44
4.6.1 Mối liên hệ và mối quan hệ nhân quả trong một ví dụ về đồ chơi . . . . .	44

4.6.2 Một ví dụ hoàn chỉnh với ước tính . . . . .	45
4.7 Xem lại các giả định . . . . .	47
5 Thí nghiệm ngẫu nhiên 5.1 Khả năng so sánh và cân bằng hiệp phương sai . . . . .	49
5.2 Khả năng trao đổi . . . . .	50
5.3 Không có đường dẫn cửa sau . . . . .	51
6 Nhận dạng phi tham số 6.1 Điều chỉnh cửa trước . . . . .	52
6.2 thực hiện phép tính . . . . .	52
6.3 Xác định khả năng nhận dạng từ đồ thị . . . . .	58
7 Ước tính 7.1	62
Sơ bộ . . . . .	62
7.2 Mô hình hóa kết quả có điều kiện (COM) . . . . .	63
7.3 Mô hình hóa kết quả có điều kiện được nhóm (GCOM) . . . . .	64
7.4 Tăng hiệu quả dữ liệu . . . . .	65
7.4.1 TARNet . . . . .	65
7.4.2 X-Learner . . . . .	66
7.5 Điểm Xu hướng . . . . .	67
7.6 Trọng số xác suất nghịch đảo (IPW) . . . . .	68
7.7 Phương pháp mạnh mẽ gấp đôi . . . . .	70
7.8 Các phương pháp khác . . . . .	70
7.9 Nhận xét kết luận . . . . .	71
7.9.1 Khoảng tin cậy . . . . .	71
7.9.2 So sánh với các thí nghiệm ngẫu nhiên . . . . .	72
8 Nhiều không được quan sát: Phân tích giới hạn và độ nhạy	73
8.1 Giới hạn . . . . .	73
8.1.1 Ràng buộc Không giả định . . . . .	74
8.1.2 Đáp ứng điều trị đơn điệu . . . . .	76
8.1.3 Lựa chọn xử lý đơn điệu . . . . .	78
8.1.4 Lựa chọn phương pháp xử lý tối ưu . . . . .	79
8.2 Phân tích độ nhạy . . . . .	82
8.2.1 Cơ bản về độ nhạy trong cài đặt tuyến tính . . . . .	82
8.2.2 Cài đặt chung khác . . . . .	85
9 biến công cụ	86
9.1 Dụng cụ là gì? . . . . .	86
9.2 Không nhận dạng theo tham số của phần khởi động ATE 9.3: Cài đặt tuyến tính nhị phân . . . . .	87
9.4 Cài đặt tuyến tính liên tục . . . . .	88
9.5 Nhận dạng không tham số của ATE cục bộ	90
9.5.1 Ký hiệu tiềm năng mới với các dụng cụ . . . . .	90
9.5.2 Sự phân tầng chính . . . . .	90
9.5.3 ATE cục bộ . . . . .	91
9.6 Cài đặt chung khác để nhận dạng ATE . . . . .	94
10 Sự khác biệt trong sự khác biệt 10.1 Sơ bộ . . . . .	95

10.2 Thời gian giới thiệu . . . . .	96
10.3 Nhận dạng . . . . .	96
10.3.1 Giả định . . . . .	96
10.3.2 Kết quả chính và chứng minh . . . . .	97
10.4 Các vấn đề chính . . . . .	98
11 Khám phá nhân quả từ dữ liệu quan sát 11.1	100
Khám phá nhân quả dựa trên độc lập . . . . .	100
11.1.1 Giả định và định lý . . . . .	100
11.1.2 Thuật toán PC . . . . .	102
11.1.3 Chúng ta có thể nhận được thông tin nhận dạng nào tốt hơn không? . . . . .	104
11.2 Khám phá nhân quả bán tham số . . . . .	104
11.2.1 Không thể nhận dạng nếu không có giả định tham số . . . . .	105
11.2.2 Nhiều tuyến tính phi Gaussian . . . . .	105
11.2.3 Mô hình phi tuyến . . . . .	108
11.3 Tài nguyên bổ sung . . . . .	109
12 Khám phá nhân quả từ dữ liệu can thiệp 12.1	110
Can thiệp mang tính cấu trúc . . . . .	110
12.1.1 Can thiệp vào một nút . . . . .	110
12.1.2 Can thiệp vào nhiều nút . . . . .	110
12.2 Can thiệp tham số . . . . .	110
12.2.1 Sắp ra mắt . . . . .	110
12.3 Tương đương Markov can thiệp . . . . .	110
12.3.1 Sắp ra mắt . . . . .	110
12.4 Các cài đặt khác . . . . .	110
12.4.1 Sắp ra mắt . . . . .	110
13 Chuyển tiếp học tập và khả năng di chuyển	111
13.1 Hiểu biết sâu sắc về nguyên nhân của việc học chuyển tiếp . . . . .	111
13.1.1 Sắp ra mắt . . . . .	111
13.2 Khả năng vận chuyển của tác động nhân quả giữa các quần thể . . . . .	111
13.2.1 Sắp ra mắt . . . . .	111
14 Phản thực tế và Hòa giải	112
14.1 Cơ bản về phản thực tế . . . . .	112
14.1.1 Sắp ra mắt . . . . .	112
14.2 Ứng dụng quan trọng: Hòa giải . . . . .	112
14.2.1 Sắp ra mắt . . . . .	112
ruột thừa	113
A Bằng chứng	114
A.1 Chứng minh phương trình 6.1 từ Mục 6.1 . . . . .	114
A.2 Bằng chứng về Định lý Điểm Xu hướng (7.1) . . . . .	114
A.3 Bằng chứng về ước tính IPW (7.18) . . . . .	115
Thư mục	117
Danh mục theo thứ tự chữ cái	123

## Danh sách các hình

1.1 Cấu trúc nhân quả về thời điểm nên áp dụng phương pháp điều trị B cho COVID-27 . . . . .	2
1.2 Cấu trúc nhân quả về thời điểm nên áp dụng phương pháp điều trị A cho COVID-27 . . . . .	2
1.3 Số lượng phim của Nicolas Cage tương quan với số vụ chết đúối ở bể bơi . . . . .	3
1.4 Cấu trúc nhân quả với việc được thắp sáng như một yêu tố gây nhiễu . . . . .	4
2.2 Cấu trúc nhân quả của cơ chế phân công điều trị không thể bỏ qua . . . . .	9
2.1 Cấu trúc nhân quả của gây nhiễu tác động của . . . . .	9
lên 2.3 Cấu trúc nhân quả của gây nhiễu . . . . .	11
thông qua 2.4 Cấu trúc nhân quả cho khả năng trao đổi có . . . . .	11
điều kiện đã cho 2.5 Sơ đồ nhận dạng-Ước tính... . . . . .	16
3.3 Đồ thị có hướng . . . . .	19
3.1 Thuật ngữ súng máy . . . . .	19
3.2 Đồ thị vô hướng . . . . .	19
3.4 Đồ thị có hướng có chu trình . . . . .	19
3.5 Đồ thị có hướng có tính vô đạo đức . . . . .	20
3.6 DAG bốn nút trong đó 4 cục bộ chỉ phụ thuộc vào 3 . . . . .	20
3.7 DAG bốn nút với nhiều nút độc lập . . . . .	21
3.8 Hai nút kết nối DAG . . . . .	22
3.9 Các khối xây dựng đồ thị cơ bản . . . . .	24
3.11 Hai nút kết nối DAG . . . . .	24
3.12 Chuỗi có liên kết . . . . .	24
3.10 Hai nút DAG không được kết nối . . . . .	24
3.13 Ngã ba có liên kết . . . . .	25
3.14 Chuỗi có liên kết bị chặn . . . . .	25
3.15 Ngã ba có liên kết bị chặn . . . . .	25
3.16 Vô đạo đức với sự liên kết bị chặn bởi máy va chạm . . . . .	26
3.17 Vô đạo đức với sự liên kết được bỏ chặn . . . . .	26
3.18 Đàm ông đẹp trai là ví dụ điển hình . . . . .	27
3.19 Đồ thị của bài tập tách d . . . . .	30
3.20 Mối liên hệ nhân quả và mối liên kết gây nhiễu . . . . .	30
3.21 Sơ đồ giả định từ độc lập thống kê đến phụ thuộc nhân quả	31
4.1 Sơ đồ nhận dạng-Ước tính (mở rộng) . . . . .	32
4.2 Minh họa sự khác biệt giữa điều hòa và can thiệp . . . . .	33
4.3 Cơ chế nhân quả . . . . .	34
4.4 Can thiệp bằng cách xóa cạnh trong đồ thị nhân quả . . . . .	35
4.5 Cấu trúc nhân quả ứng dụng hệ số rút gọn . . . . .	36
4.6 Đồ thị thao tác cho ba nút . . . . .	37
4.7 Đồ thị phương trình cấu trúc . . . . .	40
4.8 Đồ thị nhân quả của một số phương trình cấu trúc . . . . .	41
4.9 Cấu trúc nhân quả trước can thiệp đơn giản . . . . .	42
4.10 Cấu trúc nhân quả sau can thiệp đơn giản . . . . .	42
4.11 Đồ thị nhân quả chặn hoàn toàn dòng nhân quả . . . . .	43
4.12 Đồ thị nhân quả cho dòng nhân quả chặn một phần . . . . .	43
4.13 Đồ thị nhân quả trong đó máy va chạm có điều kiện gây ra độ lệch . . . . .	43
4.14 Biểu đồ nhân quả trong đó con của người hòa giải bị quy định trên . . . . .	44

4.15 Đồ thị nhân quả phóng to trong đó con của người hòa giải được điều chỉnh theo . . . . .	44
4.16 Đồ thị nhân quả của độ lệch M . . . . .	44
4.17 Đồ thị nhân quả cho ví dụ về đồ chơi . . . . .	45
4.18 Đồ thị nhân quả cho ví dụ về huyết áp bằng máy và chạm . . . . .	46
4.19 Đồ thị nhân quả của độ lệch M với các biến không quan sát được . . . . .	47
5.1 Cấu trúc nhân quả gây nhiễu thông qua . . . . .	51
Cấu trúc nhân quả khi chúng ta phân ngẫu nhiên điều trị . . . . .	51
6.1 Đồ thị nhân quả cho tiêu chí frontdoor . . . . .	52
6.2 Minh họa việc tập trung phân tích vào một người hòa giải . . . . .	52
6.3 Minh họa các bước điều chỉnh cửa trước . . . . .	52
6.5 Phương trình . . . . .	53
6.4 Đồ thị nhân quả cho tiêu chí frontdoor . . . . .	53
6.6 Đồ thị nhân quả cho tiêu chí frontdoor . . . . .	54
6.7 Đồ thị nhân quả cho tiêu chí frontdoor . . . . .	57
6.10 Đồ thị nhân quả cho tiêu chí cửa trước . . . . .	58
6.8 Đồ thị nhân quả cho cửa trước loại bỏ cạnh . . . . .	58
6.9 Đồ thị nhân quả cho cửa trước loại bỏ cạnh . . . . .	58
6.11 Biểu đồ trong đó việc chặn một đường dẫn cửa sau sẽ mở khóa một đường dẫn cửa sau khác . . . . .	59
6.12 Đồ thị mẫu thỏa mãn tiêu chí con không bị nhiễu . . . . .	60
6.13 Đồ thị cho câu hỏi về tiêu chí trẻ em không bị nhầm lẫn . . . . .	61
7.1 Sơ đồ nhận dạng-ước tính . . . . .	63
7.2 Các mạng thần kinh khác nhau dành cho các loại công cụ ước tính khác nhau . . . . .	66
7.3 Đồ thị đơn giản trong đó thỏa mãn tiêu chí cửa sau . . . . .	68
7.5 Biểu đồ đơn giản trong đó làm nhiều tác động của lên . . . . .	68
7.6 Biểu đồ hiệu quả cho quần thể giả mà chúng ta có được bằng cách tính lại trọng số dữ liệu được tạo theo biểu đồ trong Hình 7.5 bằng cách sử dụng trọng số xác suất nghịch đảo. . . . .	68
7.4 Chứng minh bằng đồ thị của định lý điểm xu hướng . . . . .	68
8.1 Đồ thị gây nhiễu không quan sát được . . . . .	73
8.2 Biểu đồ gây nhiễu đơn giản không quan sát được . . . . .	82
8.3 Đồ thị gây nhiễu đơn giản không quan sát được . . . . .	82
8.4 Đồ thị gây nhiễu đơn giản không quan sát được . . . . .	84
8.5 Đồ thị đường viền có độ nhạy gây nhiễu không quan sát được . . . . .	84
9.1 Đồ thị biến công cụ . . . . .	86
9.2 Đồ thị biến công cụ . . . . .	87
9.3 Đồ thị biến công cụ . . . . .	88
9.4 Đồ thị biến công cụ . . . . .	89
9.5 Đồ thị biến công cụ . . . . .	89
9.6 Biểu đồ nhân quả của người tuân thủ và người thách thức . . . . .	91
9.7 Biểu đồ nhân quả cho người luôn nhận và không bao giờ nhận . . . . .	91
11.1 Biểu đồ phản ví dụ về độ tin cậy. . . . .	100
11.3 Sự vô đạo đức Lớp tương đương Markov . . . . .	101
11.2 Ba đồ thị tương đương Markov . . . . .	101
11.5 Biểu đồ hoàn chỉnh. . . . .	102
11.6 Ví dụ về đồ thị thực của PC. . . . .	102
11.4 Bộ xương xích/ngã ba. . . . .	102

11.8 Đồ thị từ PC sau khi chúng ta đã định hướng được những điều trái đạo đức. . . . .	103
11.9 Đồ thị từ PC sau khi chúng ta đã định hướng các cạnh sẽ hình thành các điểm vô đạo đức nếu chúng được định hướng theo hướng khác (không chính xác). . . . .	103
11.7 Minh họa quy trình bước 1 của PC, nơi chúng tôi bắt đầu với phần hoàn thiện đồ thị (trái) và loại bỏ các cạnh cho đến khi chúng tôi xác định được khung của đồ thị (phải), cho rằng đồ thị đúng là đồ thị trong Hình 11.6. . . . .	103
11.10Sự phù hợp tuyến tính của dữ liệu phi Gaussian tuyến tính . . . . .	108
11.11 Phần dư của mô hình tuyến tính phù hợp với dữ liệu phi Gaussian tuyến tính . . . . .	108
A.1 Đồ thị nhân quả cho tiêu chí cửa trước . . . . .	114

## Danh sách các bảng

1.1 Nghịch lý Simpson trong dữ liệu COVID-27 . . . . .	1
2.1 Suy luận nhân quả dưới dạng bài toán thiểu dữ liệu . . . . .	9
3.1 Số mũ của tham số cho các hệ số mô hình . . . . .	20

## Danh sách

2.1 Mã Python để ước tính ATE 2.2 Mã Python để ước tính ATE bằng hệ số hồi quy tuyến tính 17	17
4.1 Mã Python để ước tính ATE mà không cần điều chỉnh máy va chạm . . . . .	46

# Động lực: Tại sao bạn có thể quan tâm

# 1

## 1.1 Nghịch lý của Simpson

Hãy xem xét một tương lai hoàn toàn giả định, nơi có một căn bệnh mới được biết đến như COVID-27 phổ biến trong dân số loài người. Trong việc này hoàn toàn trong tương lai giả định, có hai phương pháp điều trị đã được phát triển: điều trị A và điều trị B. Điều trị B khan hiếm hơn điều trị A, do đó sự phân chia giữa những người hiện đang được điều trị A và điều trị B là khoảng 73%/27%. Bạn có trách nhiệm lựa chọn phương pháp điều trị nào đất nước của bạn sẽ độc quyền sử dụng, ở một đất nước chỉ quan tâm đến giảm thiểu thiệt hại về nhân mạng.

Bạn có dữ liệu về tỷ lệ phần trăm số người chết vì COVID-27, dựa trên phương pháp điều trị mà họ đã được chỉ định và đưa ra tình trạng của họ tại thời gian điều trị đã được quyết định. Điều kiện của chúng là một biến nhị phân: hoặc nhẹ hoặc nặng. Trong dữ liệu này, 16% những người nhận được A chết, trong khi 19% số người nhận B chết. Tuy nhiên, khi chúng tôi kiểm tra những người người bệnh nhẹ tách biệt với người bệnh nặng, các số đảo ngược thứ tự. Ở nhóm dân số nhẹ, 15% số người nhận được A sê chét, trong khi 10% số người nhận được B sê chét. Trong tình trạng nghiêm trọng trong một nhóm nhỏ, 30% số người nhận được thuộc A sê chét, trong khi 20% số người nhận được thuộc B sê chét. Chúng tôi mô tả những tỷ lệ phần trăm này và tương ứng tính ở bảng 1.1.

Tình trạng			
	Nhẹ	Nghiêm trọng	Tổng cộng
MỘT	15%	30%	16%
	(210/1400)	(30/100)	(240/1500)
B	10%	20%	19%
	(5/50)	(100/500)	(105/550)

Nghịch lý rõ ràng này bắt nguồn từ thực tế là, trong Bảng 1.1, "Tổng" cột có thể được hiểu là chúng ta nên ưu tiên điều trị A, trong khi cột "Nhẹ" và "Nặng" đều có thể được hiểu có nghĩa là chúng ta nên ưu tiên cách xử lý B. Trên thực tế, câu trả lời là nếu chúng ta biết tình trạng của ai đó, chúng ta nên điều trị B cho họ, và nếu chúng ta không biết tình trạng của họ, chúng ta nên điều trị cho họ A. Chỉ cần điều đó không có ý nghĩa gì cả. Vậy thực sự nên điều trị như thế nào bạn chọn cho đất nước của bạn?

Phương pháp điều trị A hoặc phương pháp điều trị B có thể là câu trả lời đúng, tùy thuộc vào về cấu trúc nhân quả của dữ liệu. Nói cách khác, quan hệ nhân quả là cần thiết để giải quyết nghịch lý của Simpson. Hiện tại, chúng tôi sẽ chỉ đưa ra trực giác về thời điểm bạn nên thích phương pháp điều trị A hơn khi bạn nên thích phương pháp điều trị B, nhưng nó sẽ được trình bày trang trọng hơn ở Chương 4.

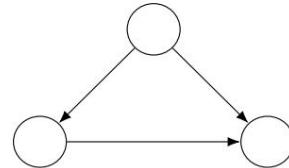
1.1 Nghịch lý của Simpson . . . . .	1
1.2 Ứng dụng của suy luận nhân quả vậy . . . . .	2
1.3 Mối tương quan không ngũ ý Nhân quả . . . . .	3
Nicolas Cage và hòn bơi Đuối nước . . . . .	3
Tại sao Hiệp hội không phải là nguyên nhân vị trí? . . . . .	4
1.4 Chủ đề chính . . . . .	5

Bảng 1.1: Nghịch lý Simpson trong Covid-27 dữ liệu. Tỷ lệ phần trăm biểu thị tỷ lệ tử vong giá ở mỗi nhóm. Thấp hơn là tốt hơn. Các số trong ngoặc đơn là số đếm tương ứng. Điều nghịch lý rõ ràng này bắt nguồn từ việc giải thích rằng phương pháp điều trị A có vẻ tốt hơn khi kiểm tra toàn bộ dân số, nhưng điều trị B có vẻ tốt hơn ở tất cả các quần thể.

<sup>1</sup> Một thành phần quan trọng cần thiết để tìm ra nghịch lý Simpson là tính không đồng nhất của phân bổ người vào các nhóm. 1400 trong số 1500 người được điều trị A có tình trạng nhẹ, trong khi 500 550 người được điều trị B bị bệnh nặng. Bởi vì mọi người với tình trạng nhẹ ít có khả năng tử vong, điều này có nghĩa là tổng tỷ lệ tử vong đối với những người có điều trị A thấp hơn điều gì sẽ xảy ra nếu nhẹ và nặng các điều kiện được chia đều cho họ. Xu hướng ngược lại đúng với điều trị B.

Tình huống 1 Nếu tình trạng  $\alpha$  là nguyên nhân của phương pháp điều trị  $\beta$  (Hình 1.1), phương pháp điều trị  $\beta$  sẽ hiệu quả hơn trong việc giảm tỷ lệ tử vong  $\gamma$ . Một tình huống ví dụ là khi các bác sĩ quyết định áp dụng phương pháp điều trị A cho hầu hết những người mắc bệnh nhẹ. Vì họ tiết kiệm phương pháp điều trị B đắt tiền hơn và hạn chế hơn cho những người mắc bệnh nặng. Bởi vì tình trạng bệnh nặng khiến một người có nhiều khả năng tử vong ( $\alpha$  trong Hình 1.1) và khiến một người có nhiều khả năng nhận được điều trị B ( $\beta$  trong Hình 1.1), điều trị  $\beta$  sẽ có tỷ lệ tử vong cao hơn trong Tổng dân số. Nói cách khác, phương pháp điều trị B có liên quan đến tỷ lệ tử vong cao hơn đơn giản vì tình trạng bệnh là nguyên nhân phổ biến của cả phương pháp điều trị và tỷ lệ tử vong. Ở đây, tình trạng làm ảnh hưởng đến tác dụng của việc điều trị đối với tỷ lệ tử vong. Để khắc phục sự nhầm lẫn này, chúng ta phải kiểm tra mối quan hệ của  $\alpha$  và  $\beta$  giữa những bệnh nhân có cùng tình trạng bệnh lý.

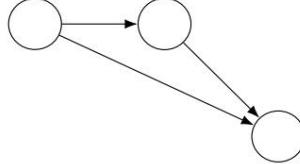
Điều này có nghĩa là phương pháp điều trị tốt hơn là phương pháp mang lại tỷ lệ tử vong thấp hơn ở mỗi nhóm quần thể (cột "Nhẹ" và "Nặng" trong Bảng 1.1): điều trị B.



Hình 1.1: Cấu trúc nhân quả của tình huống 1, trong đó tình trạng  $\alpha$  là nguyên nhân phổ biến của việc điều trị  $\beta$  và tỷ lệ tử vong  $\gamma$ . Với cấu trúc nhân quả này, phương pháp điều trị B được ưu tiên hơn.

Tình huống 2 Nếu đơn thuốc  $\alpha$  điều trị  $\beta$  là nguyên nhân gây ra tình trạng  $\gamma$  (Hình 1.2), thì điều trị A sẽ hiệu quả hơn. Một tình huống ví dụ là phương pháp điều trị B khan hiếm đến mức bệnh nhân phải đợi một thời gian dài sau khi được kê đơn điều trị trước khi có thể được điều trị. Phương pháp điều trị A không gặp phải vấn đề này. Do tình trạng bệnh nhân mắc COVID-19 xâm nhập theo thời gian nên việc kê đơn thuốc điều trị B thực sự khiến bệnh nhân mắc bệnh nhẹ chuyển sang bệnh nặng, gây ra tỷ lệ tử vong cao hơn. Do đó, ngay cả khi điều trị B có hiệu quả hơn điều trị A một lần (tác động tích cực dọc theo  $\beta \rightarrow \gamma$  trong Hình 1.2), do việc kê đơn điều trị B gây ra tình trạng tồi tệ hơn (tác động tiêu cực dọc theo  $\alpha \rightarrow \beta$  trong Hình 1.2), điều trị B nói chung là kém hiệu quả hơn. Lưu ý: Vì điều trị B đắt hơn nên điều trị B được chỉ định với xác suất 0,27, trong khi điều trị A được chỉ định với xác suất 0,73; quan trọng là việc kê đơn điều trị không phụ thuộc vào tình trạng bệnh trong trường hợp này.

2 để cập đến việc chỉ định điều trị, hơn là việc tiếp nhận điều trị sau đó.



Hình 1.2: Cấu trúc nhân quả của kịch bản 2, trong đó việc điều trị  $\alpha$  là nguyên nhân gây ra tình trạng  $\gamma$ . Với cấu trúc nhân quả này, phương án A được ưu tiên hơn.

Tóm lại, việc điều trị hiệu quả hơn hoàn toàn phụ thuộc vào cấu trúc nguyên nhân của vấn đề. Trong Tình huống 1, trong đó  $\alpha$  là nguyên nhân của  $\beta$  (Hình 1.1), phương pháp điều trị  $\beta$  hiệu quả hơn. Trong Tình huống 2, trong đó  $\alpha$  là nguyên nhân của  $\gamma$  (Hình 1.2), phương pháp điều trị  $\alpha$  hiệu quả hơn. Nếu không có quan hệ nhân quả thì nghịch lý Simpson không thể giải quyết được. Với quan hệ nhân quả thì điều đó không hề nghịch lý chút nào.

## 1.2 Ứng dụng của suy luận nhân quả

Suy luận nhân quả là điều cần thiết đối với khoa học, vì chúng ta thường muốn đưa ra những tuyên bố mang tính nhân quả, thay vì chỉ đơn thuần là những tuyên bố mang tính liên kết. Ví dụ, nếu chúng ta đang lựa chọn giữa các phương pháp điều trị cho một căn bệnh, chúng ta muốn chọn phương pháp điều trị giúp nhiều người được chữa khỏi nhất mà không gây ra quá nhiều tác dụng phụ xấu. Nếu chúng ta muốn một thuật toán học tăng cường tối đa hóa phần thưởng, thì chúng ta muốn thuật toán đó thực hiện các hành động để đạt được phần thưởng tối đa. Nếu chúng ta đang nghiên cứu tác động của mạng xã hội đối với sức khỏe tâm thần, chúng ta đang cố gắng hiểu nguyên nhân chính dẫn đến kết quả sức khỏe tâm thần nhất định là gì và sắp xếp các nguyên nhân này theo tỷ lệ phần trăm kết quả có thể được quy cho mỗi nguyên nhân.

Suy luận nhân quả là điều cần thiết cho việc ra quyết định nghiêm ngặt. Ví dụ: giả sử chúng tôi đang xem xét thực hiện một số chính sách khác nhau nhằm giảm phát thải khí nhà kính và chúng tôi chỉ phải chọn một chính sách do hạn chế về ngân sách. Nếu muốn đạt hiệu quả tối đa, chúng ta nên tiến hành phân tích nguyên nhân để xác định chính sách nào sẽ giúp giảm lượng khí thải nhiều nhất. Một ví dụ khác, giả sử chúng ta đang xem xét một số biện pháp can thiệp để giảm nghèo toàn cầu. Chúng tôi muốn biết chính sách nào sẽ giúp giảm nghèo nhiều nhất.

Bây giờ chúng ta đã xem qua ví dụ chung về nghịch lý Simpson và một vài ví dụ cụ thể trong khoa học và ra quyết định, chúng ta sẽ chuyển sang phần suy luận nhân quả khác với dự đoán như thế nào.

### 1.3 Mối tương quan không ngụ ý quan hệ nhân quả

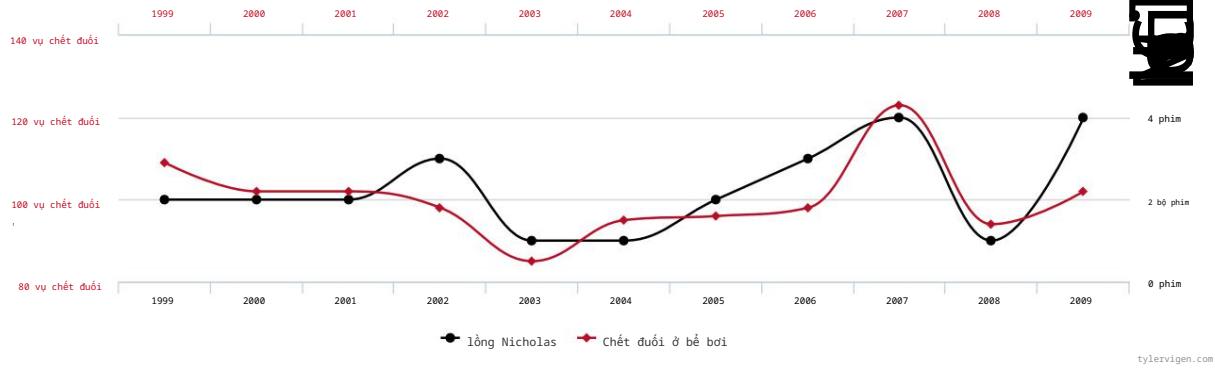
Nhiều người trong số các bạn chắc hẳn đã từng nghe câu thần chú "sự tương quan không hàm ý quan hệ nhân quả". Trong phần này, chúng tôi sẽ nhanh chóng xem xét điều đó và cung cấp cho bạn trực quan hơn một chút về lý do tại sao lại như vậy.

#### 1.3.1 Nicolas Cage và chết đuối ở bể bơi

Hóa ra số người chết đuối do rơi xuống bể bơi hàng năm có mối tương quan cao với số lượng phim mà Nicolas Cage xuất hiện hàng năm [1]. Xem Hình 1.3 để biết biểu đồ của dữ liệu này. Phải chăng điều này có nghĩa là Nicolas Cage khuyến khích những người bơi kém nhảy xuống hồ bơi trong phim của mình? Hay Nicolas Cage cảm thấy có động lực hơn để đóng nhiều phim hơn khi anh thấy có bao nhiêu vụ đuối nước xảy ra trong năm đó, có lẽ là để cô gắng ngăn chặn nhiều vụ đuối nước hơn? Hoặc có một số lời giải thích khác? Ví dụ, có thể Nicolas Cage quan tâm đến việc tăng mức độ phổ biến của anh ấy trong số những người thực hành suy luận nhân quả, vì vậy anh ấy du hành ngược thời gian để thuyết phục bản thân trong quá khứ của mình chỉ làm đủ số lượng phim để chúng ta thấy mối tương quan này, nhưng không quá trùng rơi xuống bể bơi và có khả năng vì điều đó sẽ khơi dậy sự tò mò và giờ là **Có người chết đuối do** tương quan với **liệu theo cách này**. Chúng ta có thể không bao giờ biết chắc chắn.

[1]: Vigen (2015), Tương quan giả

Các phim Nicolas Cage từng xuất hiện



Hình 1.3: Số lượng phim Nicolas Cage xuất hiện hàng năm tương quan với số vụ đuối nước trong bể bơi hàng năm [1].

Tất nhiên, tất cả những lời giải thích có thể có trong đoạn trước dường như khá khó xảy ra. Đúng hơn, có khả năng đây là mối tương quan giả, không có mối quan hệ nhân quả. Chúng ta sẽ sớm chuyển sang phần khác

ví dụ minh họa sẽ giúp làm rõ các mối tương quan giả có thể như thế nào này sinh.

### 1.3.2 Tại sao Hiệp hội không phải là Nhân quả?

Trước khi chuyển sang ví dụ tiếp theo, hãy nói rõ hơn một chút về thuật ngữ. "Tương quan" thường được sử dụng một cách thông tục như một từ đồng nghĩa với sự phụ thuộc thống kê. Tuy nhiên, "sự tương quan" về mặt kỹ thuật chỉ là thước đo sự phụ thuộc thống kê tuyến tính. Phần lớn chúng ta sẽ sử dụng thuật ngữ liên kết để đề cập đến sự phụ thuộc thống kê kể từ bây giờ.

Nhân quả không phải là tất cả hoặc không có. Đối với bất kỳ mức độ liên kết nào, nó không cần phải là "tất cả các liên kết đều là nhân quả" hoặc "không có liên kết nào". sự liên kết là nhân quả." Đúng hơn, có thể có rất nhiều mối liên hệ mà chỉ một số trong đó là quan hệ nhân quả. Cụm từ "sự liên kết không phải là quan hệ nhân quả" chỉ đơn giản có nghĩa là mức độ liên kết và mức độ quan hệ nhân quả có thể khác nhau. Một số lượng liên kết và không có quan hệ nhân quả là một trường hợp đặc biệt của "liên kết không phải là quan hệ nhân quả".

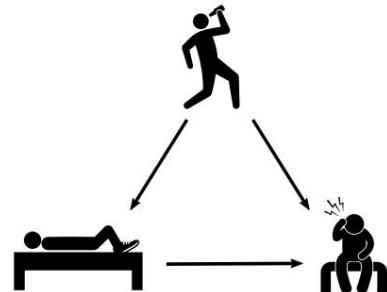
Giả sử bạn tình cờ thấy một số dữ liệu liên quan đến việc mang giày đi ngủ và thức dậy với cơn đau đầu, giống như một trường hợp. Hỏa ra hầu hết khi ai đó đi giày đi ngủ, người đó thức dậy với cơn đau đầu.

Và hầu hết khi ai đó không mang giày đi ngủ, người đó sẽ không thức dậy với cơn đau đầu. Không có gì lạ khi mọi người giải thích những dữ liệu như thế này (có liên tưởng) có nghĩa là việc đi giày đi ngủ khiến mọi người thức dậy với cơn đau đầu, đặc biệt nếu họ đang tìm lý do để biện minh cho việc không đi giày đi ngủ. Một nhà báo thận trọng có thể đưa ra những tuyên bố như "đi giày đi ngủ có liên quan đến đau đầu" hoặc "những người đi giày đi ngủ có nguy cơ thức dậy với cơn đau đầu cao hơn". Tuy nhiên, lý do chính để đưa ra những tuyên bố như vậy là hầu hết mọi người sẽ nội tâm hóa những tuyên bố như thế "nếu tôi mang giày đi ngủ, có thể tôi sẽ thức dậy với cơn đau đầu".

Chúng ta có thể giải thích việc mang giày đi ngủ có liên quan như thế nào với chứng đau đầu mà không phải nguyên nhân nào gây ra nguyên nhân khác. Hỏa ra cả hai đều có nguyên nhân chung: uống rượu vào đêm hôm trước. Chúng tôi mô tả điều này trong Hình 1.4. Bạn cũng có thể nghe thấy loại biến này được gọi là "biến gây nhiễu" hoặc "biến tiềm ẩn". Chúng tôi sẽ gọi loại liên kết này là liên kết gây nhiễu vì sự liên kết này được tạo điều kiện bởi một yếu tố gây nhiễu.

Tổng mối liên hệ được quan sát có thể được tạo thành từ cả mối liên hệ gây nhiễu và mối liên hệ nhân quả. Có thể việc mang giày đi ngủ có một số tác động nhân quả nhỏ khiến bạn thức dậy với cơn đau đầu.

Khi đó, sự liên kết tổng thể sẽ không chỉ là sự liên kết gây nhiễu hay liên kết nhân quả duy nhất. Nó sẽ là sự kết hợp của cả hai. Ví dụ, trong Hình 1.4, mối liên hệ nhân quả diễn ra dọc theo mũi tên từ lúc ngủ quên đến lúc thức dậy với cơn đau đầu. Và sự liên tưởng gây nhiễu diễn ra dọc theo con đường từ ngủ quên trong giày, uống rượu đến đau đầu (thức dậy với cơn đau đầu). Chúng ta sẽ giải thích rõ ràng bằng đồ họa về các loại liên kết khác nhau này trong Chương 3.



Hình 1.4: Cấu trúc nhân quả, trong đó uống rượu vào đêm hôm trước là nguyên nhân phổ biến dẫn đến việc đi ngủ mà vẫn mang giày và thức dậy với cơn đau đầu.

Vấn đề chính Vấn đề chính thúc đẩy suy luận nhân quả là sự liên kết không phải là quan hệ nhân quả.<sup>3</sup> Nếu cả hai đều giống nhau thì suy luận nhân quả sẽ dễ dàng.

Thống kê truyền thống và học máy đã giải quyết được suy luận nhân quả, vì việc đo lường quan hệ nhân quả sẽ đơn giản như chỉ xem xét các biện pháp như mối tương quan và hiệu suất dự đoán trong dữ liệu. Phần lớn cuốn sách này sẽ nói về việc hiểu rõ hơn và giải quyết vấn đề này.

<sup>3</sup> Như chúng ta sẽ thấy trong Chương 5, nếu chúng ta chỉ định ngẫu nhiên phương pháp điều trị trong một thử nghiệm có kiểm soát thì mối liên hệ thực sự là quan hệ nhân quả.

## 1.4 Chủ đề chính

Có một số chủ đề bao quát sẽ tiếp tục xuất hiện xuyên suốt cuốn sách này. Những chủ đề này phần lớn sẽ là sự so sánh của hai loại khác nhau. Khi bạn đọc, điều quan trọng là bạn phải hiểu các phần khác nhau của cuốn sách phù hợp với những danh mục nào và chúng không phù hợp với những danh mục nào.

Thống kê và nhân quả Ngay cả với lượng dữ liệu vô hạn, đôi khi chúng ta không thể tính toán một số đại lượng nhân quả. Ngược lại, phần lớn số liệu thống kê là về việc giải quyết sự không chắc chắn trong các mẫu hữu hạn. Khi được cung cấp dữ liệu vô hạn, không có sự không chắc chắn. Tuy nhiên, sự liên kết, một khái niệm thống kê, không phải là quan hệ nhân quả. Còn nhiều việc phải làm trong suy luận nhân quả, ngay cả sau khi bắt đầu với dữ liệu vô hạn. Đây là sự khác biệt chính thúc đẩy suy luận nhân quả. Chúng tôi đã phân biệt điều này trong chương này và sẽ tiếp tục phân biệt điều này trong suốt cuốn sách.

Xác định so với ước tính Việc xác định các tác động nhân quả là duy nhất đối với suy luận nhân quả. Đó là vấn đề vẫn cần giải quyết, ngay cả khi chúng ta có dữ liệu vô hạn. Tuy nhiên, suy luận nhân quả cũng chia sẻ ước tính với thống kê truyền thống và học máy. Phần lớn chúng ta sẽ bắt đầu bằng việc xác định các tác động nhân quả (trong Chương 2, 4 và 6) trước khi chuyển sang ước tính các tác động nhân quả (trong Chương 7). Các trường hợp ngoại lệ là Phần 2.5 và Phần 4.6.2, trong đó chúng tôi thực hiện các ví dụ hoàn chỉnh kèm theo ước tính để cung cấp cho bạn ý tưởng về toàn bộ quá trình ngay từ đầu.

Can thiệp và quan sát Nếu chúng ta có thể can thiệp/thử nghiệm, việc xác định các tác động nhân quả là tương đối dễ dàng. Điều này đơn giản là vì chúng ta thực sự có thể thực hiện hành động mà chúng ta muốn do lưỡng tác động nhân quả và chỉ đơn giản là đo lưỡng tác động sau khi chúng ta thực hiện hành động đó. Dữ liệu quan sát là nơi nó trở nên phức tạp hơn vì sự nhiễu loạn hầu như luôn được đưa vào dữ liệu.

Giả định Sẽ có sự tập trung lớn vào những giả định mà chúng ta đang sử dụng để đạt được kết quả mà chúng ta nhận được. Mỗi giả định sẽ có hộp riêng để giúp bạn khó có thể không chú ý. Các giả định rõ ràng sẽ giúp bạn dễ dàng nhận ra những lời phê bình về một phân tích nhân quả hoặc mô hình nhân quả nhất định sẽ ở đâu. Người ta hy vọng rằng việc trình bày các giả định một cách rõ ràng sẽ dẫn đến những cuộc thảo luận sáng suốt hơn về quan hệ nhân quả.

# Kết quả tiềm năng

Trong chương này, chúng ta sẽ dễ dàng đi vào thế giới của quan hệ nhân quả. Chúng ta sẽ thấy rằng những khái niệm mới và ký hiệu tương ứng cần được giới thiệu cho mô tả rõ ràng các khái niệm nhân quả. Những khái niệm này là "mới" theo nghĩa rằng chúng có thể không tồn tại trong thống kê hoặc toán học truyền thống, nhưng chúng nên quen thuộc ở chỗ chúng ta sử dụng chúng trong suy nghĩ của mình và mô tả chúng bằng ngôn ngữ tự nhiên mọi lúc.

Ký hiệu thống kê quen thuộc Chúng ta sẽ sử dụng để biểu thị biến ngẫu nhiên có khả năng điều trị, để biểu thị biến ngẫu nhiên cho kết quả của lãi suất và để biểu thị hiệp phương sai. Nói chung, chúng ta sẽ sử dụng chữ hoa các chữ cái để biểu thị các biến ngẫu nhiên (có thể ngoại trừ một trường hợp) và các chữ cái viết thường để biểu thị các giá trị mà các biến ngẫu nhiên đảm nhận. Phần lớn những gì chúng tôi xem xét sẽ có các cài đặt trong đó là nhị phân. Biết rằng, nói chung, chúng tôi có thể mở rộng mọi thứ để hoạt động trong các cài đặt trong đó có thể đảm nhận nhiều hơn hai các giá trị hoặc trong đó là liên tục.

## 2.1 Kết quả tiềm năng và cá nhân

### Tác dụng điều trị

Bây giờ chúng tôi sẽ giới thiệu khái niệm nhân quả đầu tiên xuất hiện trong cuốn sách này. Những khái niệm này đôi khi được coi là duy nhất đối với Mô hình nhân quả Neyman-Rubin [2-4] (hoặc khung kết quả tiềm năng), nhưng họ không như vậy. Ví dụ, những khái niệm tương tự vẫn còn tồn tại (chỉ dưới ký hiệu khác) trong khuôn khổ sử dụng đồ thị nhân quả (Chương 3 và 4). Điều quan trọng là bạn phải dành thời gian để đảm bảo rằng bạn hiểu những khái niệm nhân quả ban đầu này. Nếu bạn chưa học suy luận nhân quả trước đây, chúng sẽ không quen thuộc với toán học bối cảnh, mặc dù chúng có thể khá quen thuộc về mặt trực giác bởi vì chúng ta thường suy nghĩ và giao tiếp bằng ngôn ngữ nhân quả.

Tình huống 1 Hãy xem xét tình huống mà bạn không hài lòng. Và bạn là cân nhắc xem có nên nuôi một chú chó để giúp bạn hạnh phúc hay không. nếu bạn trở nên hạnh phúc sau khi bạn có được con chó, điều này có nghĩa là con chó đã gây ra cho bạn hạnh phúc? Chà, điều gì sẽ xảy ra nếu bạn cũng trở nên hạnh phúc nếu bạn không nhận được con chó? Trong trường hợp đó, con chó không nhất thiết phải khiến bạn hạnh phúc, vì vậy tuyên bố của nó về tác động nhân quả đến hạnh phúc của bạn là yếu.

Kịch bản 2 Hãy thay đổi mọi thứ một chút. Hãy cân nhắc rằng bạn vẫn sẽ hạnh phúc nếu bạn có được một con chó, nhưng bây giờ, nếu bạn không nuôi một con chó, bạn sẽ vẫn ở lại không vui. Trong kịch bản này, con chó có một yêu cầu khá mạnh mẽ về nguyên nhân ảnh hưởng đến hạnh phúc của bạn.

Trong cả hai kịch bản trên, chúng tôi đã sử dụng khái niệm nhân quả được gọi là kết quả tiềm năng. Kết quả của bạn là hạnh phúc: = 1 tương ứng với hạnh phúc trong khi = 0 tương ứng với không hạnh phúc. Việc điều trị của bạn là liệu hay không bạn nhận được một con chó: = 1 tương ứng với việc bạn nhận được một con chó trong khi = 0

2.1 Kết quả tiềm năng và chỉ số Hiệu ứng điều trị individual .	6
2.2 Vấn đề cơ bản của suy luận nhân quả . . . . .	7
2.3 Đi vòng quanh Quỹ- vẫn đề tâm thần . . . . .	số 8
Hiệu quả điều trị trung bình và Phiên dịch dữ liệu bị thiếu . . . . .	số 8
Khả năng bỏ qua và khả năng trao đổi . . . . .	9
Khả năng trao đổi có điều kiện và Sự vô căn cứ .	10
Tích cực/Chồng chéo và Ngoại suy . . . . .	12
Không có sự can thiệp, tính nhất quán và SUTVA . . . . .	13
Buộc tất cả lại với nhau . . . . .	14
2.4 Thuật ngữ thống kê ưa thích Bị phá sản . . . . .	15
2.5 Một ví dụ hoàn chỉnh với Ước tính . . . . .	16

[2]: Splawa-Neyman (1923 [1990]), 'Trên  
Ứng dụng lý thuyết xác suất vào các thí  
nghiệm nông nghiệp. Tiểu luận về Nguyên tắc.  
Phần 9.'

[3]: Rubin (1974), 'Ước tính tác động nhân quả  
của các phương pháp điều trị trong các nghiên cứu ngẫu  
nhân và không ngẫu nhiên.'

[4]: Sekhon (2008), 'Neyman-Rubin  
Mô hình suy luận và ước lượng nhân quả  
qua Phương pháp so khớp'

tương ứng với việc bạn không nuôi chó. Chúng tôi biểu thị bằng (1) kết quả hạnh phúc tiềm ẩn mà bạn sẽ quan sát được nếu bạn nuôi một con chó ( $= 1$ ). Tương tự, chúng tôi biểu thị bằng (0) kết quả hạnh phúc tiềm ẩn mà bạn sẽ quan sát được nếu bạn không nuôi một con chó ( $= 0$ ). Trong kịch bản 1, (1) = 1 và (0) = 1. Ngược lại, trong kịch bản 2, (1) = 1 và (0) = 0.

Tổng quát hơn, kết quả tiềm năng ( ) biểu thị kết quả của bạn sẽ như thế nào nếu bạn thực hiện điều trị . Kết quả tiềm năng ( ) khác với kết quả được quan sát ở chỗ không phải tất cả các kết quả tiềm năng đều được quan sát. Đúng hơn là tất cả các kết quả tiềm năng đều có thể được quan sát thấy. Giá trị thực sự được quan sát phụ thuộc vào giá trị mà việc xử lý đảm nhận.

Trong các tình huống trước, chỉ có một cá thể duy nhất trong toàn bộ quần thể bạn. Tuy nhiên, nhìn chung, có rất nhiều cá nhân trong quần thể được quan tâm<sup>1</sup>. Chúng tôi sẽ biểu thị cách xử lý, đồng biến và kết quả của cá nhân thứ bằng cách sử dụng , và . Sau đó, chúng ta có thể xác định hiệu quả điều trị riêng lẻ (ITE)<sup>2</sup> cho cá nhân :

(1) (0) (2.1)

Bất cứ khi nào có nhiều hơn một cá thể trong một quần thể, ( ) là một biến ngẫu nhiên vì các cá thể khác nhau sẽ có kết quả tiềm năng khác nhau. Ngược lại, ( ) thường được coi là không ngẫu nhiên3 vì chỉ số dưới có nghĩa là chúng ta đang điều chỉnh quá nhiều thông tin cá nhân hóa (và theo ngữ cảnh cụ thể), đến mức chúng ta hạn chế tập trung vào một cá nhân duy nhất (trong một bối cảnh) có kết quả tiềm năng mang tính quyết định.

<sup>1</sup> “Đơn vị” thường được dùng thay cho “cá nhân” vì đơn vị của dân số không phải lúc nào cũng là con người.

2 ITE còn được gọi là tác động nhân quả cá nhân, tác động nhân quả ở cấp độ đơn vị hoặc tác động điều trị ở cấp độ đơn vị.

<sup>3</sup> Tuy nhiên, ( ) có thể được coi là ngẫu nhiên.

ITE là một số đại lượng chính mà chúng ta quan tâm trong suy luận nhân quả. Ví dụ, trong kịch bản 2 ở trên, bạn sẽ chọn nuôi một con chó vì tác động nhân quả của việc nuôi một con chó đối với hạnh phúc của bạn là tích cực: (1) (0)  
 $= 1 \quad 0 = 1$ . Ngược lại, trong kịch bản 1, bạn có thể chọn không nuôi chó vì việc nuôi chó không có tác động nhân quả nào đến hạnh phúc của bạn: (1) (0)  
 $= 1 \quad 1 = 0$ .

Bây giờ chúng tôi đã giới thiệu các kết quả tiềm năng và ITE, chúng tôi có thể giới thiệu các vấn đề chính này sinh trong suy luận nhân quả không xuất hiện trong các lĩnh vực mà trọng tâm chính là liên kết hoặc dự đoán.

## 2.2 Vấn đề căn bản về nhân quả

## Sự suy luận

Không thể quan sát được tất cả các kết quả có thể xảy ra đối với một cá nhân nhất định [3]. Hãy xem xét ví dụ về con chó. Bạn có thể quan sát (1) bằng cách nuôi một con chó và quan sát niềm hạnh phúc của bạn sau khi có được một con chó. Ngoài ra, bạn có thể quan sát (0) bằng cách không nuôi chó và quan sát hạnh phúc của bạn. Tuy nhiên, bạn không thể quan sát cả (1) và (0), trừ khi bạn có cỗ máy thời gian cho phép bạn quay ngược thời gian và chọn phương pháp điều trị mà bạn đã không thực hiện lần đầu tiên. Bạn không thể đơn giản bắt một con chó, quan sát (1), đưa con chó đi và sau đó quan sát (0) vì quan sát thứ hai sẽ bị ảnh hưởng bởi tất cả các hành động bạn thực hiện giữa hai lần quan sát và bất kỳ điều gì khác đã thay đổi kể từ lần quan sát đầu tiên quan sát.

[3]: Rubin (1974), 'Ước tính tác động nhân quả của các phương pháp điều trị trong các nghiên cứu ngẫu nhiên và không ngẫu nhiên.'

Đây được gọi là vấn đề cơ bản của suy luận nhân quả [5]. Nó cơ bản vì nếu chúng ta không thể quan sát được cả (1) và (0), thì chúng ta không thể quan sát được hệ quả nhân quả (1) (0). Vấn đề này là duy nhất đối với suy luận nhân quả bởi vì, trong suy luận nhân quả, chúng ta quan tâm đến việc đưa ra các tuyên bố nhân quả, được xác định theo các kết quả tiềm năng. Ngược lại, hãy xem xét học máy. Trong học máy, chúng ta thường chỉ quan tâm đến việc dự đoán kết quả quan sát được, do đó không cần đến kết quả tiềm năng, điều đó có nghĩa là học máy không phải giải quyết vấn đề cơ bản mà chúng ta phải giải quyết trong suy luận nhân quả.

[5]: Holland (1986), 'Thống kê và nhân quả Suy suy luận'

Những kết quả tiềm ẩn mà bạn không (và không thể) quan sát được được gọi là phản thực tế vì chúng đi ngược lại với thực tế (thực tế). " Các kết quả tiềm năng" đôi khi được gọi là "các kết quả phản thực tế", nhưng chúng tôi sẽ không bao giờ làm điều đó trong cuốn sách này vì một kết quả tiềm năng ( ) không phản tác dụng với thực tế cho đến khi một kết quả tiềm năng khác ( ) được quan sát thấy. Kết quả tiềm năng được quan sát đôi khi được gọi là kết quả thực tế. Lưu ý rằng không có phản thực tế hoặc thực tế nào cho đến khi kết quả được quan sát. Trước đó, chỉ có những kết quả tiềm năng .

## 2.3 Giải quyết vấn đề cơ bản

Tôi nghĩ ngờ phần này là nơi chương này có thể bắt đầu hơi không rõ ràng. Nếu trường hợp của bạn là như vậy thì đừng quá lo lắng mà hãy tiếp tục đọc chương tiếp theo vì nó sẽ xây dựng các khái niệm song song theo một cách hy vọng là trực quan hơn.

### 2.3.1 Hiệu quả điều trị trung bình và giải thích dữ liệu bị thiếu

Chúng tôi biết rằng chúng tôi không thể tiếp cận các hiệu quả điều trị riêng lẻ, nhưng còn hiệu quả điều trị trung bình thì sao? Chúng ta có được hiệu quả điều trị trung bình (ATE)<sup>4</sup> bằng cách lấy giá trị trung bình của ITE:

<sup>4</sup> ATE còn được gọi là "giá trị trung bình tác động nhân quả (ACE).

$$[ (1) \quad (0)] = [ (1) \quad (0)], \quad (2.2)$$

trong đó giá trị trung bình lớn hơn số cá thể nếu ( ) có tính xác định. Nếu ( ) là ngẫu nhiên thì giá trị trung bình cũng cao hơn bất kỳ ngẫu nhiên nào khác.

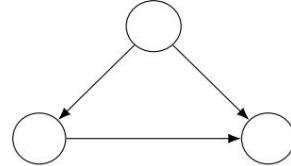
Được rồi, nhưng chúng ta thực sự tính toán ATE như thế nào? Hãy xem xét một số dữ liệu tổng hợp trong Bảng 2.1 để biết điều này. Nếu bạn thích các ví dụ, vui lòng thay thế bằng ví dụ về COVID-27 từ Phần 1.1 hoặc ví dụ về hạnh phúc của loài chó từ Phần 2.1. Chúng ta sẽ coi bảng này là toàn bộ tổng thể được quan tâm. Do vấn đề cơ bản của suy luận nhân quả nên về cơ bản đây là vấn đề thiếu dữ liệu. Tất cả các dấu chấm hỏi trong bảng biểu thị rằng chúng ta không quan sát được ô đó.

Một đại lượng tự nhiên mà tôi nghĩ đến là hiệu liên kết:  $[ | | = 1] \quad [ | | = 0]$ . Theo tính tuyến tính của kỳ vọng, chúng ta có ATE  $[ (1) \quad (0)] = [ (1) \quad (0)]$ . Khi đó, có thể  $[ (1)] \quad [ (0)]$  bằng  $[ | | = 1] \quad [ | | = 0]$ . Thật không may, điều này nói chung không đúng.

Nếu đúng như vậy thì điều đó có nghĩa là nhân quả chỉ đơn giản là sự liên kết.  $[ | | = 1] \quad [ | | = 0]$  là đại lượng liên kết, trong khi  $[ (1)] \quad [ (0)]$

(1)	(0)	(1)	(0)
1 0 0 ? 2 1 1 1	0		?
3 1 0 0 4 0 0 ? 5			?
0 1 ? 6 1 1 1			?
			?
			?
			?

Bảng 2.1: Dữ liệu ví dụ minh họa điều đó vấn đề cơ bản của suy luận nhân quả có thể được hiểu là thiếu dữ liệu vấn đề.



Hình 2.1: Cấu trúc nhân quả của làm ảnh hưởng đến lên.

là một đại lượng nhân quả. Chúng không bằng nhau do gây nhiễu, mà chúng tôi được thảo luận ở Phần 1.3. Việc giải thích bằng đồ họa về điều này, được mô tả trong Hình 2.1, có phải làm xáo trộn tác dụng của lên vì có cái này  
con đường mà sự liên kết phi nhân quả chảy dọc theo.

### 2.3.2 Tính dễ bị bỏ qua và khả năng trao đổi

Chà, (những) giả định nào sẽ khiến ATE chỉ đơn giản là sự khác biệt liên kết? Điều này tương đương với việc nói “điều gì làm cho nó hợp lệ để tính ATE bằng cách lấy giá trị trung bình của cột (0), bỏ qua các dấu chấm hỏi và trừ đi số đó từ mức trung bình của (1) cột, bỏ qua các dấu chấm hỏi?”<sup>6</sup> Việc bỏ qua câu hỏi này dấu hiệu (thiếu dữ liệu) được gọi là không thể biết được. Giả sử sự thiếu hiểu biết là như phớt lờ việc mọi người cuối cùng đã lựa chọn phương pháp điều trị mà họ đã chọn và chỉ giả sử rằng họ được chỉ định điều trị một cách ngẫu nhiên; chúng tôi mô tả điều này bằng đồ họa trong Hình 2.2 do thiếu mui tên nhân quả từ tôi. Bây giờ chúng tôi sẽ tuyên bố giả định này một cách chính thức.

Giả định 2.1 (Có thể bỏ qua/Có thể trao đổi)

$$(\quad (1), \quad (0))$$

Giả định này là chìa khóa cho suy luận nhân quả vì nó cho phép chúng ta giảm ATE đến sự khác biệt liên kết:

$$[ \quad (1) ] \quad [ \quad (0) ] = \quad [ \quad (1) \mid \quad = 1] \quad [ \quad (0) \mid \quad = 0] \quad (2.3)$$

$$= [ \quad | \quad = 1] \quad [ \quad | \quad = 0] \quad (2.4)$$

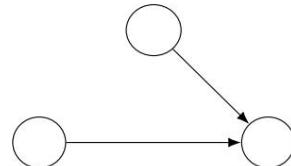
Giả định về khả năng bỏ qua được sử dụng trong Công thức 2.3. Chúng ta sẽ nói chuyện nhiều hơn về Phương trình 2.4 khi chúng ta đến Phần 2.3.5.

Một quan điểm khác về giả định này là khả năng trao đổi. Khả năng thay đổi có nghĩa là các nhóm điều trị có thể được trao đổi trong cảm giác rằng nếu chúng được đổi chỗ, nhóm điều trị mới sẽ quan sát kết quả tương tự như nhóm điều trị cũ và nhóm điều trị mới nhóm kiểm soát sẽ quan sát các kết quả tương tự như nhóm kiểm soát cũ nhom. Về mặt hình thức, giả định này có nghĩa là  $[ (1) | = 0 ] = [ (1) | = 1 ]$  và  $[ (0) | = 1 ] = [ (0) | = 0 ]$ , tương ứng. Sau đó, điều này ngụ ý  $[ (1) | = ] = [ (1) ]$  và  $[ (0) | = ] = [ (0) ]$ , với mọi , tức là gần tương đương7 Giả sử 2.1.

Một trực giác quan trọng cần có về khả năng trao đổi là nó đảm bảo rằng các nhóm điều trị có thể so sánh được. Nói cách khác, việc điều trị các nhóm đều giống nhau về tất cả các khía cạnh liên quan ngoại trừ việc điều trị. Cái này trực giác là nền tảng của khái niệm “kiểm soát” hoặc “điều chỉnh”

5 Hãy tiếp tục đọc tới Chương 3, nơi chúng ta sẽ bổ sung và chính thức hóa đồ họa này diễn dịch.

6 **Bài tập đọc tích cực:** xác minh rằng điều này thủ tục tương đương với [ ] = 1  
[ ] = 0 trong dữ liệu ở Bảng 2.1.



Hình 2.2: Cấu trúc nhân quả khi  
cơ chế chỉ định điều trị là không thể chấp nhận  
được. Đáng chú ý, điều này có nghĩa là không có  
mối tên từ đến , nghĩa là có  
không gây nhầm lẫn.

7 Về mặt kỹ thuật, đây là khả năng trao đổi trung bình, là một giả định yếu hơn so với giả định khả năng trao đổi đầy đủ mà chúng tôi mô tả trong Giả định 2.1 vì nó chỉ hạn chế thời điểm đầu tiên của quá trình phân phối. Nói chung là, chúng ta chỉ cần khả năng có thể bỏ qua/trao đổi trung bình để có được hiệu quả điều trị trung bình, nhưng nó người ta thường giả định sự độc lập hoàn toàn, như trong Giả định 2.1.

for" các biến mà chúng ta sẽ thảo luận ngay khi đề cập đến khả năng trao đổi có điều kiện.

Chúng tôi đã tận dụng Giả định 2.1 để xác định các tác động nhân quả. Xác định tác động nhân quả là giảm thiểu thức nhân quả thành biểu thức thông kê thuận túy. Trong chương này, điều đó có nghĩa là giảm thiểu thức từ biểu thức sử dụng ký hiệu kết quả tiềm năng thành biểu thức chỉ sử dụng ký hiệu thống kê như , , , kỳ vọng và điều kiện. Điều này có nghĩa là chúng ta có thể tính toán tác động nhân quả chỉ từ phân phối quan sát ( , , ).

**Định nghĩa 2.1 (Khả năng nhận dạng)** Một đại lượng nhân quả (ví dụ [ ( )]) có thể xác định được nếu chúng ta có thể tính toán nó từ một đại lượng thống kê thuận túy (ví dụ [ | ]).

Chúng ta đã thấy rằng tính không thể biết được là cực kỳ quan trọng (Công thức 2.3), nhưng giả định đó thực tế đến mức nào? Nói chung, nó hoàn toàn không thực tế vì có khả năng xảy ra nhiều trong hầu hết dữ liệu chúng ta quan sát được (cấu trúc nhân quả được trình bày trong Hình 2.1). Tuy nhiên, chúng ta có thể biến giả định này thành hiện thực bằng cách thực hiện các thử nghiệm ngẫu nhiên, buộc việc xử lý không được gây ra bởi bất cứ điều gì ngoại trừ việc tung đồng xu, do đó chúng ta có cấu trúc nhân quả như trong Hình 2.2. Chúng tôi đề cập sâu hơn về các thí nghiệm ngẫu nhiên trong Chương 5.

Chúng tôi đã đề cập đến hai quan điểm nổi bật về giả định chính (2.1) này: tính không thể biết được và khả năng trao đổi. Về mặt toán học, những điều này có nghĩa giống nhau, nhưng tên của chúng tương ứng với những cách suy nghĩ khác nhau về cùng một giả định. Khả năng trao đổi và khả năng bỏ qua chỉ là hai tên cho giả định này. Chúng ta sẽ thấy nhiều bí danh hơn sau khi đề cập đến phiên bản thực tế hơn, có điều kiện hơn của giả định này.

### 2.3.3 Khả năng trao đổi có điều kiện và tính vô căn cứ

Trong dữ liệu quan sát, sẽ không thực tế nếu cho rằng các nhóm can thiệp có thể thay đổi được. Nói cách khác, không có lý do gì để kỳ vọng rằng các nhóm đều giống nhau ở tất cả các biến liên quan ngoài cách xử lý. Tuy nhiên, nếu chúng ta kiểm soát các biến liên quan bằng điều hòa thì có thể các nhóm con sẽ có thể thay đổi được. Chúng ta sẽ làm rõ "các biến liên quan" là gì trong Chương 3, nhưng bây giờ, hãy tạm coi chúng đều là các biến đồng thời . Sau đó, chúng ta có thể nêu khả năng trao đổi có điều kiện một cách chính thức.

**Giả định 2.2 (Khả năng trao đổi có điều kiện / Tính vô căn cứ)**

( (1), (0)) |

Ý tưởng là mặc dù việc điều trị và các kết quả tiềm ẩn có thể liên quan vô điều kiện (do gây nhiễu), trong phạm vi mức , chúng không liên quan. Nói cách khác, không có sự nhiễu loạn trong các mức vì việc kiểm soát đã làm cho các nhóm can thiệp có thể so sánh được. Bây giờ chúng ta sẽ cung cấp một chút trực giác đồ họa cho phần trên. Chúng ta sẽ không rút ra mối liên hệ chặt chẽ giữa trực quan đồ họa và Giả định 2.2 cho đến Chương 3; hiện tại, nó chỉ nhằm mục đích hỗ trợ trực giác.

Chúng tôi không có khả năng trao đổi dữ liệu vì là nguyên nhân phổ biến của và . Chúng tôi minh họa điều này trong Hình 2.3. Vì là nguyên nhân chung của và nên có mối liên hệ phi nhân quả giữa và . Sự liên kết phi nhân quả này chảy dọc theo con đường ; chúng tôi mô tả điều này bằng một vòng cung nét đứt màu đỏ.

Tuy nhiên, chúng tôi có khả năng trao đổi có điều kiện trong dữ liệu. Điều này là do, khi chúng ta điều kiện hóa , không còn bất kỳ mối liên hệ phi nhân quả nào giữa và . Sự liên kết vô nhân quả hiện đã bị "chặn" tại bằng cách điều hòa . Chúng tôi minh họa việc chặn này trong Hình 2.4 bằng cách tô bóng để biểu thị nó được điều chỉnh và bằng cách hiển thị vòng cung nét đứt màu đỏ bị chặn ở đó.

Khả năng trao đổi có điều kiện là giả định chính cần thiết cho suy luận nhân quả. Với giả định này, chúng ta có thể xác định tác động nhân quả trong phạm vi các mức , giống như chúng ta đã làm với khả năng trao đổi (vô điều kiện):

$$\begin{aligned} [ (1) \quad (0) | ] &= [ (1) | ] \quad [ (0) | ] \quad (2.5) \\ &= [ (1) | = 1, ] \quad [ (0) | = 0, ] \quad (2.6) = \\ &[ | = 1, ] \quad [ | = 0, ] \quad (2.7) \end{aligned}$$

Song song với trước, ta thu được phương trình 2.5 bằng tính tuyến tính của kỳ vọng. Và bây giờ chúng ta có được phương trình 2.6 bằng khả năng trao đổi có điều kiện. Nếu chúng ta muốn hiệu ứng cận biên mà chúng ta đã có trước đây khi giả định khả năng trao đổi (vô điều kiện), chúng ta có thể đạt được điều đó bằng cách loại trừ :

$$\begin{aligned} [ (1) \quad (0) ] &= [ (1) \quad (0) | ] \quad (2.8) \\ &= [ [ | = 1, ] \quad [ | = 0, ] ] \quad (2.9) \end{aligned}$$

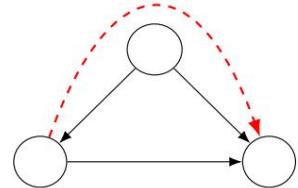
Điều này đánh dấu một kết quả quan trọng cho suy luận nhân quả, vì vậy chúng ta sẽ đặt cho nó một命题 riêng. Bằng chứng chúng tôi đưa ra ở trên bù qua một số chi tiết. Đọc qua Phần 2.3.6 (nơi chúng tôi làm lại bằng chứng với tất cả các chi tiết được chỉ định) để biết các chi tiết còn lại. Chúng ta sẽ gọi kết quả này là công thức điều chỉnh .

**Định lý 2.1 (Công thức điều chỉnh)** Với các giả định về tính không nhiễu, tích cực, nhất quán và không can thiệp, chúng ta có thể xác định được hiệu quả can thiệp trung bình:

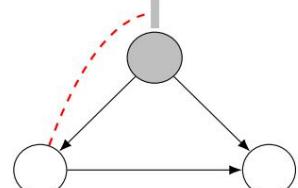
$$[ (1) \quad (0) ] = [ [ | = 1, ] \quad [ | = 0, ] ]$$

Khả năng trao đổi có điều kiện (Giả định 2.2) là giả định cốt lõi cho suy luận nhân quả và có nhiều tên gọi. Ví dụ: những từ sau đây được sử dụng phổ biến một cách hợp lý để đề cập đến cùng một giả định: tính không bị nhầm lẫn, tính không thể hiểu được có điều kiện, không gây nhiễu không được quan sát, lựa chọn dựa trên các yếu tố có thể quan sát được, không có sai lệch biến bị bỏ qua, v.v. Chúng tôi sẽ sử dụng khái niệm "không có căn cứ" xuyên suốt Cuốn sách này.

Lý do chính để chuyển từ khả năng trao đổi (Giả định 2.1) sang khả năng trao đổi có điều kiện (Giả định 2.2) là nó có vẻ giống như một giả định thực tế hơn. Tuy nhiên, chúng ta thường không thể biết chắc chắn liệu khả năng trao đổi có điều kiện có được giữ hay không. Có thể có một số yếu tố gây nhiễu không được quan sát và không thuộc , nghĩa là khả năng trao đổi có điều kiện bị vi phạm. May mắn thay, đó không phải là vấn đề trong các thí nghiệm ngẫu nhiên



Hình 2.3: Cấu trúc nhân quả của kết hợp với tác động của lên . Chúng tôi mô tả sự gây nhiễu bằng một đường nét đứt màu đỏ.



Hình 2.4: Minh họa điều hòa trên dẫn đến không gây nhiễu.

(Chương 5). Thật không may, đó là điều mà chúng ta luôn phải lưu ý trong dữ liệu quan sát. Theo trực giác, điều tốt nhất chúng ta có thể làm là quan sát và ghép càng nhiều đồng biến vào càng tốt để có gắng đảm bảo tính vô căn cứ.

Như chúng ta sẽ thấy trong Chương 3 và 4, không nhất thiết việc điều kiện hóa nhiều đồng biến hơn luôn giúp ích cho quan hệ nhân quả của chúng ta. Ước lượng ít sai lệch hơn.

### 2.3.4 Tích cực/Chồng chéo và ngoại suy

Mặc dù việc quy định nhiều biến đồng thời là điều hấp dẫn để đạt được tính vô căn cứ, nhưng nó thực sự có thể gây bất lợi vì một lý do khác liên quan đến một giả định quan trọng khác mà chúng ta vẫn chưa thảo luận: tích tích cực. Chúng ta sẽ biết lý do tại sao ở cuối phần này. Tích tích cực là điều kiện mà tất cả các nhóm con của dữ liệu với các biến số khác nhau có một số xác suất nhận được bất kỳ giá trị can thiệp nào. Về mặt hình thức, chúng tôi định nghĩa tích tích cực đối với xử lý nhị phân như sau.

**Giả định 2.3 (Dương/Chồng chéo/Hỗ trợ chung)** Đối với tất cả các giá trị của hiệp phương sai có trong tổng thể quan tâm (tức là sao cho  $(\cdot = \cdot) > 0$ ),

$$0 < (\cdot = 1 | \cdot = \cdot) < 1$$

Để hiểu tại sao tích tích cực lại quan trọng, chúng ta hãy xem xét kỹ hơn Công thức 2.9:

$$[(1) - (0)] = [(\cdot = 1 | \cdot = \cdot) - (\cdot = 0 | \cdot = \cdot)] \quad (xem lại 2.9)$$

Nói tóm lại, nếu chúng ta vi phạm tính tích cực thì chúng ta sẽ điều chỉnh sự kiện có xác suất bằng 0. Điều này là do sẽ có một số giá trị của với xác suất khác 0 mà  $(\cdot = 1 | \cdot = \cdot) = 0$  hoặc  $(\cdot = 0 | \cdot = \cdot) = 0$ . Điều này có nghĩa là với một số giá trị của mà chúng ta đang loại trừ trong phương trình trên,  $(\cdot = 1, \cdot = \cdot) = 0$  hoặc  $(\cdot = 0, \cdot = \cdot) = 0$ , và đây là hai sự kiện mà chúng ta đặt điều kiện vào Phương trình 2.9.

Để thấy rõ sự vi phạm dương được chuyển thành phép chia cho 0 như thế nào, hãy viết lại vé phái của phương trình 2.9. Đối với các hiệp phương sai và kết quả rời rạc, nó có thể được viết lại như sau:

$$( \cdot = \cdot ) = (\cdot = \cdot | \cdot = 1, \cdot = \cdot) + (\cdot = \cdot | \cdot = 0, \cdot = \cdot) \quad (2.10)$$

Sau đó, áp dụng quy tắc Bayes, điều này có thể được viết lại:

$$( \cdot = \cdot ) = \frac{(\cdot = \cdot, \cdot = 1, \cdot = \cdot)}{(\cdot = 1 | \cdot = \cdot) (\cdot = \cdot)} = \frac{(\cdot = \cdot, \cdot = 0, \cdot = \cdot)}{(\cdot = 0 | \cdot = \cdot) (\cdot = \cdot)} \quad (2.11)$$

Trong phương trình 2.11, chúng ta có thể thấy rõ tại sao tích tích cực lại quan trọng.

Nếu  $(\cdot = 1 | \cdot = \cdot) = 0$  đối với bất kỳ mức hiệp phương sai nào có xác suất khác 0, thì sẽ có phép chia cho số 0 ở số hạng đầu tiên trong phương trình, vì vậy  $[(\cdot = 1 | \cdot = \cdot) = 1, \cdot = \cdot]$  không được xác định. Tương tự, nếu  $(\cdot = 1 | \cdot = \cdot) = 1$  cho bất kỳ mức nào của , thì  $(\cdot = 0 | \cdot = \cdot) = 0$ , do đó có phép chia cho 0 trong số hạng thứ hai và  $[(\cdot = 0 | \cdot = \cdot) = 0, \cdot = \cdot]$  không được xác định. Với một trong những vi phạm giả định tích cực này, tác động nhân quả là không xác định được.

Trực giác Đó là phép toán giải thích tại sao chúng ta cần giả định tích cực, nhưng trực giác là gì? Chà, nếu chúng ta vi phạm tính tích cực, điều đó có nghĩa là trong một số nhóm con dữ liệu, mọi người luôn nhận được sự điều trị hoặc mọi người luôn nhận được sự kiểm soát. Sẽ là vô nghĩa nếu có thể ước tính tác động nhân quả của việc điều trị so với kiểm soát trong phân nhóm đó vì chúng ta chỉ thấy điều trị hoặc chỉ kiểm soát. Chúng tôi không bao giờ thấy sự thay thế trong nhóm con đó.

Một tên gọi khác của sự tích cực là sự chồng chéo. Trực giác của tên này là chúng tôi muốn phân phối đồng biến của nhóm điều trị trùng với phân phối đồng biến của nhóm đối chứng. Cụ thể hơn, chúng tôi muốn  $(\quad | \quad = 1)$  là lý do tại sao một <sup>9</sup> có cùng độ hỗ trợ như  $(\quad | \quad = 0)$ . <sup>10</sup> Cái này bí danh phổ biến khác cho sự tích cực là hỗ trợ chung.

Sự đánh đổi giữa tính tích cực và tính không bị nhầm lẫn Mặc dù việc điều chỉnh nhiều đồng biến hơn có thể dẫn đến cơ hội thỏa mãn tính không bị nhầm lẫn cao hơn, nhưng nó có thể dẫn đến cơ hội vi phạm tính tích cực cao hơn. Khi chúng tôi tăng thứ nguyên của các hiệp phương sai, chúng tôi tạo ra các nhóm con cho bất kỳ cấp độ nào của các hiệp phương sai nhỏ hơn.<sup>11</sup> Khi mỗi nhóm con trở nên nhỏ hơn, khả năng toàn bộ nhóm con sẽ được điều trị hoặc toàn bộ nhóm nhỏ sẽ ngày càng cao hơn điều khiển. Ví dụ: một khi kích thước của bất kỳ nhóm con nào giảm xuống một, thì tính tích cực được đảm bảo sẽ không còn giữ nguyên. Xem [6] để biết lập luận chặt chẽ về các hiệp phương sai chiều cao dẫn đến vi phạm tính tích cực.

Phép ngoại suy Việc vi phạm giả định tích cực thực sự có thể dẫn đến việc đòi hỏi quá nhiều từ các mô hình và đòi lại nhận được những hành vi rất xấu. Nhiều công cụ ước tính hiệu ứng nhân quả<sup>12</sup> khớp mô hình với  $[ \quad | \quad, \quad ]$  bằng cách sử dụng các bộ dữ liệu  $(\quad, \quad, \quad)$  làm dữ liệu. Đầu vào của các mô hình này là cặp  $(\quad, \quad)$  và đầu ra là kết quả tương ứng. Các mô hình này sẽ buộc phải ngoại suy ở các vùng (sử dụng các giả định tham số của chúng) trong đó  $(\quad = 1, \quad = \quad) = 0$  và các vùng trong đó  $(\quad = 0, \quad = \quad) = 0$  khi chúng được sử dụng trong điều chỉnh công thức (Định lý 2.1) thay cho các kỳ vọng có điều kiện tương ứng.

### 2.3.5 Không can thiệp, nhất quán và SUTVA

Có một vài giả định bổ sung mà chúng tôi đã đưa vào trong suốt chương này. Chúng tôi sẽ chỉ định tất cả các giả định còn lại trong phần này. Giả định đầu tiên trong phần này là không có sự can thiệp. Không can thiệp có nghĩa là kết quả của tôi không bị ảnh hưởng bởi cách đổi xử của bất kỳ ai khác. Đúng hơn, kết quả của tôi chỉ là một chức năng của cách điều trị của chính tôi. Chúng ta đã ngầm sử dụng giả định này trong suốt chương này. Đây giờ chúng ta sẽ chính thức hóa nó.

#### Giả định 2.4 (Không can thiệp)

$$(\quad 1, \dots, \quad 1, \dots, \quad \dots, \quad \dots, \quad \dots, \quad ) = (\quad )$$

Tất nhiên, giả định này có thể bị vi phạm. Ví dụ: nếu phương pháp điều trị là "kiểm một con chó" và kết quả là hạnh phúc của tôi, thì có thể hạnh phúc của tôi bị ảnh hưởng bởi việc bạn bè tôi có nuôi chó hay không vì cuối cùng chúng tôi có thể đi chơi nhiều hơn để cho chó chơi cùng nhau. . Như bạn

<sup>9</sup> Bất cứ khi nào chúng ta sử dụng một biến ngẫu nhiên (được ký hiệu bằng chữ in hoa) làm đối số cho , chúng ta đang đề cập đến toàn bộ phân phối, thay vì chỉ đại số vô hướng như  $(\quad | \quad = 1)$  để cập đến.

<sup>10</sup> **Bài tập đọc tích cực:** hãy thuyết phục bản thân rằng công thức chồng chéo/tích cực này tương đương với công thức trong Giả định 2.3.

<sup>11</sup> Điều này có liên quan đến lời nguyên của chiều không gian .

[6]: D'Amour và cộng sự. (2017), Sự chồng chéo trong các nghiên cứu quan sát với các hiệp phương sai chiều cao

<sup>12</sup> "Công cụ ước tính" là một hàm thực hiện một tập dữ liệu làm đầu vào và đưa ra một ước tính. Chúng ta sẽ thảo luận thêm về thuật ngữ thống kê này trong Phần 2.4.

có thể mong đợi, việc vi phạm giả định không bị nhiễu đang lan tràn trong dữ liệu mạng.

Giả định cuối cùng là tính nhất quán. Tính nhất quán là giả định rằng kết quả mà chúng ta quan sát được thực sự là kết quả tiềm năng theo cách xử lý được quan sát .

Giả định 2.5 (Tính nhất quán) Nếu can thiệp là thì kết quả quan sát được là kết quả tiềm năng của can thiệp . chính thức,

$$= = = ( ) \quad (2.12)$$

Chúng ta có thể viết điều này tương đương như sau:

$$= ( ) \quad (2.13)$$

Lưu ý rằng khác với và (< ) khác với ( ). là biến ngẫu nhiên tương ứng với can thiệp được quan sát, trong khi là giá trị cụ thể của can thiệp. Tương tự, (< ) là kết quả tiềm năng đối với một số giá trị can thiệp cụ thể, trong khi ( ) là kết quả tiềm năng đối với giá trị thực tế của biến pháp can thiệp mà chúng tôi quan sát được.

Khi chúng tôi sử dụng khả năng trao đổi để chứng minh khả năng nhận dạng, chúng tôi thực sự đã giả định tính nhất quán trong Công thức 2.4 để có được đằng sau:

$$[ (1) | = 1] \quad [ (0) | = 0] = [ | = 1] \quad [ | = 0]$$

Tương tự, khi chúng tôi sử dụng khả năng trao đổi có điều kiện để chứng minh khả năng nhận dạng, chúng tôi giả định tính nhất quán trong Công thức 2.7.

Có vẻ như tính nhất quán rõ ràng là đúng, nhưng không phải lúc nào cũng như vậy . Ví dụ: nếu đặc điểm xử lý chỉ đơn giản là “bắt một con chó” hoặc “không nuôi một con chó”, thì điều này có thể quá thô thiển để mang lại sự nhất quán. Có thể là nếu tôi nuôi một con chó con, tôi sẽ quan sát = 1 (hạnh phúc) vì tôi cần một người bạn tràn đầy năng lượng, nhưng nếu tôi nuôi một con chó già, năng lượng thấp, tôi sẽ quan sát = 0 (bất hạnh) . Tuy nhiên, cả hai cách xử lý này đều thuộc danh mục “bắt một con chó”, vì vậy cả hai đều tương ứng với = 1. Điều này có nghĩa là (1) không được xác định rõ ràng, vì nó sẽ là 1 hoặc 0, tùy thuộc vào điều gì đó không được nắm bắt bởi đặc điểm kỹ thuật xử lý. Theo nghĩa này, tính nhất quán bao hàm giả định đôi khi được gọi là “không có nhiều phiên bản điều trị” . Xem Phần 3.4 và 3.5 của Hernán và Robins [7] và các tài liệu tham khảo trong đó để thảo luận thêm về chủ đề này.

[7]: Hernán và Robins (2020), Suy luận nhân quả : Chuyện gì xảy ra nếu

SUTVA Bạn cũng sẽ thường thấy giả định giá trị điều trị đơn vị ổn định (SUTVA) trong tài liệu. SUTVA hài lòng nếu kết quả của đơn vị (cá nhân) chỉ đơn giản là một hàm trong cách xử lý của đơn vị . Do đó, SUTVA là sự kết hợp giữa tính nhất quán và không có sự can thiệp (và cả những kết quả tiềm năng mang tính xác định).13

<sup>13</sup> **Bài tập đọc tích cực:** hãy thuyết phục bản thân rằng SUTVA là sự kết hợp của tính nhất quán và không có suy luận

### 2.3.6 Buộc tất cả lại với nhau

Chúng tôi giới thiệu tính vô căn cứ (khả năng trao đổi có điều kiện) trước tiên vì đây là giả định nhân quả chính. Tuy nhiên, tất cả các giả định đều cần thiết:

1. Tính vô căn cứ (Giả định 2.2)
2. Tính tích cực (Giả định 2.3)
3. Không bị nhiễu (Giả định 2.4)
4. Tính nhất quán (Giả định 2.5)

Bây giờ chúng ta sẽ xem lại cách chứng minh công thức điều chỉnh (Định lý 2.1) được thực hiện từ Công thức 2.5 đến Công thức 2.9 và liệt kê các giả định được sử dụng cho từng bước. Ngay cả trước khi chúng ta đi đến các phương trình này, chúng ta sử dụng giả định không có nhiễu để chứng minh rằng đại lượng mà chúng ta nên xem xét để suy luận nhân quả là [ (1) (0)], thay vì một thứ phức tạp hơn như bên trái- vé giả định 2.4.

Trong chứng minh dưới đây, hai đẳng thức đầu tiên xuất phát từ các sự kiện toán học, trong khi hai đẳng thức cuối cùng xuất phát từ các giả định chính này.

Chứng minh Định lý 2.1.

$$\begin{aligned}
 [ (1) (0)] &= [ (1)] [ (0)] \quad (\text{độ tuyển tính của kỳ vọng}) = \\
 &[ [ (1) | ] [ (0) | ] ] \\
 &\quad (\text{lý thuyết kỳ vọng lặp lại}) \\
 &= [ [ (1) | ] = 1, [ (0) | ] = 0, \\
 &\quad ] ] \quad (\text{vô căn cứ và dương}) \\
 &= [ [ | ] = 1, [ | ] = 0, \\
 &\quad ] ] \quad (\text{nhất quán})
 \end{aligned}$$

□

Đó là cách tất cả các giả định này liên kết với nhau để giúp chúng ta có thể nhận dạng được ATE. Chúng ta sẽ sớm xem cách sử dụng kết quả này để có được con số ước tính thực tế cho ATE.

## 2.4 Thuật ngữ thống kê ưa thích không được chuẩn hóa

Trước khi bắt đầu tính toán những con số cụ thể cho ATE, chúng ta phải nhanh chóng giới thiệu một số thuật ngữ từ số liệu thống kê để giúp làm rõ cuộc thảo luận. Ước tính là số lượng mà chúng ta muốn ước tính. Ví dụ:  $[ [ | ] = 1, [ | ] = 0, ] ]$  là ước tính mà chúng tôi quan tâm để ước tính ATE. Ước tính (danh từ) là giá trị gần đúng của một số ước tính mà chúng tôi nhận được bằng cách sử dụng dữ liệu. Chúng ta sẽ thấy những con số cụ thể ở phần tiếp theo; đây là những ước tính. Cho trước một số ước tính  $\hat{\theta}$ , chúng ta viết ước tính cho ước tính đó bằng cách đơn giản đặt mũ lên nó:

.. Vẽ công cụ ước tính là một hàm ánh xạ tập dữ liệu tới ước tính của ước tính.

Quá trình chúng ta sẽ sử dụng để chuyển từ dữ liệu + ước tính sang một con số cụ thể được gọi là ước tính. Ước tính (động từ) là đưa dữ liệu vào công cụ ước tính để có được ước tính.

Trong cuốn sách này, chúng tôi sẽ sử dụng ngôn ngữ cụ thể hơn nữa để cho phép chúng tôi phân biệt giữa đại lượng nhân quả và đại lượng thống kê.

Chúng tôi sẽ sử dụng cụm từ ước tính nhân quả để đề cập đến bất kỳ ước tính nào chứa kết quả tiềm năng trong đó. Chúng ta sẽ sử dụng cụm từ ước tính thống kê để biểu thị phần bù: bất kỳ ước tính nào không chứa kết quả tiềm năng.<sup>14</sup> Ví dụ: hãy nhớ lại công thức điều chỉnh

<sup>14</sup> Như chúng ta sẽ thấy trong Chương 4, chúng ta sẽ coi ước tính nhân quả là bất kỳ ước tính nào có chứa toán tử do và chúng ta sẽ coi ước tính thống kê là bất kỳ ước tính nào không chứa toán tử do.

(Định lý 2.1):

$$[ \quad (1) \quad (0) ] = [ \quad [ \quad | \quad = 1, \quad ] \quad [ \quad | \quad = 0, \quad ] ] \quad (2.14)$$

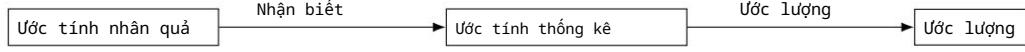
[ (1) (0) ] là ước tính nhân quả mà chúng ta quan tâm. Để thực sự ước tính ước tính nhân quả này, chúng ta phải chuyển nó thành ước lượng thống kê: [ [ | = 1, ] ] [ | = 0, ] ].

15

Khi chúng tôi nói “nhận dạng” trong cuốn sách này, chúng tôi đang đề cập đến quá trình chuyển từ ước tính nhân quả sang ước tính thống kê tương đương.

Khi chúng tôi nói “Ước tính”, chúng tôi đang đề cập đến quá trình chuyển từ ước tính thống kê sang ước tính. Chúng tôi minh họa điều này trong sơ đồ trong

Hình 2.5.



Hình 2.5: Sơ đồ nhận dạng-Ước tính - sơ đồ minh họa quá trình chuyển từ ước tính nhân quả mục tiêu sang ước tính tương ứng, thông qua nhận dạng và ước lượng.

Chúng ta phải làm gì khi thực sự ước tính số lượng như [ [ | = 1, ] ] [ | = 0, ]]? Chúng ta sẽ thường sử dụng một mô hình (ví dụ: hồi quy tuyến tính hoặc một số công cụ dự đoán thú vị hơn từ học máy) thay cho các kỳ vọng có điều kiện [ | = , , = ]. Chúng tôi sẽ đề cập đến các công cụ ước tính sử dụng các mô hình như thế này làm công cụ ước tính được hỗ trợ bởi mô hình.

Bây giờ chúng ta đã hiểu rõ một số thuật ngữ này, chúng ta có thể chuyển sang ví dụ về ước tính ATE.

## 2.5 Một ví dụ hoàn chỉnh với ước tính

Định lý 2.1 và bù sao tương ứng gần đây trong phương trình 2.14 cho chúng ta nhận dạng. Tuy nhiên, chúng tôi chưa hề thảo luận về việc ước tính. Trong phần này, chúng tôi sẽ đưa ra một ví dụ ngắn kèm theo ước tính. Chúng ta sẽ đề cập đến chủ đề ước tính tác động nhân quả một cách đầy đủ hơn trong Chương 7.

Chúng tôi sử dụng Luque-Fernandez et al. [8] ví dụ từ dịch tễ học. Kết quả quan tâm là huyết áp (tâm thu). Đây là một kết quả quan trọng vì khoảng 46% người Mỹ bị huyết áp cao và huyết áp cao có liên quan đến việc tăng nguy cơ tử vong [9]. “Phương pháp điều trị” được quan tâm là lượng natri nạp vào. Lượng natri tiêu thụ là một biến số liên tục; để dễ dàng áp dụng Công thức 2.14, được chỉ định cho xử lý nhị phân, chúng tôi sẽ nhị phân hóa bằng cách cho  $= 1$  biểu thị lượng natri tiêu thụ hàng ngày trên 3,5 gam và để  $= 0$  biểu thị lượng natri tiêu thụ hàng ngày dưới 3,5 gam.16 Chúng tôi sẽ ước tính lượng natri tiêu thụ hàng ngày dưới 3,5 gam. tác dụng nhân quả của lượng natri ăn vào đối với huyết áp. Trong dữ liệu của chúng tôi, chúng tôi cũng có tuổi của các cá nhân và lượng protein trong nước tiểu của họ dưới dạng Luque-Fernandez và cộng sự. [8] chạy một mô phỏng, chú ý đảm bảo rằng phạm vi giá trị “hợp lý về mặt sinh học và càng gần với thực tế càng tốt”.

Vì chúng tôi đang sử dụng dữ liệu từ mô phỏng nên chúng tôi biết rằng ATE thực sự của natri đối với huyết áp là 1,05. Cụ thể hơn, dòng mã tạo ra huyết áp trong như sau:

<sup>1</sup>  $\text{huyết áp} = 1,05 * \text{natri} + \dots$

15 [Bài tập đọc tích cực](#): Tại sao chúng ta không thể ước tính trực tiếp một ước tính nhân quả mà không chuyển nó sang ước tính thống kê trước?

[8]: Luque-Fernandez và cộng sự. (2018), ‘Lưu ý giáo dục: Hiệu ứng máy và chậm nghịch lý trong phân tích dữ liệu dịch tễ học về bệnh không lây nhiễm: một minh họa có thể tái tạo và ứng dụng web’

[9]: Virani và cộng sự. (2020), ‘Thống kê về bệnh tim và đột quỵ-Cập nhật năm 2020: Báo cáo từ Hiệp hội Tim mạch Hoa Kỳ’

16 Như chúng ta sẽ thấy, việc nhị phân hóa này hoàn toàn mang tính sự phạm và không phản ánh bất kỳ hạn chế nào trong việc điều chỉnh các yếu tố gây nhiễu.

Bây giờ, làm thế nào để chúng ta thực sự ước tính ATE? Đầu tiên, chúng tôi giả định tính nhất quán, tích tích cực và tính vô căn cứ đã cho . Như chúng tôi vừa nhắc lại trong phương trình 2.14, điều này có nghĩa là chúng tôi đã xác định ATE là

$$\frac{1}{[ \quad | \quad ]} = 1, \quad [ \quad | \quad ] = 0, \quad ]]$$

Sau đó, chúng tôi lấy kỳ vọng bên ngoài đó thay cho và thay thế nó bằng giá trị trung bình thực nghiệm đối với dữ liệu, mang lại cho chúng tôi kết quả sau:

$$\frac{1}{[ \quad | \quad ]} = 1, \quad [ \quad | \quad ] = 0, \quad ]] \quad (2.15)$$

Để hoàn thiện công cụ ước tính, chúng tôi điều chỉnh một số mô hình học máy phù hợp với kỳ vọng có điều kiện [ | ]. Việc giảm thiểu sai số bình phương trung bình (MSE) của việc dự đoán từ các cặp ( , ) tương đương với việc lập mô hình kỳ vọng có điều kiện này [xem, ví dụ, 10, Phần 2.4]. Do đó, chúng ta có thể áp dụng bất kỳ mô hình học máy nào cho [ | ], cung cấp cho chúng ta một công cụ ước tính được hỗ trợ bởi mô hình. Ở đây chúng ta sẽ sử dụng hồi quy tuyến tính, phương pháp này hoạt động rất tốt vì huyết áp được tạo ra dưới dạng kết hợp tuyến tính của các biến khác trong mô phỏng này. Chúng tôi cung cấp mã Python cho phần này bên dưới, trong đó dữ liệu của chúng tôi nằm trong Khung dữ liệu Pandas có tên là df. Chúng tôi phù hợp với mô hình cho [ | , ] ở dòng 8 và chúng tôi lấy giá trị trung bình thực nghiệm trên ở dòng 10-14.

```
1 nhập numpy dưới dạng
np 2 nhập gấu trúc dưới
dạng pd 3 từ sklearn.Tuyến_model nhập LinearRegression 4

5 Xt = df[['sodium', 'age', 'proteinuria']] 6 y =
df['blood_Pressure'] 7 model =
LinearRegression() 8 model.fit(Xt,
y) 9

10 Xt1 = pd.DataFrame.copy(Xt)
11 Xt1['natri'] = 1 12
Xt0 = pd.DataFrame.copy(Xt)
13 Xt0['sodium'] = 0 14
eat_est = np.mean(model.predict(Xt1) - model.predict(Xt0)) 15 print('ATE
ước tính:', eat_est)
```

Điều này mang lại ước tính ATE là 0,85. Nếu chúng ta đơn giản hồi quy trên , tương ứng với việc thay thế dòng 5 trong Liệt kê 2.1 bằng chúng ta sẽ nhận được ước  $100\% = 40\%$  lỗi! Ngược <sup>17</sup> tính ATE là 5,33. Đó là  $Xt = df[['natri']]$ ,  $|5,33 \ 1,05| \times$  lại, khi chúng ta kiểm soát (như trong  $1,05 \ + 0,85 \ 1,05|$  Liệt kê 2.1), sai số phần trăm của chúng ta chỉ là  $\times 100\% = 19\%.$  1,05

Tất cả những điều trên được thực hiện bằng cách sử dụng công thức điều chỉnh với ước tính được hỗ trợ bởi mô hình, trong đó trước tiên chúng tôi điều chỉnh mô hình cho kỳ vọng có điều kiện [ | , ], và sau đó chúng ta lấy trung bình thực nghiệm trên , sử dụng mô hình đó. Tuy nhiên, vì chúng ta đang sử dụng mô hình tuyến tính nên điều này tương đương với việc chỉ lấy hệ số ở phía trước trong hồi quy tuyến tính làm ước tính ATE . Đây là những gì chúng tôi thực hiện trong đoạn mã sau (đưa ra ước tính ATE chính xác như nhau):

```
1 Xt = df[['sodium', 'age', 'proteinuria']] 2 y =
df['blood_Pressure'] 3 mô hình
= LinearRegression()
```

[10]: Hastie và cộng sự. (2001), Các yếu tố của việc học thống kê

Liệt kê 2.1: Mã Python để ước tính ATE

Mã đầy đủ , hoàn chỉnh với mô phỏng, có sẵn tại [https://github.com;bradyneal/causal-book-code/blob/master/sodium\\_example.py](https://github.com;bradyneal/causal-book-code/blob/master/sodium_example.py).

<sup>17</sup> Bài tập đọc tích cực: Phiên bản đơn giản này tương đương với việc chỉ lấy sự khác biệt liên kết: [ | = 1] [ | = 0]. Tại sao?

Liệt kê 2.2: Mã Python để ước tính ATE bằng hệ số hồi quy tuyến tính

```

4 model.fit(Xt, y)
5 eat_est = model.coef_[0]
6 print(' Ước tính ATE:', eat_est)

```

Điều trị liên tục Điều gì sẽ xảy ra nếu chúng ta cho phép điều trị, lượng natrì hàng ngày lượng tiêu thụ, để duy trì liên tục, thay vì nhị phân hóa nó? Điều tuyệt vời về việc chỉ lấy hệ số hồi quy làm ước tính ATE là không cần phải tính đến sự khác biệt giữa hai giá trị xử lý (ví dụ  $\alpha = 1$  và  $\beta = 0$ ), do đó nó khái quát hóa một cách tầm thường khi liên tục. Khi liên tục, chúng ta quan tâm đến việc  $[ \alpha + \beta x ]$  thay đổi như thế nào với  $x$ . Từ chúng tôi giả sử  $[ \alpha + \beta x ]$  là tuyến tính, sự thay đổi này hoàn toàn được nắm bắt bởi  $\beta$ .<sup>18</sup>  $(\beta)$  là tuyến tính, hóa ra đại lượng này chính xác là những gì lấy hệ số từ ước tính hồi quy tuyến tính. Có vẻ như thật kỳ diệu, chúng tôi đã nén tất cả  $[ \alpha + \beta x ] = [ \alpha | \beta x ]$ , là hàm của  $x$ , thành một giá trị duy nhất.

Tuy nhiên, việc nén tất cả  $[ \alpha | \beta x ]$  cho liên tục xuất hiện dưới dạng chi phí: dạng tham số tuyến tính mà chúng tôi đã giả định. Nếu mô hình này được xác định sai,<sup>19</sup> ước tính ATE của chúng tôi sẽ bị sai lệch. Và bởi vì tuyến tính các mô hình quá đơn giản nên chúng có thể sẽ bị xác định sai. Ví dụ, Giả định sau đây ngầm hàm ý rằng một mô hình tuyến tính được xác định rõ ràng: hiệu quả can thiệp là nhau đối với tất cả các cá nhân. Xem Morgan và Winship [12, Phần 6.2 và 6.3] để có sự phê bình đầy đủ hơn về sử dụng hệ số trước điều trị làm ước tính ATE.

<sup>18</sup> Tóm tắt chính xác hàm phi tuyến  $[ \alpha + \beta x ]$  là một bài toán mở. Xem, ví dụ, Janzing và cộng sự. [11].

[11]: Janzing và cộng sự. (2013), 'Định lượng tác động nhân quả'

<sup>19</sup> Khi nói "xác định sai", chúng tôi muốn nói rằng dạng hàm của mô hình không phù hợp với dạng chức năng của quá trình tạo dữ liệu.

[12]: Morgan và Winship (2014), Suy luận phản thực tế và nhân quả: Phương pháp và Nguyên tắc nghiên cứu xã hội

# Dòng chảy của Hiệp hội và Quan hệ nhân quả trong đồ thị

# 3

Chúng ta đã sử dụng đồ thị nhân quả trong các chương trước để hỗ trợ trực giác. Trong chương này, chúng tôi sẽ giới thiệu các hình thức làm nền tảng cho vấn đề này. Hy vọng rằng chúng ta đã có đủ động lực cho chương này và làm cho tiện ích của các mô hình đồ họa trở nên rõ ràng với tất cả các tính năng đồ họa giải thích các khái niệm ở các chương trước.

## 3.1 Thuật ngữ đồ thị

Trong phần này chúng ta sẽ sử dụng thuật ngữ súng máy (xem Hình 3.1). ĐẾN có thể sử dụng ngôn ngữ đồ thị tiện lợi trong các phần sau, Thật không may, việc sử dụng nhanh rất nhiều thuật ngữ đồ thị là một điều xấu cần thiết.

Thuật ngữ "đồ thị" thường được sử dụng để mô tả nhiều hình ảnh trực quan khác nhau. Ví dụ: "biểu đồ" có thể đề cập đến hình ảnh trực quan của một biến duy nhất hàm ( ), trong đó được vẽ trên trực và ( ) được vẽ trên trực . Hoặc "biểu đồ thanh" có thể được sử dụng làm từ đồng nghĩa với thanh đồ thị. Tuy nhiên, trong lý thuyết đồ thị, thuật ngữ "đồ thị" đề cập đến một đối tượng toán học.

Đồ thị là một tập hợp các nút (còn gọi là "định") và các cạnh kết nối các nút. Ví dụ: trong Hình 3.2, , , và là các nút của đồ thị và các đường nối chúng là các cạnh. Hình 3.2 là gọi là đồ thị vô hướng vì các cạnh không có hướng. TRONG Ngược lại, Hình 3.3 là đồ thị có hướng. Các cạnh của đồ thị có hướng đi ra ngoài của nút cha và vào nút con, với các mũi tên biểu thị nút nào hướng các cạnh đang đi. Chúng ta sẽ biểu thị nút cha của nút với pa( ). Chúng ta sẽ sử dụng cách viết tắt thậm chí còn đơn giản hơn khi các nút được sắp xếp để chúng ta có thể biểu thị nút bởi ; trong trường hợp đó, chúng tôi sẽ cũng biểu thị cha mẹ của bằng pa . Hai nút được gọi là liền kề nếu chúng được kết nối bởi một cạnh. Ví dụ, trong cả Hình 3.2 và Hình 3.3, và liền kề nhau nhưng và thì không.

Đường dẫn trong biểu đồ là bất kỳ chuỗi các nút liền kề nào, bắt đầu từ hướng của các cạnh nối chúng. Ví dụ: - - là đường dẫn trong Hình 3.2, và là một đường dẫn trong Hình 3.3. Một đường dẫn có hướng là một đường đi bao gồm các cạnh được định hướng đều có hướng giống nhau hướng (không có hai cạnh dọc theo đường dẫn đều trỏ vào hoặc cả hai điểm ra khỏi cùng một nút). Ví dụ: là đường đi có hướng trong

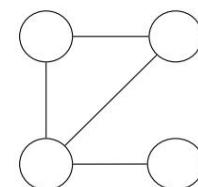
Hình 3.3 nhưng và thì không.

Nếu có một đường dẫn có hướng bắt đầu tại nút và kết thúc tại nút , thì là tổ tiên của , và là hậu duệ của . Chúng ta sẽ biểu thị con cháu của bởi de( ). Ví dụ, trong Hình 3.3, là tổ tiên của và , và đều là hậu duệ của (de( )). Nếu là tổ tiên của chính nó, sau đó một số chuyển đổi hành thời gian thú vị đã diễn ra. Nói một cách nghiêm túc, một đường đi có hướng từ nút nào đó quay lại chính nó được gọi là một chu trình (xem Hình 3.4). Nếu không có chu trình trong đồ thị có hướng thì đồ thị đó đã biết

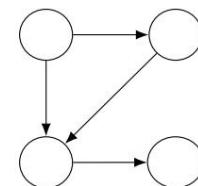
3.1 Thuật ngữ đồ thị . . . . .	19
3.2 Mạng Bayes . . . . .	20
3.3 Đồ thị nhân quả . . . . .	22
3.4 Đồ thị hai nút và Khối xây dựng đồ họa . . . . .	23
3.5 Xích và Nĩa . . . . .	24
3.6 Máy và chạm và sự di chuyển của chúng dants . . . . .	26
3,7 d-tách . . . . .	28
3.8 Dòng chảy của Hiệp hội và Cau- . . . . .	30



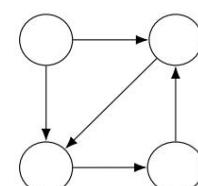
Hình 3.1: Thuật ngữ súng máy



Hình 3.2: Đồ thị vô hướng



Hình 3.3: Đồ thị có hướng



Hình 3.4: Đồ thị có hướng có chu trình

dưới dạng đồ thị chu kỳ có hướng (DAG). Các đồ thị mà chúng tôi tập trung vào trong cuốn sách này sẽ chủ yếu là DAG.

Nếu hai cha mẹ và có chung một số con, nhưng không có cạnh nào kết nối và thì được coi là vô đạo đức. Nghiêm túc; đó là một thuật ngữ thực tế trong các mô hình đồ họa. Ví dụ: nếu chúng ta loại bỏ từ Hình 3.3 để có được Hình 3.5 thì là trái đạo đức.

## 3.2 Mạng Bayes

Hóa ra phần lớn công việc dành cho các mô hình đồ họa nhân quả đã được thực hiện trong lĩnh vực mô hình đồ họa xác suất. Đồ họa xác suất mô hình là mô hình thống kê trong khi mô hình đồ họa nhân quả là mô hình nhân quả các mô hình. Mạng Bayesian là mô hình đồ họa xác suất chính rằng các mô hình đồ họa nhân quả (mạng Bayes nhân quả) kế thừa hầu hết tài sản của họ từ.

Hãy tưởng tượng rằng chúng ta chỉ quan tâm đến hiệp hội người mẫu mà không có bất kỳ mô hình nhân quả. Chúng tôi muốn lập mô hình phân phối dữ liệu  $(1, 2, \dots)$ .

Nói chung, chúng ta có thể sử dụng quy tắc xác suất chuỗi để phân tích bất kỳ phân phối nào :

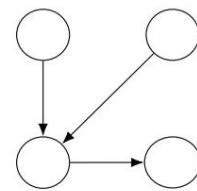
$$(1, 2, \dots) = (1) | (1, \dots, 1) \quad (3.1)$$

Tuy nhiên, nếu chúng ta mô hình hóa các yếu tố này bằng bảng, sẽ mất một số mủ của tham số. Để thấy điều này, hãy lấy mỗi thành nhị phân và xem xét cách chúng ta mô hình hóa hệ số  $(| 1, \dots, 1)$ . Từ là nhị phân nên chúng ta chỉ cần lập mô hình  $(| = 1 | 1, \dots, 1)$  vì  $(| = 0 | 1, \dots, 1)$  chỉ đơn giản là 1.  $(| = 1 | 1, \dots, 1)$  Vâng, chúng tôi 1 sẽ cần 2 các tham số để mô hình hóa điều này. Lấy một ví dụ cụ thể, hãy  $= 4$ . Như chúng ta có thể thấy trong Bảng 3.1, điều này đòi hỏi  $2^4 = 16$  tham số:  $1, \dots, 16$ . Việc tham số hóa bạo lực này nhanh chóng trở nên khó hiểu khi tăng lên.

Một cách trực quan để lập mô hình hiệu quả hơn nhiều biến cùng nhau trong một bản phân phối chung chỉ để mô hình hóa các phụ thuộc cục bộ. Ví dụ, thay vì lập mô hình thừa số 4 dưới dạng  $(4 | 3, 2, 1)$ , chúng ta có thể lập mô hình nó là  $(4 | 3)$  nếu chúng ta có lý do để tin rằng 4 chỉ phụ thuộc cục bộ trên 3. Trên thực tế, trong biểu đồ tương ứng ở Hình 3.6, nút duy nhất cấp dữ liệu vào 4 là 3. Điều này có nghĩa là biểu thị rằng 4 chỉ cục bộ phụ thuộc vào 3. Bất cứ khi nào chúng ta sử dụng đồ thị liên quan đến xác suất phân phối, sẽ luôn có ánh xạ một-một giữa các nút trong và các biến ngẫu nhiên trong , vì vậy khi chúng ta nói về các nút độc lập, ý chúng tôi là các biến ngẫu nhiên tương ứng là độc lập.

Cho một phân bố xác suất và một chu kỳ có hướng tương ứng đồ thị (DAG), chúng ta có thể chính thức hóa đặc tả tính độc lập bằng giả định Markov cục bộ:

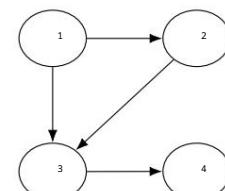
**Giả định 3.1 (Giả định Markov địa phương)** Cho cha mẹ của nó trong DAG, nút độc lập với tất cả các nút không phải là con cháu của nó.



Hình 3.5: Đồ thị có hướng có tính vô đạo đức

Bảng 3.1: Bảng cần thiết để lập mô hình thừa số đơn  $(| 1, \dots, 1)$  trong đó  $= 4$  và các biến là nhị phân. Các số lượng tham số cần thiết là số mủ trong .

1	2	3	$(  4   3, 2, 1)$
0	0	0	1
0	0	1	2
0	1	0	3
0	1	1	4
1	0	0	5
1	0	1	6
1	1	0	7
1	1	1	8



Hình 3.6: DAG bốn nút trong đó 4 cục bộ chỉ phụ thuộc vào 3.

Giả định này (cùng với các DAG cụ thể) mang lại cho chúng tôi rất nhiều điều. Chúng ta sẽ chứng minh điều này trong một vài phương trình tiếp theo. Trong ví dụ bốn biến của chúng ta, quy tắc chuỗi xác suất cho chúng ta biết rằng chúng ta có thể phân tích bất kỳ nào sao cho

$$(-1, -2, -3, -4) = (-1)(-2 | -1)(-3 | -2, -1)(-4 | -3, -2, -1). \quad (3.2)$$

Nếu là Markov đối với đồ thị 1 trên Hình 3.6, thì chúng ta có thể đơn giản hóa thừa số cuối cùng:

$$(-1, -2, -3, -4) = (-1)(-2)(-3 | -1)(-4 | -3). \quad (3.3)$$

Nếu chúng ta loại bỏ thêm các cạnh, loại bỏ 1 2 và 2 3 như trong Hình 3.7, chúng ta có thể đơn giản hóa hơn nữa việc phân tích nhân tử của :

$$(-1, -2, -3, -4) = (-1)(-2)(-3 | -1)(-4 | -3). \quad (3.4)$$

Với sự hiểu biết mà chúng tôi hy vọng đã xây dựng được từ một vài ví dụ, bây giờ chúng tôi sẽ nêu một trong những hậu quả chính của tình trạng cục bộ.

Giả định Markov:

**Định nghĩa 3.1 (Hệ số hóa mạng Bayes)** Cho phân bố xác suất và DAG , phân tích theo nếu

$$(-1, \dots, -n) = (-1 | p_1, \dots, p_n)$$

Hy vọng rằng bạn thấy sự giống nhau giữa việc chuyển từ Công thức 3.2 sang Công thức 3.3 hoặc chuyển sang Công thức 3.4 và sự khái quát hóa của điều này được trình bày trong Định nghĩa 3.1.

Hệ số hóa mạng Bayes còn được gọi là quy tắc chuỗi cho mạng Bayesian hoặc khả năng tương thích Markov. Ví dụ, nếu phân tích nhân tử theo , thì và tương thích với Markov.

Chúng ta đã đưa ra trực giác về cách giả định Markov cục bộ bao hàm việc phân tích hệ số mạng Bayes và hóa ra cả hai thực tế là tương đương nhau. Nói cách khác, chúng ta có thể bắt đầu với việc phân tích hệ số mạng Bayes làm giả định chính (và gắn nhãn nó là một giả định) và chỉ ra rằng nó bao hàm giả định Markov cục bộ.

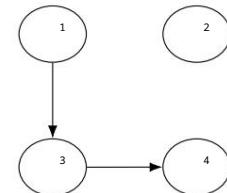
Xem Koller và Friedman [13, Chương 3] để biết những chứng minh này và biết thêm thông tin về chủ đề này.

Mặc dù giả định Markov cục bộ cũng quan trọng nhưng nó chỉ cung cấp cho chúng ta thông tin về tính độc lập trong mà DAG ngũ ý. Nó thậm chí còn không cho chúng ta biết rằng nếu và liền kề trong DAG thì và là phụ thuộc. Và thông tin bổ sung này thường được giả định trong DAG nhân quả. Để có được sự phụ thuộc được đảm bảo này giữa các nút lân cận, thông thường chúng ta sẽ giả định một giả định mạnh hơn một chút so với giả định Markov cục bộ: tính tối thiểu.

**Giả định 3.2 (Giả định tối thiểu)** 1. Với cha mẹ của nó trong DAG, nút độc lập với tất cả các nút không con cháu của nó (Giả định 3.1).

2. Các nút liền kề trong DAG phụ thuộc vào nhau.

<sup>1</sup> Một phân phối xác suất được gọi là Markov (cục bộ) đối với DAG nếu chúng thỏa mãn giả định Markov cục bộ.



Hình 3.7: Bốn nút DAG với nhiều tính độc lập hơn.

<sup>2</sup> **Bài tập đọc tích cực:** đảm bảo rằng bạn biết cách chúng ta đi từ Công thức 3.2 đến Công thức 3.3 và Công thức 3.4 bằng cách sử dụng giả định Markov cục bộ.

[13]: Koller và Friedman (2009), Mô hình đồ họa xác suất: Nguyên tắc và kỹ thuật

<sup>3</sup> Điều này thường được phát biểu tương đương theo cách sau: nếu chúng ta loại bỏ bất kỳ cạnh nào khỏi DAG, sẽ không phải là Markov đối với đồ thị có các cạnh bị loại bỏ [xem, ví dụ, 14, Phần 6.5.3]

[14]: Peters và cộng sự. (2017), Các yếu tố của suy luận nhân quả: Nền tảng và thuật toán học tập

Để biết lý do tại sao giả định này được đặt tên là "tối thiểu", hãy xem xét những gì chúng ta biết khi biết rằng là Markov đối với DAG . Chúng ta biết rằng thỏa mãn một tập độc lập đặc trưng cho cấu trúc của . Nếu và cũng thỏa mãn tính tối thiểu thì tập các tính độc lập này là tối thiểu theo nghĩa là không thỏa mãn bất kỳ tính độc lập bổ sung nào.

Điều này tương đương với việc nói rằng các nút lân cận là phụ thuộc.

Ví dụ: nếu DAG chỉ đơn giản là hai nút được kết nối và như trong Hình 3.8, giả định Markov cục bộ sẽ cho chúng ta biết rằng chúng ta có thể phân tích ( , ) thành ( ) ( | ), nhưng nó cũng sẽ cho phép chúng ta phân tích ( , ) thành ( ) ( ), nghĩa là nó cho phép phân phối trong đó và độc lập. Ngược lại, giả định tối thiểu không cho phép tính độc lập bổ sung này. Tính tối thiểu sẽ bảo chúng ta phân tích ( , ) thành ( ) ( | ), và nó sẽ cho chúng ta biết rằng không có sự độc lập bổ sung nào ( ) tồn tại trong là tối thiểu đối với Hình 3.8 .

Bởi vì việc loại bỏ các cạnh trong mạng Bayes tương đương với việc thêm các cạnh độc lập<sup>4</sup>, giả định về mức tối thiểu tương đương với việc nói rằng chúng ta không thể loại bỏ thêm bất kỳ cạnh nào khỏi biểu đồ. Theo một nghĩa nào đó, mọi cạnh đều "hoạt động". Cụ thể hơn, hãy xem xét rằng và tương thích với Markov và là những gì chúng ta nhận được khi loại bỏ một số cạnh khỏi . Nếu cũng là Markov đối với thì không tối thiểu đối với .

Được trang bị giả định về tính tối thiểu và ý nghĩa của nó về cách phân tích thành thửa só khi chúng là Markov đối với một số DAG (Định nghĩa 3.1), giờ đây chúng ta đã sẵn sàng thảo luận về luồng liên kết trong DAG. Tuy nhiên, vì mọi thứ trong phần này hoàn toàn mang tính thống kê nên chúng tôi chưa sẵn sàng thảo luận về luồng quan hệ nhân quả trong DAG. Để làm được điều đó, chúng ta phải đưa ra các giả định nhân quả. Về mặt sư phạm, điều này cũng sẽ cho phép chúng ta sử dụng ngôn ngữ nhân quả trực quan khi giải thích dòng liên tưởng.

### 3.3 Đồ thị nhân quả

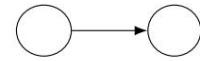
Phần trước là tất cả về mô hình thống kê và liên kết mô hình. Trong phần này, chúng ta sẽ tăng cường các mô hình này bằng các giả định nhân quả, biến chúng thành mô hình nhân quả và cho phép chúng ta nghiên cứu nhân quả. Để đưa ra các giả định nhân quả, trước tiên chúng ta phải hiểu ý nghĩa của việc là nguyên nhân của .

**Định nghĩa 3.2 (Nguyên nhân là gì?)** Biến được gọi là nguyên nhân của một biến nếu có thể thay đổi để đáp ứng với những thay đổi trong <sup>5</sup>.

Một cụm từ khác thường được sử dụng để mô tả nguyên thủy này là "lắng nghe" . Với điều này, giờ đây chúng ta có thể xác định rõ giả định nhân quả chính mà chúng ta sẽ sử dụng trong suốt cuốn sách này.

**Giả định 3.3 (Giả định các cạnh nhân quả (nghiêm ngặt))** Trong đồ thị có hướng, mỗi phụ huynh là nguyên nhân trực tiếp của tất cả con cái của nó.

Ở đây, tập hợp các nguyên nhân trực tiếp của là mọi thứ mà phản ứng trực tiếp ; nếu chúng ta khắc phục tất cả nguyên nhân trực tiếp của thì việc thay đổi bất kỳ nguyên nhân nào khác của sẽ không gây ra bất kỳ thay đổi nào trong . Giả định này "nghiêm ngặt" theo nghĩa



Hình 3.8: Hai nút được kết nối

<sup>4</sup> **Bài tập đọc tích cực:** tại sao việc loại bỏ các cạnh trong mạng Bayes tương đương với việc thêm các phần độc lập?

<sup>5</sup> Xem Phần 4.5.1 để biết định nghĩa sử dụng ký hiệu toán học.

rằng mọi cạnh đều “hoạt động”, giống như trong DAG thỏa mãn mức tối thiểu. Nói cách khác, vì định nghĩa về nguyên nhân (Định nghĩa 3.2) ngụ ý rằng nguyên nhân và kết quả của nó phụ thuộc vào nhau và bởi vì chúng ta cho rằng tất cả cha mẹ đều là nguyên nhân của con cái họ, nên chúng ta đang giả định rằng cha mẹ và con cái của họ đều phụ thuộc. Vì vậy, phần thứ hai của tính tối thiểu (Giả định 3.2) được đưa vào giả định biên nhân quả chặt chẽ.

Ngược lại, giả định về các khía cạnh nhân quả không chặt chẽ sẽ cho phép một số cha mẹ không phải là nguyên nhân của con cái họ. Nó sẽ chỉ cho rằng con cái không phải là nguyên nhân của cha mẹ chúng. Điều này cho phép chúng ta vẽ đồ thị có nhiều cạnh hơn để đưa ra ít giả định hơn, giống như cách chúng ta làm trong mạng Bayesian, trong đó nhiều cạnh hơn có nghĩa là ít giả định độc lập hơn. Đồ thị nhân quả đôi khi được vẽ với loại ý nghĩa không tối thiểu này, nhưng trong phần lớn trường hợp, khi ai đó vẽ đồ thị nhân quả, điều đó có nghĩa là cha mẹ là nguyên nhân của con cái họ. Do đó, trừ khi chúng tôi có quy định khác, xuyên suốt cuốn sách này, chúng tôi sẽ sử dụng “đồ thị nhân quả” để chỉ một DAG thỏa mãn giả định biên nhân quả nghiêm ngặt. Và chúng ta sẽ thường bỏ qua từ “nghiêm ngặt” khi đề cập đến giả định này.

Khi chúng tôi thêm giả định về các cạnh nhân quả, các đường dẫn có hướng trong DAG mang một ý nghĩa rất đặc biệt; chúng tương ứng với quan hệ nhân quả. Điều này trái ngược với các đường dẫn khác trong biểu đồ, sự liên kết có thể diễn ra theo nhưng chắc chắn là không có quan hệ nhân quả. Điều này sẽ trở nên rõ ràng hơn khi chúng ta đi vào chi tiết về các loại đường dẫn khác này trong Phần 3.5 và 3.6.

Tiếp theo, bây giờ chúng ta sẽ coi các cạnh của đồ thị là quan hệ nhân quả, để mô tả các khái niệm một cách trực quan bằng ngôn ngữ nhân quả. Tuy nhiên, tất cả các tuyên bố liên kết về tính độc lập thống kê sẽ vẫn đúng, ngay cả khi các cạnh không có ý nghĩa nhân quả như trong mạng Bayesian vani ở Phần 3.2.

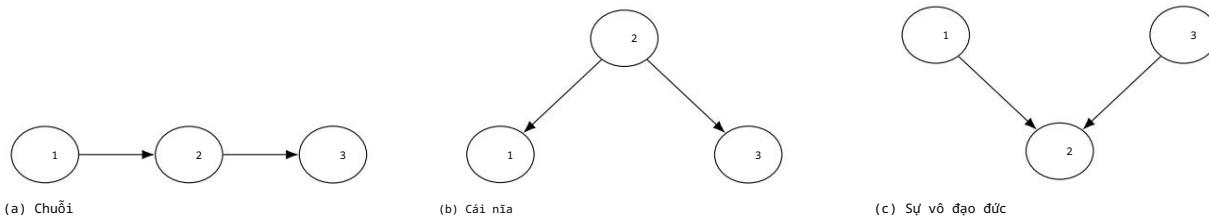
Như chúng ta sẽ thấy trong một số phần tiếp theo, các giả định chính mà chúng ta cần cho các mô hình đồ họa nhân quả để cho chúng ta biết mối liên hệ và luồng nhân quả giữa các biến như sau:

1. Giả định Markov địa phương (Giả định 3.1)
2. Giả định về biên nhân quả (Giả định 3.3)

Chúng ta sẽ thảo luận về những giả định này trong một số phần tiếp theo và quay lại thảo luận chúng đầy đủ hơn trong Phần 3.8 sau khi chúng ta đã thiết lập những bước sơ bộ cần thiết.

### 3.4 Đồ thị hai nút và khối xây dựng đồ họa

Bây giờ chúng ta đã hiểu rõ các giả định và định nghĩa cơ bản, chúng ta có thể đi đến cốt lõi của chương này: dòng liên kết và quan hệ nhân quả trong DAG. Chúng ta có thể hiểu luồng này trong DAG nói chung bằng cách hiểu luồng trong các khối xây dựng tối thiểu của biểu đồ. Các khối xây dựng tối thiểu này bao gồm các chuỗi (Hình 3.9a), các nhánh (Hình 3.9b), các điểm vô đạo đức (Hình 3.9c), hai nút không được kết nối (Hình 3.10) và hai nút được kết nối (Hình 3.11).



Hình 3.9: Các khái niệm cơ bản

Theo “luồng liên kết”, chúng tôi muốn nói đến việc liệu hai nút bất kỳ trong biểu đồ có được liên kết hay không liên kết. Một cách khác để nói điều này là liệu hai nút có phụ thuộc (thống kê) hay độc lập (thống kê). Ngoài ra, chúng ta sẽ nghiên cứu xem hai nút có độc lập có điều kiện hay không hay không.

Đối với mỗi khái niệm, chúng tôi sẽ đưa ra trực giác về lý do tại sao hai nút có độc lập (có điều kiện) hay không và chúng tôi cũng sẽ đưa ra bằng chứng.

Chúng ta có thể chứng minh rằng hai nút và độc lập có điều kiện với một số tập hợp nút bằng cách chỉ ra rằng  $(\cdot, \cdot | \cdot)$  phân tích thành nhân tử như  $(\cdot | \cdot)$  ( $\cdot | \cdot$ ). Nay giờ chúng ta sẽ thực hiện điều này trong trường hợp khái niệm cơ bản đơn giản nhất: hai nút không được kết nối.

Cho một biểu đồ chỉ có hai nút không được kết nối, như được mô tả trong Hình 3.10, các nút này không được liên kết đơn giản vì không có cạnh giữa chúng. Để chỉ ra điều này, hãy xem xét phân tích nhân tử của  $(\cdot, \cdot | \cdot)$ , mà phân tích mạng Bayesian (Định nghĩa 3.1) mang lại cho chúng ta:

$$(\cdot, \cdot | \cdot) = (\cdot | \cdot) (\cdot | \cdot) \quad (3.5)$$

Đó là nó; áp dụng hệ số hóa mạng Bayes ngay lập tức cho chúng ta bằng chứng rằng hai nút 1 và 2 không liên kết (độc lập) trong khái niệm này. Và giả định nào cho phép chúng ta chứng minh điều này? Đó là Markov đối với đồ thị trong Hình 3.10.

Ngược lại, nếu có một cạnh giữa hai nút (như trong Hình 3.11) thì hai nút đó được kết nối với nhau. Giả định mà chúng tôi tận dụng ở đây là giả định biên nhân quả (Giả định 3.3), có nghĩa là 1 là nguyên nhân của 2. Vì 1 là nguyên nhân của 2, nên 2 phải có khả năng thay đổi để đáp ứng với những thay đổi trong 1, do đó 2 và 1 liên kết với nhau. Nói chung, bắt cứ khi nào hai nút kề nhau trong đồ thị nhân quả, chúng liên kết với nhau.<sup>6</sup> Chúng ta sẽ thấy khái niệm tương tự này nhiều lần nữa trong Phần 3.5 và Phần 3.6.

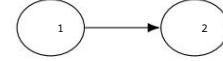
Nay giờ chúng ta đã đề cập đến các biểu đồ hai nút có liên quan, chúng ta sẽ đề cập đến luồng liên kết trong các khái niệm đồ họa còn lại (biểu đồ ba nút trong Hình 3.9), bắt đầu bằng biểu đồ chuỗi.

### 3.5 Dây xích và nĩa

Chuỗi (Hình 3.12) và nhánh (Hình 3.13) có cùng bộ phụ thuộc. Trong cả hai cấu trúc, 1 và 2 đều phụ thuộc, còn 2 và 3 phụ thuộc vì cùng một lý do mà chúng ta đã thảo luận ở cuối Phần 3.4. Các nút liền kề luôn phụ thuộc khi chúng ta đưa ra giả định về các cạnh nhân quả (Giả định 3.3). Còn 1 và 3 thì sao?

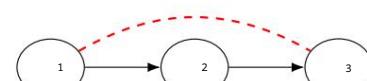


Hình 3.10: Hai nút chưa được kết nối



Hình 3.11: Hai nút được kết nối

<sup>6</sup> Hai nút liền kề trong biểu đồ nhân quả không chia sẻ có thể không được kết nối.

Hình 3.12: Chuỗi có dòng **liên kết** được vẽ bằng một vòng cung màu đỏ nét dày.

mặc dù? Sự liên kết có chảy từ 1 đến 3 trong chuỗi và nhánh không?

Thông thường, có, 1 và 3 được liên kết trong cả chuỗi và nhánh. Trong biểu đồ chuỗi, 1 và 3 thường phụ thuộc đơn giản vì 1 gây ra những thay đổi trong 2 và sau đó gây ra những thay đổi trong 3. Trong đồ thị phân nhánh, 1 và 3 cũng thường phụ thuộc. Điều này là do cùng một giá trị mà 2 nhận được sử dụng để xác định cả giá trị 1 nhận và giá trị 3 nhận. Nói cách khác, 1 và 3 được liên kết thông qua nguyên nhân chung (được chia sẻ) của chúng. Chúng tôi sử dụng từ "thường" trong suốt đoạn này vì tồn tại những trường hợp bệnh lý trong đó phân bố có điều kiện  $(2 | 1)$  và  $(3 | 2)$  bị căn chỉnh sai theo cách cụ thể khiến cho 1 và 3 không thực sự liên quan [xem, ví dụ: 15, Phần 2.2].

Một cách suy nghĩ bằng đồ họa trực quan về việc 1 và 3 được liên kết trong chuỗi và nhánh là trực quan hóa luồng liên kết. Chúng ta hình dung điều này bằng một đường nét đứt màu đỏ trong Hình 3.12 và Hình 3.13. Trong biểu đồ chuỗi (Hình 3.12), sự liên kết đi từ 1 đến 3 đọc theo đường dẫn 1 2 3. Một cách đối xứng, sự liên kết chảy từ 3 đến 1 đọc theo cùng một đường dẫn đó, chỉ chạy đối diện với các mũi tên.

Trong đồ thị ngã ba (Hình 3.13), sự liên kết đi từ 1 đến 3 đọc theo đường dẫn 1 2 3. Và tương tự, chúng ta có thể nghĩ về sự liên kết chảy từ 3 đến 1 đọc theo cùng một đường dẫn đó, giống như trường hợp với chuỗi. Nói chung, dòng liên kết là đối xứng.

Chuỗi và nhánh cũng có chung tính độc lập. Khi chúng ta đặt điều kiện vào 2 trong cả hai đồ thị, nó sẽ chặn luồng liên kết từ 1 đến 3. Điều này là do giả định Markov cục bộ; mỗi biến chỉ có thể phụ thuộc cục bộ vào cha mẹ của nó. Vì vậy, khi chúng ta đặt điều kiện vào 2 (mẹ của 3 trong cả hai đồ thị), 3 trở nên độc lập với 1 (và ngược lại).

Chúng ta sẽ đề cập đến tính độc lập này như ví dụ về đường dẫn bị chặn. Chúng tôi minh họa các đường dẫn bị chặn này trong Hình 3.14 và Hình 3.15. Điều hòa ngăn chặn dòng liên kết trong chuỗi và nhánh. Không có điều kiện, sự liên tưởng có thể tự do tuôn chảy trong những chuỗi và nhánh; chúng tôi sẽ coi đây là một đường dẫn không bị chặn. Tuy nhiên, tình hình hoàn toàn khác với những hành vi vô đạo đức, như chúng ta sẽ thấy trong phần tiếp theo.

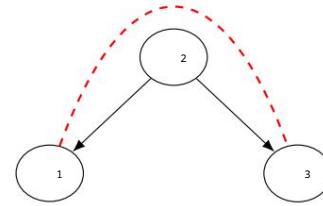
Tất cả đều là trực giác tốt đẹp, nhưng còn bằng chứng thì sao? Ta có thể chứng minh rằng  $1 \rightarrow 3 | 2$  chỉ sử dụng giả định Markov cục bộ. Chúng ta sẽ làm điều này bằng cách chứng minh rằng  $(1, 3 | 2) = (1 | 2) (3 | 2)$ . Chúng tôi sẽ hiển thị bằng chứng cho biểu đồ chuỗi. Việc bắt đầu với hệ số hóa Bayes thường rất hữu ích. Đối với chuỗi, chúng ta có thể phân tích  $(1, 2, 3)$  như sau:

$$(1, 2, 3) = (1) (2 | 1) (3 | 2) \quad (3.6)$$

Quy tắc Bayes cho chúng ta biết rằng  $(1, 3 | 2) = \frac{(1, 2, 3)}{(1 | 2)}$  nên ta có  $(3 | 2)$ ,

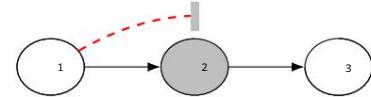
$$\frac{2) (1, 3 | 2)}{2) = \frac{(1) (2 | 1) (3 | 2)}{(1 | 2)} \quad (3.7)$$

Vì chúng ta đang muốn kết thúc với  $(1 | 2)$  ( $3 | 2$ ) và chúng ta đã có  $(3 | 2)$ , nên chúng ta phải biến phần còn lại thành  $(1 | 2)$ . Chúng ta có thể làm điều này bằng cách

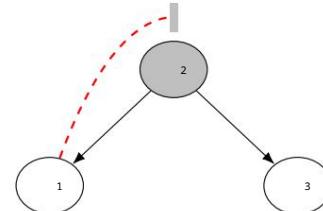


Hình 3.13: Fork với dòng **liên kết** được vẽ bằng một vòng cung màu đỏ nét đứt.

[15]: Ngọc trai và cộng sự. (2016), Suy luận nhân quả trong thống kê: Sơ lược



Hình 3.14: Chuỗi **liên kết** bị chặn bởi điều hòa trên 2.



Hình 3.15: Ngã ba có **liên kết** bị chặn bởi điều hòa trên 2.

một ứng dụng khác của quy tắc Bayes:

$$(1, 3 | 2) = \frac{(1, 2)}{(3 | 2)} = (1 | 2) \quad (3.8)$$

$$(3 | 2) \quad (3.9)$$

Bằng cách đó, chúng ta đã chứng minh được rằng  $1 | 3 | 2$ . Hãy tự mình thử; chứng minh sự tương tự trong fork.7

Dòng nhân quả Dòng liên tưởng có tính đối xứng, trong khi dòng nhân quả thì không. Theo giả định về ranh giới nhân quả (Giả định 3.3), quan hệ nhân quả chỉ diễn ra theo một hướng duy nhất. Nhân quả chỉ chảy dọc theo những con đường được định hướng. Hiệp hội chảy theo bất kỳ con đường nào không chứa đựng sự vô đạo đức.

7 [Bài tập đọc tích cực](#): chứng minh điều đó  
1 | 3 | 2 cho nia (Hình 3.15).

### 3.6 Máy va chạm và con cháu của chúng

Nhớ lại Phần 3.1 rằng chúng ta có một hành vi vô đạo đức khi có một đứa con mà cha và mẹ không có cạnh nối giữa họ (Hình 3.16). Và trong cấu trúc biểu đồ này, đứa trẻ được coi là một đứa con hoang. Không, chỉ đứa thôi; nó được gọi là máy va chạm.

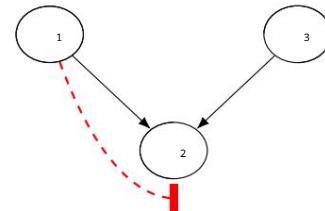
Ngược lại với xiềng xích và nia, trong vô đạo đức, 1 | 3. Nhìn vào cấu trúc biểu đồ và suy nghĩ về nó một chút. Tại sao 1 và 3 lại liên kết với nhau? Người này không phải là hậu duệ của người kia như trong dây chuyền và họ không có chung một mục đích như trong nia. Đúng hơn, chúng ta có thể coi 1 và 3 đơn giản là những sự kiện không liên quan xảy ra, xảy ra với cả hai đều góp phần tạo ra một số hiệu ứng chung (2). Để chỉ ra điều này, chúng tôi áp dụng hệ số hóa mạng Bayes và loại trừ 2:

$$(1, 3) = (1, 2, 3) \quad (3.10)$$

$$= (1) (3) (2 | 1, 3) \quad (3.11)$$

$$= (1) (3) (2 | 1, 3) \quad (3.12)$$

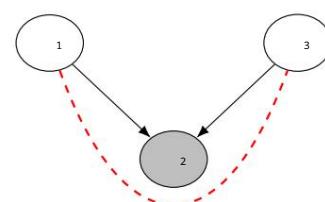
$$= (1) (3) \quad (3.13)$$



Hình 3.16: Vô đạo đức với **sự liên kết** bị chặn bởi máy va chạm.

Chúng ta minh họa tính độc lập của 1 và 3 trong Hình 3.16 bằng cách chỉ ra rằng liên kết mà chúng ta có thể tưởng tượng là chảy dọc theo đường 1 | 2 | 3 thực ra bị chặn tại 2. Bởi vì chúng ta có một máy va chạm trên đường dẫn nối 1 và 3, nên sự liên kết không đi qua đường dẫn đó. Đây là một ví dụ khác về con đường bị chặn, nhưng lần này con đường không bị chặn bởi điều kiện; đường đi bị chặn bởi một máy va chạm.

Đàn ông đẹp trai lại là những kẻ ngu ngốc Thật kỳ lạ, khi chúng ta tạo điều kiện cho máy va chạm 2, cha mẹ của nó 1 và 3 trở nên phụ thuộc (được mô tả trong Hình 3.17). Một ví dụ là cách dễ nhất để biết lý do tại sao lại như vậy. Hãy tưởng tượng rằng bạn đang hẹn hò với đàn ông và bạn nhận thấy rằng hầu hết những người đàn ông tử tế mà bạn gặp đều không đẹp trai lắm, và hầu hết những người đàn ông đẹp trai mà bạn gặp đều là những kẻ khốn nạn. Có vẻ như bạn phải lựa chọn giữa ngoại hình và lòng tốt. Nói cách khác, có vẻ như lòng tốt và ngoại hình có mối liên hệ tiêu cực. Tuy nhiên, điều gì sẽ xảy ra nếu tôi cũng nói với bạn rằng có một biến số quan trọng thứ ba ở đây: tính sẵn có (liệu nam giới có



Hình 3.17: Sự vô đạo đức với **sự liên kết** được bỏ chặn bằng cách điều hòa trên máy va chạm.

đã có quan hệ tình cảm hay chưa)? Và điều gì sẽ xảy ra nếu tôi nói với bạn rằng sự sẵn sàng của một người đàn ông phần lớn được quyết định bởi vẻ ngoài và lòng tốt của họ; nếu cả hai đều đẹp trai và tốt bụng thì họ đang yêu nhau. Những người đàn ông sẵn có là những người còn lại, những người không đẹp trai hoặc không tử tế. Bạn nhận thấy mối liên hệ giữa ngoại hình và lòng tốt bởi vì bạn đã có điều kiện về máy và chạm (sự sẵn có). Bạn chỉ nhìn vào những người đàn ông không có mối quan hệ. Bạn có thể thấy cấu trúc nhân quả của ví dụ này bằng cách lấy Hình 3.17 và thay thế 1 bằng "ngoại hình", 3 bằng "sự tử tế" và 2 bằng "sự sẵn có".

Ví dụ trên gợi ý một cách tự nhiên rằng, khi hẹn hò với đàn ông, có lẽ bạn nên cân nhắc việc không đặt điều kiện vào 2 = "không có quan hệ tình cảm" mà thay vào đó, đặt điều kiện vào 2 = "đang có mối quan hệ". Tuy nhiên, bạn có thể gặp phải các biến số khác 4 đưa đến những hành vi trái đạo đức mới ở đó.

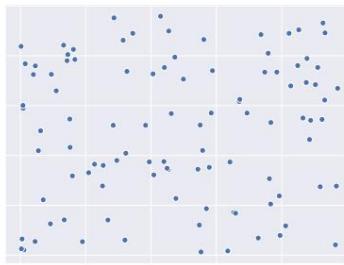
Những câu hỏi đạo đức như vậy nằm ngoài phạm vi của cuốn sách này.

Quay trở lại phạm vi của cuốn sách này, chúng ta thấy rằng việc điều hòa trên máy và chạm có thể biến một đường đi bị chặn thành một đường đi không bị chặn. Cha mẹ 1 và 3 không liên kết trong quần thể nói chung, nhưng khi chúng ta đặt điều kiện cho con chung 2 của chúng nhận một giá trị cụ thể, chúng sẽ liên kết. Điều hòa trên máy và chạm 2 cho phép liên kết chảy dọc theo đường 1 2 3, mặc dù thực tế là nó không xảy ra khi chúng ta không điều hòa trên 2. Chúng tôi minh họa điều này khi chuyển từ Hình 3.16 sang Hình 3.17.

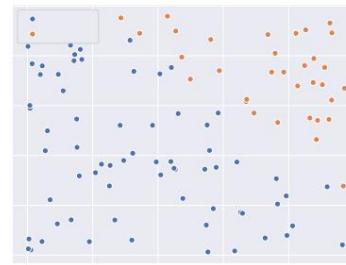
Chúng tôi cũng minh họa điều này bằng biểu đồ phân tán trong Hình 3.18. Trong Hình 3.18a, chúng ta vẽ biểu đồ toàn bộ dân số, với sự tử tế trên trực x và nhìn trên trực y. Như bạn có thể thấy, các biến không liên quan đến tổng thể chung. Tuy nhiên, nếu chúng ta loại bỏ những người đã có mối quan hệ (những người màu cam trong Hình 3.18b), chúng ta sẽ còn lại mối liên hệ tiêu cực rõ ràng mà chúng ta thấy trong Hình 3.18c. Hiện tượng này được gọi là nghịch lý Berkson. Việc chúng ta thấy mối liên hệ tiêu cực này đơn giản chỉ vì chúng ta đang chọn một tập hợp con thiên vị của tổng thể để xem xét, đó là lý do tại sao điều này đôi khi được gọi là thiên kiến lựa chọn [xem, ví dụ, 7, Chương 8]

**Bài tập đọc tích cực:** Hãy đưa ra ví dụ của riêng bạn về sự vô đạo đức và cách điều kiện tác động lên máy và chạm tạo ra sự liên kết giữa cha mẹ của nó. Gợi ý: hãy nghĩ về các sự kiện hiếm gặp cho 1 và 3 trong đó, nếu một trong hai sự kiện đó xảy ra thì kết quả 2 nào đó sẽ xảy ra.

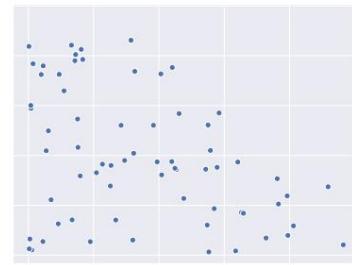
[7]: Hernán và Robins (2020), Suy luận nhân quả : Chuyện gì xảy ra nếu



(a) Dữ liệu về ngoại hình và lòng tốt của toàn bộ dân số. Vẽ bê ngoài và lòng tốt là độc lập.



(b) Dữ liệu về ngoại hình và lòng tốt được nhóm lại theo việc người đó có rắn hay không. Trong mỗi nhóm, có một mối tương quan tiêu cực.



(c) Dữ liệu về ngoại hình và lòng tốt chỉ dành cho những người có sẵn. Nay giờ, có một mối tương quan tiêu cực.

Hình 3.18: Dữ liệu ví dụ cho ví dụ "đàn ông đẹp trai lại là kẻ khốn nạn". Cả ngoại hình và lòng tốt đều là những giá trị liên tục trên thang điểm từ 0 đến 10.

Ví dụ bằng số Tất cả những điều trên nhằm cung cấp cho bạn trực giác về lý do tại sao việc điều kiện hóa một máy và chạm lại gây ra sự liên kết giữa cha mẹ của nó, nhưng chúng ta vẫn chưa đưa ra một ví dụ bằng số cụ thể về điều này. Chúng tôi sẽ đưa ra một cái đơn giản ở đây. Hãy xem xét việc tạo dữ liệu sau đây

(DGP), trong đó 1 và 3 được rút ra độc lập với phân phối chuẩn hóa và sau đó được sử dụng để tính 2:

$$1 \quad (\theta, 1), \quad 3 \quad (\theta, 1) \quad 2 \quad (3.14)$$

$$= 1 + 3 \quad (3.15)$$

Chúng ta đã tuyên bố rằng 1 và 3 là độc lập, nhưng để đặt hai phép tính cạnh nhau, hãy tính hiệp phương sai của chúng:

$$\begin{aligned} \text{Cov}(1, 3) &= [(1 \quad [1])(3 \quad [3])] = \\ &= [1 \quad 3] \quad (\text{trung bình}) \\ &= [1] \quad [3] \quad (\text{bằng } 0) \quad (\text{độc lập}) \\ &= 0 \end{aligned}$$

Bây giờ, hãy tính hiệp phương sai của chúng, có điều kiện theo 2:

$$\text{Cov}(1, 3 | 2 = ) = [1 \quad 3 | 2 = ] \quad (3.16)$$

$$= [1(1)] \quad (3.17)$$

$$= [1] \quad [2] \quad (3.18)$$

$$= 1 \quad (3.19)$$

Điều quan trọng là trong Công thức 3.17, chúng ta đã sử dụng Công thức 3.15 để thế vào 3 theo 1 và 2 (điều kiện theo ). Điều này dẫn đến số hạng bậc hai, dẫn đến phép tính cho một số khác 0, nghĩa là 1 và 3 liên kết với nhau, có điều kiện là 2.

Hậu duệ của máy va chạm Điều kiện hóa đối với con cháu của máy va chạm cũng gây ra sự liên kết giữa cha mẹ của máy va chạm. Trực giác là nếu chúng ta tìm hiểu điều gì đó về hậu duệ của máy va chạm, thì chúng ta cũng thường học được điều gì đó về bản thân máy va chạm vì có một đường nhân quả trực tiếp từ máy va chạm đến con cháu của nó và chúng ta biết rằng các nút trong chuỗi thường được liên kết với nhau (xem Phần 3.5), giả sử mức tối thiểu (Giả định 3.2). Nói cách khác, một hậu duệ của một máy va chạm có thể được coi là một đại diện cho máy va chạm đó, do đó, việc điều hòa đối với hậu duệ cũng tương tự như việc điều hòa đối với chính máy va chạm đó.

**Bài tập đọc tích cực:** Chúng tôi đã cung cấp một số kỹ thuật để suy nghĩ về máy va chạm: ví dụ cấp cao, ví dụ số và lý luận trừu tượng. Hãy sử dụng ít nhất một trong số chúng để thuyết phục bản thân rằng việc điều kiện hóa con cháu của máy va chạm có thể tạo ra sự liên kết giữa cha mẹ của máy va chạm.

## 3,7 d-tách

Trước khi định nghĩa sự phân tách d, chúng ta sẽ hệ thống hóa ý nghĩa của khái niệm "đường dẫn bị chặn", mà chúng ta đã thảo luận ở phần trước. phần:

**Định nghĩa 3.3** (đường dẫn bị chặn) Đường dẫn giữa các nút và bị chặn bởi bộ điều hòa (có khả năng trống) nếu một trong hai điều sau đây là đúng:

1. Dọc theo đường đi có một chuỗi . . . . . hoặc một ngã ba . . . . . , trong đó bị điều hòa bởi ( ).

2. Có một máy va chạm trên đường đi không bị điều hòa bởi ( ) và không có con cháu nào của nó bị điều hòa bởi (de( ) ).

Sau đó, một đường dẫn không bị chặn chỉ đơn giản là phần bổ sung; một con đường không bị chặn là một

con đường không bị chặn. Trực giác đồ họa cần lưu ý là sự liên kết diễn ra đọc theo các đường dẫn không bị chặn và liên kết không chảy đọc theo các đường dẫn bị chặn. Nếu bạn không có trực giác này trong đầu thì có lẽ bạn nên đọc lại hai phần trước với mục tiêu đạt được trực giác này. Bây giờ, chúng tôi đã sẵn sàng giới thiệu một khái niệm rất quan trọng: **sự tách d**.

**Định nghĩa 3.4 (tách d)** Hai (tập hợp) nút và được phân tách d bởi một tập hợp nút nếu tất cả các đường dẫn giữa (bất kỳ nút nào trong) và (bất kỳ nút nào trong) bị chặn bởi [16]

Nếu tất cả các đường dẫn giữa hai nút và bị chặn thì chúng ta nói rằng và được phân tách bằng d. Tương tự, nếu tồn tại ít nhất một đường đi giữa và không bị chặn thì chúng ta nói rằng và được kết nối d.

Như chúng ta sẽ thấy trong Định lý 3.1, phép tách d là một khái niệm quan trọng vì nó bao hàm sự độc lập có điều kiện. Chúng ta sẽ sử dụng ký hiệu | để biểu thị rằng và được phân tách bằng d trong đồ thị khi điều hòa trên . Tương tự, chúng ta sẽ sử dụng ký hiệu | để biểu thị rằng và độc lập trong phân bố khi điều hòa trên .

**Định lý 3.1** Cho rằng là Markov đối với (thỏa mãn giả định Markov địa phương, Giả định 3.1), nếu và phân tách d trong phụ thuộc vào , thì và độc lập trong phụ thuộc vào . Chúng ta có thể viết ngắn gọn điều này như sau:

| = | (3.20)

Bởi vì điều này rất quan trọng nên chúng ta sẽ đặt tên cho Phương trình 3.20 : giả định Markov toàn cục. Định lý 3.1 cho chúng ta biết rằng giả định Markov cục bộ bao hàm giả định Markov toàn cầu.

Giống như chúng tôi đã xây dựng trực giác gợi ý rằng giả định Markov cục bộ (Giả định 3.1) bao hàm hệ số hóa mạng Bayes (Định nghĩa 3.1) và cảnh báo bạn về thực tế rằng hệ số hóa mạng Bayes cũng bao hàm giả định Markov cục bộ (cả hai đều tương đương nhau) Ngoài ra), hóa ra giả định Markov toàn cầu cũng bao hàm giả định Markov cục bộ. Nói cách khác, giả định Markov cục bộ, giả định Markov toàn cầu và hệ số hóa mạng Bayes đều tương đương nhau [xem, ví dụ, 13, Chương 3]. Do đó, chúng tôi sẽ sử dụng cụm từ giả định Markov được rút ngắn một chút để đề cập đến các khái niệm này như một nhóm hoặc chúng tôi sẽ viết đơn giản là “ là Markov đối với ” để truyền đạt ý nghĩa tương tự.

[16]: Pearl (1988), Lý luận xác suất trong các hệ thống thông minh: Mạng lưới hợp lý Suy luận

**Bài tập đọc tích cực:** Để luyện tập về phép tách d, dưới đây là một số câu hỏi về phép tách d trong Hình 3.19.

Các câu hỏi về Hình 3.19a:

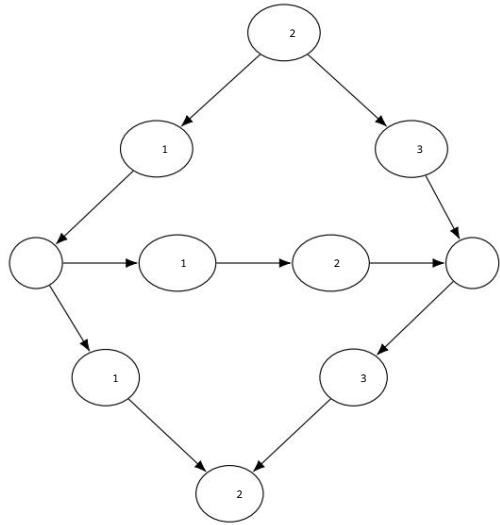
1. và d có được phân tách bằng tập rỗng không?
2. và d có cách nhau bởi 2 không?
3. và d có được phân tách bằng { 2, 1} không?
4. và d có được phân tách bằng { 1, 2} không?
5. và d có được phân tách bằng { 1, 2, 2} không?

[13]: Koller và Friedman (2009), Mô hình đồ họa xác suất: Nguyên tắc và kỹ thuật

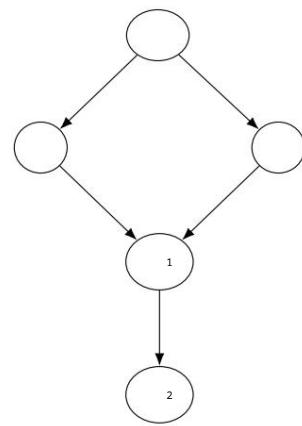
6. và d có được phân tách bằng  $\{1, 2, 2, 3\}$  không?

Các câu hỏi về Hình 3.19b:

1. và d có được phân tách bằng tập rỗng không?
2. và d có bị ngăn cách bởi không?
3. và d có được phân tách bằng  $\{1, 2\}$  không?



(Mô típ)



(b)

Hình 3.19: Đồ thị của bài tập tách d

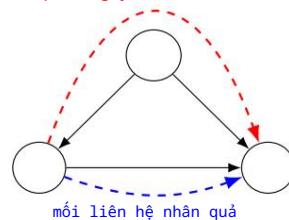
### 3.8 Dòng Hiệp hội và Nhân quả

Bây giờ chúng ta đã đề cập đến những bước sơ bộ cần thiết (chuỗi, nhánh, máy va chạm và phân tách d), điều cần nhấn mạnh là cách thức liên kết và quan hệ nhân quả trong đồ thị có hướng. Hiệp hội chảy dọc theo tất cả các con đường không bị chặn. Trong đồ thị nhân quả, quan hệ nhân quả diễn ra dọc theo các đường dẫn có hướng. Hãy nhớ lại ở Phần 1.3.2 rằng mỗi liên hệ không những không phải là mối quan hệ nhân quả mà mối quan hệ nhân quả còn là một tiểu loại của mỗi liên kết. Đó là lý do tại sao sự liên kết và nhân quả đều chảy dọc theo những con đường được định hướng.

Chúng tôi gọi dòng liên kết dọc theo các đường dẫn có định hướng là liên kết nhân quả. Một loại liên kết phi nhân quả phổ biến khiêm cho sự liên kết tổng thể không phải là mối quan hệ nhân quả là liên kết gây nhiễu. Trong biểu đồ ở Hình 3.20, chúng tôi mô tả mối liên hệ gây nhiễu màu đỏ và mối liên hệ nhân quả màu xanh lam.

Mạng Bayesian thông thường là các mô hình thống kê thuận túy, vì vậy chúng ta chỉ có thể nói về luồng liên kết trong mạng Bayesian. Tuy nhiên, sự liên kết vẫn diễn ra theo cách giống hệt nhau trong các mạng Bayesian giống như trong các biểu đồ nhân quả. Trong cả hai, sự liên kết diễn ra dọc theo chuỗi và các nhánh, trừ khi có một nút được điều chỉnh. Và trong cả hai, một máy va chạm sẽ chặn dòng liên kết, trừ khi nó bị điều kiện hóa. Kết hợp các khái niệm này, chúng ta biết được cách liên kết diễn ra trong các DAG nói chung. Chúng ta có thể biết liệu hai nút có được liên kết hay không (không có luồng liên kết nào giữa chúng) bằng cách chúng có được phân tách bằng d hay không.

hiệp hội gây nhiễu



Hình 3.20: Biểu đồ nhân quả mô tả một ví dụ về luồng liên kết gây nhiễu và liên kết nhân quả.

Đồ thị nhân quả đặc biệt ở chỗ chúng ta còn giả định thêm rằng các cạnh có ý nghĩa nhân quả (giả định cạnh nhân quả, Giả định 3.3). Giả định này là yếu tố đưa quan hệ nhân quả vào các mô hình của chúng tôi và nó làm cho một loại đường dẫn mang một ý nghĩa hoàn toàn mới: đường dẫn có hướng. Giả định này ban cho những con đường được định hướng có vai trò duy nhất là mang theo quan hệ nhân quả đọc theo chúng. Ngoài ra, giả định này không đối xứng; “ là nguyên nhân của ” không giống như nói “ là nguyên nhân của .” Điều này có nghĩa là có một sự khác biệt quan trọng giữa mối liên hệ và mối quan hệ nhân quả: sự liên kết có tính đối xứng, trong khi mối quan hệ nhân quả là không đối xứng.

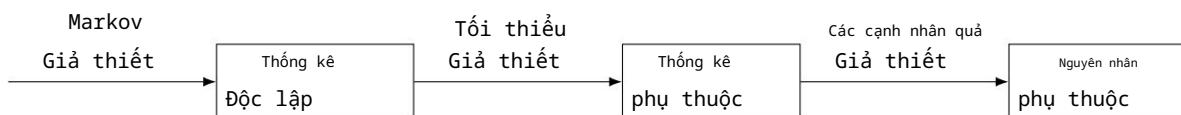
d-tách ngữ ý Hiệp hội là quan hệ nhân quả Cho rằng chúng ta có các công cụ để do lường sự liên kết, làm thế nào chúng ta có thể cài đặt quan hệ nhân quả? Nói cách khác, làm thế nào chúng ta có thể đảm bảo rằng mối liên hệ mà chúng ta đo lường là quan hệ nhân quả, chẳng hạn, để đo lường tác động nhân quả của lên ? Chà, chúng ta có thể làm điều đó bằng cách đảm bảo rằng không có mối liên hệ phi nhân quả nào giữa và . Điều này đúng nếu và được phân tách bằng d trong biểu đồ tăng cường trong đó chúng ta loại bỏ các cạnh ra khỏi . Điều này là do tất cả tác động nhân quả của lên sẽ chảy qua các cạnh hướng ra ngoài của nó, vì vậy một khi chúng bị loại bỏ, mối liên kết duy nhất còn lại hoàn toàn là mối liên kết phi nhân quả.

Trong Hình 3.21, chúng tôi minh họa những gì mà giả định quan trọng mang lại cho chúng ta về mặt diễn giải luồng liên kết này. Đầu tiên, chúng ta có giả định Markov (cục bộ/toàn cầu) (Giả định 3.1). Như chúng ta đã thấy trong Phần 3.7, giả định này cho phép chúng ta biết nút nào không được liên kết. Nói cách khác, giả định Markov cho biết mối liên kết này không đi theo những con đường nào . Khi chúng ta tăng cường một chút giả định Markov lên giả định tối thiểu (Giả định 3.2), chúng ta sẽ biết được đường liên kết nào sẽ diễn ra đọc theo (ngoại trừ trong các trường hợp cạnh nội động). Khi chúng ta bổ sung thêm giả định về các cạnh nhân quả (Giả định 3.3), chúng ta nhận được rằng quan hệ nhân quả đó chảy đọc theo các đường dẫn có hướng.

Vì vậy, hai giả định sau đây là cần thiết cho đồ họa mô hình nhân quả:

1. Giả định Markov (Giả định 3.1)
2. Giả định về biên nhân quả (Giả định 3.3)

“ Hãy nhớ lại rằng phần đầu tiên của giả định cực tiểu chỉ là giả định Markov cục bộ và phần thứ hai nằm trong giả định biên nhân quả.

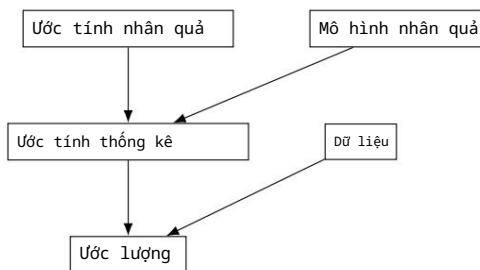


Hình 3.21: Lưu đồ minh họa loại tuyên bố nào chúng tôi có thể đưa ra về dữ liệu của mình khi chúng tôi thêm từng giả định quan trọng bổ sung.

# 4

## Mô hình nhân quả

Các mô hình nhân quả là cần thiết để xác định các đại lượng nhân quả. Khi chúng tôi đã trình bày lại Sơ đồ nhận dạng-Ước tính (Hình 2.5) trong Phần 2.4, chúng tôi đã mô tả sự nhận dạng là quá trình di chuyển từ ước tính nhân quả đến ước tính thống kê. Tuy nhiên, để làm được điều đó, chúng ta phải có một mô hình nhân quả. Chúng tôi mô tả phiên bản đầy đủ hơn này của Sơ đồ nhận dạng-Ước tính trong Hình 4.1.



Hình 4.1: Sơ đồ nhận dạng-Ước tính - sơ đồ minh họa quy trình chuyển từ ước tính nhân quả mục tiêu sang ước tính tương ứng, thông qua việc xác định và ước tính. Ngược lại với Hình 2.5, phiên bản này được tăng cường bằng mô hình nhân quả và dữ liệu.

Chương trước cung cấp trực giác đồ họa cho các mô hình nhân quả, nhưng nó không giải thích cách xác định số lượng nhân quả và chính thức hóa quan hệ nhân quả các mô hình. Chúng ta sẽ làm điều đó trong chương này.

### 4.1 Toán tử do và can thiệp

#### Phân phối

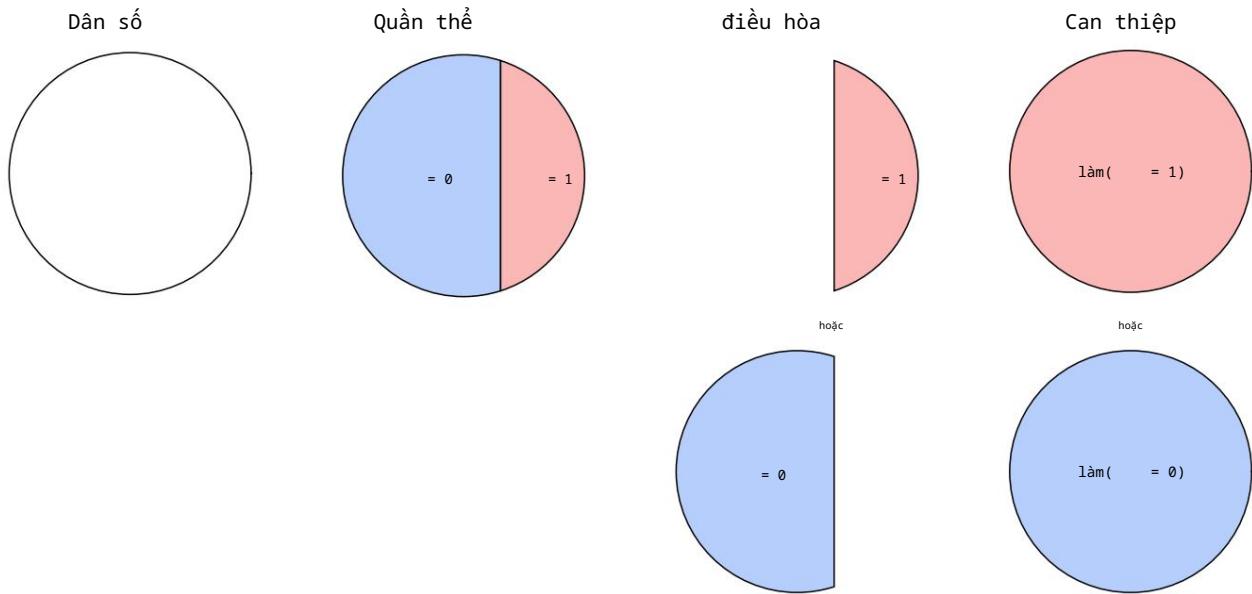
Điều đầu tiên chúng tôi sẽ giới thiệu là một toán tử toán học cho sự can thiệp. Trong ký hiệu thông thường cho xác suất, chúng ta có điều kiện, nhưng điều đó không có nghĩa là can thiệp. Điều hòa trên  $=$  chỉ có nghĩa là rằng chúng ta đang hạn chế sự tập trung của mình vào một nhóm nhỏ dân số đối với những người đó người đã được điều trị  $.$  Ngược lại, sự can thiệp sẽ là thực hiện toàn dân và điều trị cho mọi người  $.$  Chúng tôi minh họa điều này trong Hình 4.2. Chúng ta sẽ biểu thị sự can thiệp bằng toán tử do: do( $\quad = \quad$ ) $.$  Đây là ký hiệu thường được sử dụng trong các mô hình nhân quả đồ họa và nó có tương đương trong ký hiệu kết quả tiềm năng. Ví dụ, chúng ta có thể viết sự phân bổ kết quả tiềm năng  $(\quad)$  mà chúng ta đã thấy ở Chương 2 như sau:

$$( \quad ( \quad ) = \quad ) \quad ( \quad = \quad | \text{làm}( \quad = \quad )) \quad ( \quad | \text{làm}( \quad )) \quad (4.1)$$

Lưu ý rằng chúng ta rút gọn do( $\quad = \quad$ ) thành chỉ do( $\quad$ ) trong tùy chọn cuối cùng trong Phương trình 4.1. Chúng tôi sẽ sử dụng cách viết tắt này xuyên suốt cuốn sách. Chúng ta có thể tương tự viết ATE (hiệu quả xử lý trung bình) khi xử lý ở dạng nhị phân là sau:

$$[ \quad | \text{do}( \quad = 1)] \quad [ \quad | \text{làm}( \quad = 0)] \quad (4.2)$$

4.1 Toán tử do và Inter-	
Phân phối thông thường . . .	32
4.2 Giả định chính: Tính mô	
dun . . . . .	34
4.3 Hệ số rút gọn . . .	35
Ứng dụng mẫu và xem lại	
"Hiệp hội không phải là	
Nhân quả" . . . . .	36
4.4 Điều chỉnh cửa sau	37
Liên quan đến tiềm năng Out-	
đến . . . . .	39
4.5 Mô hình nhân quả mang tính cấu trúc	
(SCM) . . . . .	40
Phương trình cấu trúc . . . .	40
Can thiệp . . . . .	42
Xu hướng Collider và tại sao không nên	
Điều kiện về Hậu duệ	
của Điều trị . . . . .	43
4.6 Ví dụ ứng dụng của	
Điều chỉnh cửa sau . . . .	44
Hiệp hội và nhân quả trong	
một ví dụ về đồ chơi . . . . .	44
Một ví dụ hoàn chỉnh với	
Ước tính . . . . .	45
4.7 Xem lại các giả định . . .	47



Hình 4.2: Minh họa sự khác biệt giữa điều hòa và can thiệp

Chúng ta thường làm việc với các phân phối đầy đủ như  $(\dots | \text{do}(\dots))$ , thay vì phương tiện của chúng, vì điều này tổng quát hơn; nếu chúng ta mô tả đặc trưng  $(\dots | \text{do}(\dots))$ , thì chúng ta mô tả  $[\dots | \text{làm}(\dots)]$ . Chúng ta thường coi  $(\dots | \text{do}(\dots = \dots))$  và các biểu thức khác có toán tử do trong đó là phân phối can thiệp .

Các phân phối can thiệp như  $(\dots | \text{do}(\dots = \dots))$  về mặt khái niệm khá khác với phân phối quan sát  $(\dots)$ . Các phân phối quan sát như  $(\dots)$  hoặc  $(\dots, \dots, \dots)$  không có toán tử do trong đó. Vì chúng không có toán tử do nên chúng ta có thể quan sát dữ liệu từ chúng mà không cần thực hiện bất kỳ thử nghiệm nào. Đây là lý do tại sao chúng ta gọi dữ liệu từ dữ liệu quan sát  $(\dots, \dots, \dots)$ . Nếu chúng ta có thể rút gọn một biểu thức có do trong đó (một biểu thức can thiệp) thành một biểu thức không có do trong đó (một biểu thức quan sát), thì được cho là có thể nhận dạng được. Một biểu thức có do trong đó về cơ bản khác với một biểu thức không có do trong đó, mặc dù thực tế là trong ký hiệu do, do xuất hiện sau một thanh điều kiện thông thường. Như chúng ta đã thảo luận trong Phần 2.4, chúng ta sẽ coi ước tính là ước tính nhân quả khi nó chứa toán tử do và chúng ta coi ước tính là ước tính thống kê khi nó không chứa toán tử do.

toán tử do.

Bất cứ khi nào,  $\text{do}(\dots)$  xuất hiện sau thanh điều kiện, điều đó có nghĩa là mọi thứ trong biểu thức đó đều ở thế giới hậu can thiệp , nơi xảy ra sự can thiệp  $\text{do}(\dots)$  . Ví dụ:  $[\dots | \text{do}(\dots), \dots = \dots]$  đề cập đến kết quả mong đợi trong tiêu quần thể trong đó  $\dots = \dots$  sau khi toàn bộ tiêu quần thể đã thực hiện điều trị  $\dots$  . Ngược lại,  $[\dots | \dots = \dots]$  chỉ đơn giản đề cập đến giá trị kỳ vọng trong dân số (trước can thiệp) nơi các cá nhân thực hiện bất kỳ phương pháp điều trị nào mà họ thường thực hiện  $(\dots)$  . Sự khác biệt này sẽ trở nên quan trọng khi chúng ta đề cập đến phản ứng tinh tế trong Chương 14.

## 4.2 Giả định chính: Tính mô đun

Trước khi có thể mô tả một giả định rất quan trọng, chúng ta phải xác định cơ chế nhân quả là gì. Có một số cách khác nhau để suy nghĩ về cơ chế nhân quả. Trong phần này, chúng ta sẽ đề cập đến cơ chế nhân quả tạo ra là phân bố có điều kiện của với tất cả các nguyên nhân của nó:

( | pa ). Như chúng tôi trình bày bằng đồ họa trong Hình 4.3, cơ chế nhân quả tạo ra là tất cả cha mẹ của và các cạnh của chúng đi vào . Chúng tôi sẽ đưa ra một mô tả cụ thể hơn một chút về cơ chế nhân quả là gì trong Phần 4.5.1, nhưng hiện tại những điều này là đủ.

Để có được nhiều kết quả xác định nguyên nhân, giả định chính mà chúng tôi sẽ đưa ra là các biện pháp can thiệp mang tính cục bộ. Cụ thể hơn, chúng ta sẽ giả định rằng việc can thiệp vào một biến chỉ làm thay đổi cơ chế nhân quả của ; nó không thay đổi cơ chế nhân quả tạo ra bất kỳ biến số nào khác. Theo nghĩa này, các cơ chế nhân quả có tính mô-đun.

Các tên khác được sử dụng cho thuộc tính mô-đun là cơ chế độc lập , quyền tự chủ và tính bất biến. Bây giờ chúng ta sẽ phát biểu giả định này một cách chính thức hơn.

### Giả định 4.1 (Tính mô đun / Cơ chế độc lập / Tính bất biến)

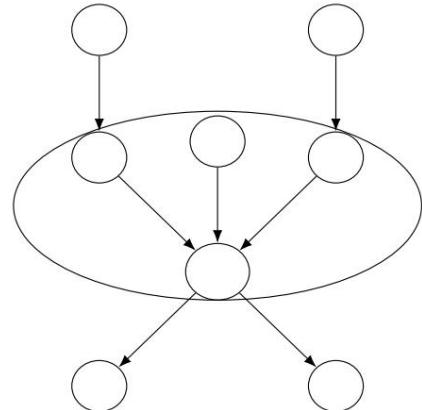
Nếu chúng ta can thiệp vào một tập hợp nút <sup>1</sup> đặt chúng thành hằng số, sau đó cho [ ], tất cả , thì chúng ta có:

1. Nếu thì ( | pa ) không thay đổi.
2. Nếu , thì ( | pa ) = 1 nếu là giá trị mà được thiết lập bởi sự can thiệp; ngược lại, ( | pa ) = 0.

Trong phần thứ hai của giả định trên, chúng ta có thể nói ( | pa ) = 1 nếu phù hợp với can thiệp2 và 0 nếu ngược lại. Rõ ràng hơn, chúng tôi sẽ nói (trong tương lai) rằng nếu , giá trị phù hợp với can thiệp nếu bằng giá trị mà được đặt trong can thiệp.

Giả định về tính mô đun là điều cho phép chúng ta mã hóa nhiều phân phối can thiệp khác nhau trong một biểu đồ duy nhất. Ví dụ: có thể xảy ra trường hợp ( ), ( | do( = ) ), ( | do( = ) ), và ( | do( 2 = 2 ) ) đều là những bản phân phối hoàn toàn khác nhau và hầu như không có gì chia sẻ. Nếu đúng như vậy thì mỗi phân phối này sẽ cần có biểu đồ riêng. Tuy nhiên, bằng cách giả định tính mô đun, chúng ta có thể mã hóa tất cả chúng bằng cùng một biểu đồ mà chúng ta sử dụng để mã hóa khớp ( , , 2, . . . . . ) và chúng ta có thể biết rằng tất cả các yếu tố (ngoại trừ những yếu tố được can thiệp ) on) được chia sẻ trên các biểu đồ này.

Biểu đồ nhân quả cho các phân bố can thiệp chỉ đơn giản là cùng một biểu đồ được sử dụng cho phân phối khớp quan sát, nhưng đã loại bỏ tất cả các cạnh đối với (các) nút bị can thiệp. Điều này là do xác suất của yếu tố can thiệp đã được đặt thành 1, vì vậy chúng ta có thể bỏ qua yếu tố đó (đây là trọng tâm của phần tiếp theo). Một cách khác để thấy rằng nút bị can thiệp không có nút gốc nhân quả là nút bị can thiệp được đặt thành một giá trị không đổi, do đó, nó không còn phụ thuộc vào bất kỳ biến nào mà nó phụ thuộc vào trong bài đặt quan sát (mẹ của nó). Biểu đồ đã loại bỏ các cạnh được gọi là biểu đồ thao tác.

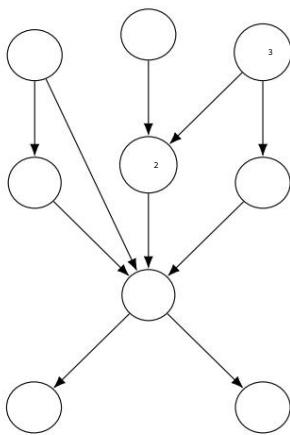


Hình 4.3: Đồ thị nhân quả với cơ chế nhân quả sinh ra được mô tả bên trong một hình elip.

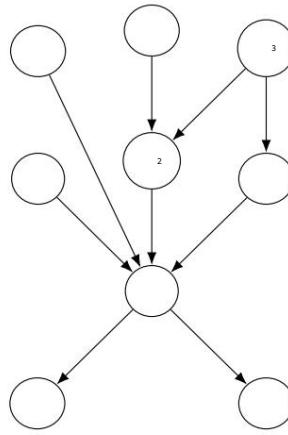
<sup>1</sup>Chúng ta sử dụng [ ] để chỉ tập hợp {1, 2, . . . , }.

<sup>2</sup> Vâng, từ “nhất quán” cực kỳ quá tải.

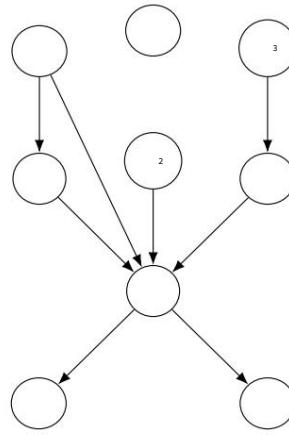
Ví dụ, hãy xem xét biểu đồ nhân quả cho một phân bố quan sát trong Hình 4.4a. Cả  $(\dots | do(\dots = \dots))$  và  $(\dots | do(\dots = \dots))$  đều tương ứng với đồ thị nhân quả trong Hình 4.4b, trong đó cạnh tói đã bị loại bỏ. Tương tự,  $(\dots | do(\dots = \dots))$  tương ứng với đồ thị trong Hình 4.4c, trong đó các cạnh tói của  $\dots = \dots$  đã bị loại bỏ. Mặc dù nó không được biểu thị trong các biểu đồ (chỉ biểu thị sự độc lập có điều kiện và quan hệ nhân quả), theo giả định tính mô đun,  $(\dots)$ ,  $(\dots | \dots = \dots)$  và  $(\dots | do(\dots = \dots))$  tất cả đều có chung những yếu tố giống nhau (không bị can thiệp).



(a) Đồ thị nhân quả cho phân bố quan sát  
sự phân chia



(b) Đồ thị nhân quả sau can thiệp trên  
(phân phôi can thiệp)



(c) Đồ thị nhân quả sau can thiệp trên  
(phân phôi can thiệp) 2

Hình 4.4: Can thiệp bằng cách xóa cạnh trong đồ thị nhân quả

Việc giả định tính mô đun bị vi phạm có nghĩa là gì?

Hãy tưởng tượng rằng bạn can thiệp vào  $\dots$  và điều này khiến cơ chế tạo ra một nút khác  $\dots$  thay đổi; sự can thiệp vào  $\dots$  thay đổi  $(\dots | pa \dots)$ , trong đó  $\dots$ . Nói cách khác, sự can thiệp không mang tính cục bộ đối với nút mà bạn can thiệp vào; cơ chế nhân quả không bắt biến khi bạn thay đổi cơ chế nhân quả khác; các cơ chế nhân quả không mang tính môđun.

Giả định này quan trọng đến mức Judea Pearl đề cập đến một phiên bản có liên quan chặt chẽ (mà chúng ta sẽ xem trong Phần 4.5.2) là Quy luật phản thực tế (và can thiệp), một trong hai nguyên tắc chính mà tất cả các kết quả nhân quả khác đều tuân theo.<sup>3</sup> Ngẫu nhiên, việc kết hợp giả định tính mô đun (Giả định 4.1) và giả định Markov (nguyên tắc then chốt khác) sẽ cho chúng ta mạng lưới Bayes nhân quả. Bây giờ chúng ta sẽ chuyển sang một trong những kết quả quan trọng rút ra từ những giả định này.

<sup>3</sup> Nguyên tắc quan trọng khác là giả định Markov toàn cục (Định lý 3.1), giả định rằng sự phân tách d hàm ý sự độc lập có điều kiện.

### 4.3 Hệ số rút gọn

Nhớ lại phép phân tích mạng Bayes (Định nghĩa 3.1), nó cho chúng ta biết rằng nếu  $\dots$  là Markov đối với đồ thị  $\dots$ , thì  $\dots$  phân tích thành nhân tử như sau:

$$(1, \dots, \dots) = (\dots | pa \dots) \quad (4.3)$$

trong đó  $pa \dots$  biểu thị cha mẹ của  $\dots$  trong  $\dots$ . Bây giờ, nếu chúng ta can thiệp vào một số tập hợp nút  $\dots$  và giả sử tính mô đun (Giả định 4.1), thì tất cả các thừa số sẽ giữ nguyên ngoại trừ các thừa số cho  $\dots$ ; những yếu tố đó

nên thay đổi thành 1 (đổi với các giá trị phù hợp với can thiệp) vì các biến đó đã được can thiệp. Đây là cách chúng ta có được hệ số cắt ngắn.

Mệnh đề 4.1 (Hệ số rút gọn) Chúng ta giả sử rằng và thỏa mãn giả định Markov và tính mâu thuẫn hóa. Cho trước một tập hợp các nút can thiệp , nếu phù hợp với can thiệp thì

$$(1, \dots, | do(\_ = \_)) = (\_ | pa\_). \quad (4.4)$$

Ngược lại,  $(1, \dots, | do(\_ = \_)) = \emptyset$ .

Điều quan trọng đã thay đổi khi chúng ta chuyển từ phân tích nhân tử thông thường trong Công thức 4.3 sang phân tích nhân tử rút gọn trong Công thức 4.4 là tích của Công thức 4.4 chỉ lớn hơn chứ không phải tất cả . Nói cách khác, các thừa số của đã bị cắt bớt.

#### 4.3.1 Ứng dụng ví dụ và xem lại "Hiệp hội là Không phải nhân quả"

Để thấy được sức mạnh mà hệ số rút gọn mang lại cho chúng ta, hãy áp dụng nó để xác định tác động nhân quả của việc điều trị lên kết quả trong một biểu đồ đơn giản. Cụ thể, chúng ta sẽ xác định đại lượng nhân quả  $(\_ | do(\_))$ . Trong ví dụ này, phân bố là Markov đối với đồ thị trong Hình 4.5. Hệ số hóa mạng Bayes (từ giả định Markov), cho chúng ta những điều sau:

$$(\_, \_ | \_) = (\_) (\_ | \_) (\_ | \_) \quad (4.5)$$

Khi chúng tôi can thiệp vào việc xử lý, việc phân tích nhân tử rút gọn (từ việc thêm giả định mâu thuẫn) cho chúng tôi kết quả sau:

$$(\_, \_ | \_ \text{ làm } \_) = (\_) (\_ | \_) \quad (4.6)$$

Sau đó, chúng ta chỉ cần loại trừ để đạt được điều mình muốn:

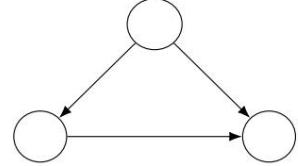
$$(\_ | \_ \text{ làm } \_) = (\_ | \_) (\_ | \_) \quad (4.7)$$

Chúng ta giả sử là rỗi rạc khi tính tổng các giá trị của nó, nhưng chúng ta có thể đơn giản thay bằng tích phân nếu là liên tục. Trong suốt cuốn sách này, điều đó sẽ xảy ra nên chúng tôi thường chỉ ra điều đó.

Nếu chúng ta điều chỉnh phương trình 4.7 một chút, chúng ta có thể thấy rõ ràng sự liên kết không phải là quan hệ nhân quả như thế nào. Đối tác liên kết thuận túy của  $(\_ | do(\_))$  là  $(\_ | \_)$ . Nếu  $(\_ | \_)$  trong Phương trình 4.7 là  $(\_ | \_)$ , thì chúng ta thực sự sẽ phục hồi được  $(\_ | \_)$ . Chúng tôi trình bày ngắn gọn điều này:

$$(\_ | \_ | \_) (\_ | \_) = (\_ | \_) (\_ | \_) \quad (4.8)$$

$$= (\_ | \_) \quad (4.9)$$



Hình 4.5: Cấu trúc nhân quả đơn giản trong đó làm xáo trộn tác dụng của lên và trong đó là yếu tố gây nhiễu duy nhất.

Điều này mang lại một số tính cụ thể cho sự khác biệt giữa sự liên kết và quan hệ nhân quả. Trong ví dụ này (đại diện cho một phạm vi rộng hơn

hiện tượng), sự khác biệt giữa  $(\quad | \text{do}(\quad))$  và  $(\quad | \quad)$  là sự khác biệt giữa  $(\quad)$  và  $(\quad | \quad)$ .

Để làm tròn ví dụ này, giả sử là biến ngẫu nhiên nhị phân và chúng tôi muốn tính ATE.  $(\quad | \text{do}(\quad = 1))$  là phân phối của  $(1)$ , vì vậy chúng ta chỉ cần lấy kỳ vọng để nhận được  $[ \quad (1) ]$ . Tương tự, chúng ta có thể làm điều tương tự với  $(0)$ . Khi đó, chúng ta có thể viết ATE như sau:

$$[ \quad (1) - (0) ] = (\quad | \text{làm}(\quad = 1)) - (\quad | \text{làm}(\quad = 0)) \quad (4.10)$$

Sau đó, nếu chúng ta thay Phương trình 4.7 vào cho  $(\quad | \text{do}(\quad = 1))$  và  $(\quad | \text{do}(\quad = 0))$ , chúng ta có một ATE được xác định đầy đủ. Với biểu đồ đơn giản trong Hình 4.5, chúng tôi đã chỉ ra cách chúng tôi có thể sử dụng hệ số rút gọn để xác định các tác động nhân quả trong các Công thức 4.5 đến 4.7. Bây giờ chúng ta sẽ khái quát quá trình nhận dạng này thành một công thức tổng quát hơn.

## 4.4 Điều chỉnh cửa sau

Nhớ lại ở Chương 3 rằng sự liên kết nhân quả chảy từ  $\rightarrow$  đến  $\rightarrow$  đọc theo các đường dẫn có hướng và mỗi liên kết phi nhân quả đó chảy đọc theo bất kỳ đường dẫn nào khác từ  $\rightarrow$  đến  $\rightarrow$  mà không bị chặn bởi 1) một yếu tố không va chạm bị điều hòa bởi hoặc 2) một máy va chạm không được điều chỉnh. Những đường dẫn không bị chặn không định hướng này từ  $\rightarrow$  đến  $\rightarrow$  được gọi là đường dẫn cửa sau vì chúng có một cạnh đi vào "cửa sau" của nút  $\rightarrow$ . Và hóa ra là nếu chúng ta có thể chặn những đường đi này bằng cách điều hòa, thì chúng ta có thể xác định được các đại lượng nhân quả như  $(\quad | \text{do}(\quad))$ .<sup>4</sup>

Đây chính xác là những gì chúng tôi đã làm trong phần trước. Chúng ta đã chặn đường dẫn cửa sau  $\rightarrow$  trong Hình 4.5 đơn giản bằng cách điều chỉnh  $\rightarrow$  và loại nó ra ngoài (Phương trình 4.7). Trong phần này, chúng ta sẽ khái quát hóa phương trình 4.7 thành các DAG tùy ý. Nhưng trước khi làm điều đó, chúng ta hãy xem xét bằng đồ họa tại sao đại lượng  $(\quad | \text{do}(\quad))$  hoàn toàn là nhân quả.

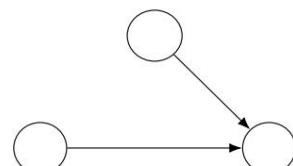
Như chúng ta đã thảo luận trong Phần 4.2, đồ thị cho phân phối can thiệp  $(\quad | \text{do}(\quad))$  giống với đồ thị cho phân phối quan sát  $(\quad, \quad, \quad)$ , nhưng có các cạnh tới LOẠI BỎ. Ví dụ: nếu lấy đồ thị từ Hình 4.5 và can thiệp vào  $\rightarrow$ , thì chúng ta thu được đồ thị bị thao tác trong Hình 4.6. Trong biểu đồ bị thao tác này, không thể có bất kỳ đường dẫn cửa sau nào vì không có cạnh nào đi vào cửa sau của  $\rightarrow$ . Do đó, tất cả mối liên hệ diễn ra từ  $\rightarrow$  đến  $\rightarrow$  trong đồ thị bị thao tác hoàn toàn là quan hệ nhân quả.

Bởi sự lắc đè do sang một bên, hãy chứng minh rằng chúng ta có thể xác định được  $(\quad | \text{do}(\quad))$ . Chúng tôi muốn biến ước tính nhân quả  $(\quad | \text{do}(\quad))$  thành ước tính thống kê (chỉ dựa vào phân phối quan sát). Chúng ta sẽ bắt đầu với việc giả sử rằng chúng ta có một tập hợp các biến thỏa mãn tiêu chí cửa sau:

**Định nghĩa 4.1 (Tiêu chí cửa sau)** Một tập hợp các biến thỏa mãn tiêu chí cửa sau liên quan đến  $\rightarrow$  và nếu các điều sau đây đúng:

1. chặn tất cả các đường dẫn cửa sau từ  $\rightarrow$  đến  $\rightarrow$ .
2. không chứa bất kỳ hậu duệ nào của  $\rightarrow$ .

<sup>4</sup> Như chúng tôi đã đề cập trong Phần 3.8, việc chặn tất cả các đường dẫn cửa sau tương đương với việc có sự phân tách  $\rightarrow$  trong biểu đồ nơi các cạnh đi ra ngoài  $\rightarrow$  bị loại bỏ. Điều này là do đây là những cạnh duy nhất mà quan hệ nhân quả chảy đọc theo, nên một khi chúng bị loại bỏ, tất cả những gì còn lại chỉ là sự liên kết phi nhân quả.



Hình 4.6: Đồ thị bị thao tác là kết quả của việc can thiệp vào  $\rightarrow$ , khi đồ thị gốc là Hình 4.5.

<sup>5</sup> **Bài tập đọc tích cực:** Trong DAG tổng quát, tập hợp nút nào liên quan đến  $\rightarrow$  sẽ luôn là tập hợp điều chỉnh đầy đủ? Tập hợp nút nào liên quan đến  $\rightarrow$  sẽ luôn là tập hợp điều chỉnh đầy đủ?

Việc thỏa mãn tiêu chí cửa sau làm cho trở thành một bộ điều chỉnh đủ.<sup>5</sup>  
 Chúng ta đã xem ví dụ về là một bộ điều chỉnh đầy đủ trong Phần 4.3.1.  
 Vì chỉ có một đường dẫn cửa sau duy nhất trong Phần 4.3.1 nên một nút ( ) là đủ để chặn tất cả các đường dẫn cửa sau, nhưng nói chung, có thể có nhiều đường dẫn cửa sau.

Để đưa vào chứng minh, chúng ta sẽ sử dụng thủ thuật điều hòa thông thường đối với các biến và loại chúng ra ngoài lề:

$$( \quad | \text{làm}( \quad ) ) = ( \quad | \text{làm}( \quad ), \quad ) ( \quad | \text{làm}( \quad ) ) \quad (4.11)$$

Cho rằng thỏa mãn tiêu chí cửa sau, chúng ta có thể viết như sau:

$$( \quad | \text{làm}( \quad ), \quad ) ( \quad | \text{làm}( \quad ) ) = ( \quad | \quad , \quad ) ( \quad | \text{làm}( \quad ) ) \quad (4.12)$$

Điều này xuất phát từ giả định tính mô đun (Giả định 4.1). Nếu là tất cả cha mẹ của (trừ ), thì rõ ràng giả định mô đun ngay lập tức ngụ ý  $( \quad | \text{do}( \quad ), \quad ) = ( \quad | \quad , \quad )$ . Nếu không phải là cha mẹ của nhưng vẫn chặn tất cả các đường dẫn cửa sau theo cách khác thì đẳng thức này vẫn đúng nhưng yêu cầu sử dụng kiến thức đồ họa mà chúng ta đã xây dựng trong Chương 3.

Trong biểu đồ được thao tác (đối với  $( \quad | \text{do}( \quad ), \quad )$ ), tất cả liên kết - chạy dọc theo (các) đường dẫn có hướng từ đến , vì không thể có bất kỳ đường dẫn cửa sau nào vì không có các cạnh đến . Tương tự, trong biểu đồ thông thường (đối với  $( \quad | \quad , \quad )$ ), tất cả liên kết - chạy dọc theo (các) đường dẫn có hướng từ đến . Điều này là do, mặc dù tồn tại các đường dẫn cửa sau, nhưng liên kết đi dọc theo chúng bị chặn bởi , khiến liên kết chỉ chạy dọc theo các đường dẫn có hướng. Trong cả hai trường hợp, sự liên kết diễn ra dọc theo các đường dẫn có hướng giống hệt nhau, tương ứng với các phân bố có điều kiện giống hệt nhau (theo giả định tính mô đun).

Mặc dù chúng ta đã chứng minh phương trình 4.12 nhưng vẫn có một do trong biểu thức:  $( \quad | \text{do}( \quad ))$ . Tuy nhiên,  $( \quad | \text{do}( \quad )) = ( \quad )$ . Để thấy điều này, hãy xem xét có thể có ảnh hưởng như thế nào đến trong biểu đồ bị thao túng. Nó không thể đi qua bất kỳ đường dẫn nào có cạnh vào vì không có bất kỳ cạnh nào trong đồ thị được thao tác. Nó không thể đi qua bất kỳ đường đi nào có cạnh đi ra ngoài bởi vì đường đi như vậy sẽ phải có một máy va chạm, không bị điều hòa, trên đường đi. Chúng ta biết bất kỳ máy va chạm nào như vậy đều không bị điều kiện hóa bởi vì chúng ta đã giả sử rằng không chứa hậu duệ của (phần thứ hai của tiêu chí cửa sau).<sup>6</sup> Vì vậy, chúng ta có thể viết bước cuối cùng:

$$( \quad | \quad , \quad ) ( \quad | \text{làm}( \quad ) ) = ( \quad | \quad , \quad ) ( \quad ) \quad (4.13)$$

Điều này được gọi là điều chỉnh cửa sau.

**Định lý 4.2 (Điều chỉnh cửa sau)** Với giả định tính mô đun (Giả định 4.1), thỏa mãn tiêu chí cửa sau (Định nghĩa 4.1), và

<sup>6</sup> Chúng ta sẽ quay lại vấn đề sai sót nếu đặt điều kiện cho hậu duệ của trong Phần 4.5.3, sau khi đề cập đến một số khái niệm quan trọng mà chúng ta cần trước khi có thể giải thích đầy đủ điều đó.

dương (Giả định 2.3), chúng ta có thể xác định tác động nhân quả của lên :

$$( | \text{làm}( )) = ( | , ) ( )$$

Đây là bản tóm tắt ngắn gọn về bằng chứng (Pt 4.11 đến 4.13) mà không có bất kỳ lời giải thích/biện minh nào:

Bằng chứng.

$$( | \text{làm}( )) = ( | \text{làm}( ), ) ( | \text{làm}( )) \quad (4.14)$$

$$= ( | , ) ( | \text{làm}( )) \quad (4.15)$$

$$= ( | , ) ( ) \quad (4.16)$$

□

Liên quan đến sự phân tách d Chúng ta có thể sử dụng điều chỉnh cửa sau nếu d- tách khỏi trong đồ thị được thao tác. Nhớ lại Phần 3.8 mà chúng ta đã đề cập rằng chúng ta có thể cô lập mối liên hệ nhân quả nếu được phân tách d khỏi trong đồ thị thao tác. "Sự cô lập của mối liên hệ nhân quả" là sự đồng nhất. Chúng ta cũng có thể cô lập mối liên hệ nhân quả nếu được tách d khỏi trong đồ thị điều khiển, có điều kiện trên . Đây là nội dung phần đầu tiên của tiêu chí cửa sau và nội dung chúng tôi đã hệ thống hóa trong quá trình điều chỉnh cửa sau.

#### 4.4.1 Mối liên hệ với các kết quả tiềm năng

Hmm, công thức điều chỉnh cửa sau (Định lý 4.2) trông khá giống với công thức điều chỉnh (Định lý 2.1) mà chúng ta đã thấy ở chương kết quả tiềm năng:

$$[ (1) (0) ] = [ [ | ] = 1, ] [ | ] = 0, ] \quad (4.17)$$

Chúng ta có thể rút ra điều này từ việc điều chỉnh cửa sau tổng quát hơn trong một vài bước. Đầu tiên, chúng ta lấy kỳ vọng về :

$$[ | \text{làm}( ) ] = [ | , ] ( ) \quad (4.18)$$

Sau đó, chúng ta nhận thấy rằng tổng trên và ( ) là một kỳ vọng (đối với rời rạc, nhưng chỉ thay thế bằng tích phân nếu không):

$$[ | \text{làm}( ) ] = [ | , ] \quad (4.19)$$

Và cuối cùng, chúng ta xem xét sự khác biệt giữa  $= 1$  và  $= 0$ :

$$[ | \text{do}( = 1) ] - [ | \text{do}( = 0) ] = [ [ | ] = 1, ] - [ | ] = 0, ] \quad (4.20)$$

Vì ký hiệu do  $[ | \text{do}( ) ]$  chỉ là một ký hiệu khác cho các kết quả tiềm năng  $[ ( ) ]$ , chúng ta đã hoàn thành! Nếu bạn còn nhớ, một trong những giả định chính mà chúng ta cần để có được Phương trình 4.17 (Định lý 2.1) là có điều kiện

khả năng trao đổi (Giả định 2.2), mà chúng tôi nhắc lại dưới đây:

$$( - (1), \quad (0)) \quad | \quad (4.21)$$

Tuy nhiên, chúng ta không có cách nào để biết cách chọn hoặc biết rằng thực sự mang lại cho chúng ta khả năng trao đổi có điều kiện. Chà, bằng cách sử dụng các mô hình nhân quả đồ họa, chúng ta biết cách chọn một hợp lệ: chúng ta chỉ cần chọn sao cho nó thỏa mãn tiêu chí của sau. Sau đó, theo các giả định được mã hóa trong biểu đồ nhân quả, khả năng trao đổi có điều kiện có thể được chứng minh là đúng; tác động nhân quả có thể được xác định rõ ràng.

#### 4.5 Mô hình nhân quả cấu trúc (SCM)

Các mô hình nhân quả đồ họa như mạng Bayesian nhân quả cho chúng ta những cách mạnh mẽ để mã hóa các giả định thống kê và nhân quả, nhưng chúng ta vẫn chưa giải thích chính xác can thiệp là gì hoặc cơ chế nhân quả là gì. Việc chuyển từ mạng Bayesian nhân quả sang các mô hình nhân quả có cấu trúc đầy đủ sẽ mang lại cho chúng ta sự rõ ràng bổ sung này cùng với khả năng tính toán các phần thực tế.

##### 4.5.1 Phương trình cấu trúc

Như Judea Pearl thường nói, dấu bằng trong toán học không truyền tải bất kỳ thông tin nhân quả nào. Nói  $=$  cũng giống như nói  $\equiv$ .  
Bình đẳng là đối xứng. Tuy nhiên, để nói về nhân quả, chúng ta phải có cái gì đó bất đối xứng. Chúng ta cần viết được rằng  $=$  là nguyên nhân của  $\equiv$ , nghĩa là việc thay đổi  $=$  dẫn đến thay đổi trong  $\equiv$ , nhưng thay đổi  $\equiv$  không dẫn đến thay đổi trong  $=$ . Đây là những gì chúng ta nhận được khi viết phương trình cấu trúc sau :

$$:= ( \quad ), \quad (4.22)$$

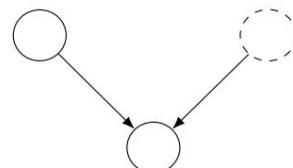
trong đó  $=$  là hàm nào đó ánh xạ  $\rightarrow$  tới  $\equiv$ . Mặc dù biểu tượng “=” thông thường không cung cấp cho chúng ta thông tin nhân quả, nhưng biểu tượng “:<=” mới này lại cung cấp cho chúng ta thông tin đó. Đây là điểm khác biệt lớn mà chúng ta thấy khi chuyển từ mô hình thống kê sang mô hình nhân quả. Bây giờ, chúng ta có sự bất đối xứng mà chúng ta cần để mô tả các mối quan hệ nhân quả. Tuy nhiên, ánh xạ giữa  $=$  và  $\equiv$  là xác định. Lý tưởng nhất là chúng tôi muốn cho phép nó mang tính xác suất, điều này cho phép có chõ cho một số nguyên nhân chưa biết của  $\equiv$  ánh hưởng đến ánh xạ này. Sau đó, chúng ta có thể viết như sau:

$$:= ( \quad , \quad ), \quad (4.23)$$

trong đó  $=$  là một biến ngẫu nhiên không được quan sát. Chúng tôi mô tả điều này trong Hình 4.7, trong đó  $=$  được vẽ bên trong một nút đứt nét để chỉ ra rằng nó không được quan sát.

Không quan sát được tương tự như tính ngẫu nhiên mà chúng ta thấy ở các đơn vị lấy mẫu (cá nhân); nó biểu thị tất cả các điều kiện nền (ôn ào) có liên quan xác định  $\equiv$ . Cụ thể hơn, có những điểm tương đồng với mọi phần của kết quả tiềm năng  $( \quad )$ :  $=$  là dạng tương tự của  $\equiv$ ,  $\equiv =$  là dạng tương tự của  $=$ , và  $\equiv$  là dạng tương tự của  $\equiv$ .

Dạng hàm của  $=$  không cần phải được chỉ định và khi không được chỉ định, chúng ta đang ở chế độ không tham số vì chúng ta không đưa ra bất kỳ giả định nào về dạng tham số. Mặc dù việc lập bản đồ



Hình 4.7: Đồ thị phương trình cấu trúc đơn giản. Nút nét đứt  $=$  có nghĩa là không được quan sát.

mang tính quyết định, vì nó lấy biến ngẫu nhiên (biến "nhiều" hoặc "điều kiện nền") làm đầu vào, nên nó có thể biểu thị bất kỳ ánh xạ ngẫu nhiên nào, do đó các phương trình cấu trúc tổng quát hóa các hệ số xác suất ( $|pa|$ ) mà chúng ta đã tìm ra sử dụng xuyên suốt chương này. Do đó, tất cả các kết quả mà chúng ta đã thấy như hệ số rút gọn và điều chỉnh cửa sau vẫn giữ nguyên khi chúng ta đưa ra các phương trình cấu trúc.

Xem lại Nguyên nhân và Cơ chế Nhân quả Bây giờ chúng ta đã đi đến những định nghĩa chính xác hơn về nguyên nhân là gì (Định nghĩa 3.2) và cơ chế nhân quả là gì (được giới thiệu trong Phần 4.2). Cơ chế nhân quả tạo ra một biến là phương trình cấu trúc tương ứng với biến đó. Ví dụ, cơ chế nhân quả của là Phương trình 4.23.

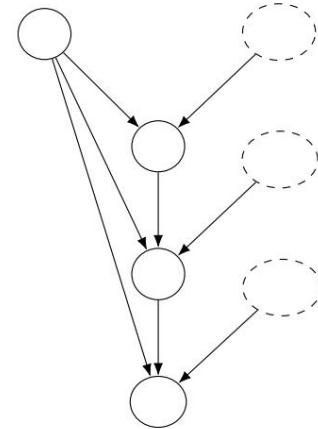
Tương tự, là nguyên nhân trực tiếp của nếu xuất hiện ở về phải của phương trình cấu trúc của . Ta nói rằng là nguyên nhân của nếu là nguyên nhân trực tiếp của bất kỳ nguyên nhân nào của<sup>7</sup>.

Chúng tôi chỉ trình bày một phương trình cấu trúc duy nhất trong Công thức 4.23, nhưng có thể có một tập hợp lớn các phương trình cấu trúc trong một mô hình duy nhất mà chúng tôi thường gọi là . Ví dụ, chúng ta viết phương trình cấu trúc cho Hình 4.8 dưới đây:

$$\begin{aligned} &:= (\ , \ ) \\ &:= (\ , \ , \ ) \\ &:= (\ , \ , \ ) \end{aligned} \quad (4.24)$$

Trong đồ thị nhân quả, các biến nhiều thường ẩn chứ không được vẽ rõ ràng. Các biến mà chúng ta viết phương trình cấu trúc được gọi là biến nội sinh. Đây là các biến có cơ chế nhân quả mà chúng tôi đang lập mô hình - các biến có cha mẹ trong biểu đồ nhân quả. Ngược lại, biến ngoại sinh là biến không có cha mẹ trong biểu đồ nhân quả; những biến này nằm ngoài mô hình nhân quả của chúng ta theo nghĩa là chúng ta chọn không lập mô hình nguyên nhân của chúng. Ví dụ, trong mô hình nhân quả được mô tả ở Hình 4.8 và Phương trình 4.24, các biến nội sinh là { , , }. Và các biến ngoại sinh là { , , , }.

<sup>7</sup> Hãy tin tôi; đệ quy kết thúc. Căn cứ trường hợp đã được chỉ định.



Hình 4.8: Đồ thị các phương trình cấu trúc trong phương trình 4.24.

**Định nghĩa 4.2 (Mô hình nhân quả cấu trúc (SCM))** Mô hình nhân quả cấu trúc là một bộ gồm các bộ sau:

1. Một tập hợp các biến nội sinh
2. Một tập hợp các biến ngoại sinh
3. Một tập hợp các hàm , một hàm tạo ra mỗi biến nội sinh dưới dạng một hàm của các biến khác

Ví dụ, , tập hợp ba phương trình trên trong Phương trình 4.24 tạo thành một SCM với đồ thị nhân quả tương ứng trong Hình 4.8. Mỗi SCM bao hàm một đồ thị nhân quả liên quan: đối với mỗi phương trình cấu trúc, hãy vẽ một cạnh từ mọi biến ở về phải đến biến ở về trái.

Nếu đồ thị nhân quả không chứa chu trình (là DAG) và các biến nhiều độc lập thì mô hình nhân quả là Markovian; phân bố là Markov đối với đồ thị nhân quả. Nếu đồ thị nhân quả không chứa chu trình nhưng các số hạng nhiều phụ thuộc thì mô hình là bán Markovian. Ví dụ, nếu có sự nhiễu loạn không quan sát được, mô hình

là bán Markovian. Cuối cùng, đồ thị của các mô hình phi Markovian chứa các chu trình. Phần lớn chúng ta sẽ xem xét các mô hình Markovian và bán Markovian trong cuốn sách này.

#### 4.5.2 Can thiệp

Các can thiệp vào SCM rất đơn giản. Sự can thiệp do( $\theta = \theta'$ ) chỉ tương ứng với việc thay thế phương trình cấu trúc của bằng  $\theta := \theta'$ . Ví dụ, xét mô hình nhân quả sau với đồ thị nhân quả tương ứng trên Hình 4.9:

$$\begin{aligned} &:= (\theta, \theta') \\ &:= (\theta, \theta', \theta'') \end{aligned} \quad (4.25)$$

Sau đó, nếu chúng ta can thiệp vào để đặt nó thành  $\theta'$ , chúng ta sẽ nhận được SCM can thiệp bên dưới và đồ thị được thao tác tương ứng trong Hình 4.10.

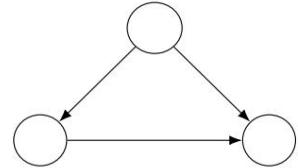
$$\begin{aligned} &:= \\ &:= (\theta, \theta', \theta'') \end{aligned} \quad (4.26)$$

Việc do( $\theta = \theta'$ ) chỉ thay đổi phương trình của và không có biến nào khác là hệ quả của giả định môđun hóa; các cơ chế nhân quả này (phương trình cấu trúc) có tính môđun. Giả định 4.1 nêu giả định về tính môđun trong bối cảnh mạng Bayes nhân quả, nhưng chúng ta cần một cách diễn giải hơi khác về giả định này đối với SCM.

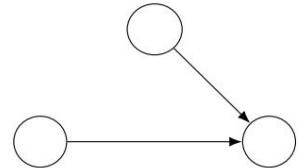
**Giả định 4.2 (Giả định môđun cho SCM)** Xét SCM và SCM can thiệp mà chúng ta có được bằng cách thực hiện phép can thiệp do( $\theta = \theta'$ ). Giả định tính môđun cho biết rằng và có chung tất cả các phương trình cấu trúc của chúng ngoại trừ phương trình cấu trúc của , là  $\theta := \theta'$  trong

Nói cách khác, sự can thiệp do( $\theta = \theta'$ ) được định vị thành . Không có phương trình cấu trúc nào khác thay đổi vì chúng có tính môđun; các cơ chế nhân quả là độc lập. Giả định tính môđun cho SCM là điều mang lại cho chúng ta cái mà Pearl gọi là Quy luật phản ứng thực tế, mà chúng tôi đã đề cập ngắn gọn ở cuối Phần 4.2, sau khi chúng tôi xác định giả định tính môđun cho các mạng Bayes nhân quả. Nhưng trước khi đạt được điều đó, trước tiên chúng ta phải giới thiệu thêm một chút ký hiệu.

Trong tài liệu suy luận nhân quả, có nhiều cách khác nhau để viết kết quả tiềm năng ở cấp độ đơn vị. Trong Chương 2, chúng ta đã sử dụng  $(\cdot)$ . Tuy nhiên, trong bài của ông còn có những cách hoặc thậm chí là  $(\cdot)$ . Ví dụ, khác như bài báo về kết quả tiềm năng nổi bật, Holland [5] sử dụng ký hiệu Trong ký hiệu này, là dạng tương tự của , giống như chúng tôi đã đề cập đến trường hợp của trong Công thức 4.23 và đoạn tiếp theo nó. Đây cũng là ký hiệu mà Pearl sử dụng cho SCM [xem, ví dụ, 17, Định nghĩa 4]. Vì vậy, ( $\cdot$ ) biểu thị kết quả mà đơn vị sẽ quan sát nếu họ thực hiện biện pháp xử lý , với điều kiện SCM là . Tương tự, chúng tôi định nghĩa ( $\cdot$ ) là kết quả mà đơn vị sẽ quan sát nếu họ thực hiện xử lý , với điều kiện SCM là (hãy nhớ rằng là SCM giống như nhưng có phương trình cấu trúc cho được đổi thành  $\theta := \theta'$ ). Nay giờ, chúng tôi đã sẵn sàng



Hình 4.9: Đồ thị nhân quả cơ bản



Hình 4.10: Nhân quả cơ bản với các cạnh mới của bị loại bỏ do có sự can thiệp do( $\theta = \theta'$ ).

[5]: Holland (1986), 'Thống kê và nhân quả Suy luận'

[17]: Pearl (2009), 'Suy luận nhân quả trong thống kê: Tổng quan'

trình bày một trong hai nguyên tắc then chốt của Pearl mà từ đó tất cả các nguyên tắc nhân quả khác kết quả như sau:

**Định nghĩa 4.3 (Quy luật phản thực tế (và can thiệp))**

$$( ) = ( ) \quad (4.27)$$

Điều này được gọi là "Quy luật phản thực tế" vì nó cung cấp cho chúng ta thông tin về phản thực tế. Đưa ra một SCM có đủ thông tin chi tiết về nó được chỉ định, chúng ta thực sự có thể tính toán các phản thực. Đây là một vấn đề lớn bởi vì đây chính xác là vấn đề cơ bản của suy luận nhân quả

(Phần 2.2) đã nói với chúng tôi rằng chúng tôi không thể làm được. Chúng tôi sẽ không nói thêm về cách thực hiện việc này cho đến khi chúng ta chuyển sang chương dành riêng cho phản thực tế: Chương 14.

#### 4.5.3 Xu hướng và chạm và tại sao không có điều kiện hậu duệ của điều trị

Trong việc xác định tiêu chí cửa sau (Định nghĩa 4.1) cho cửa sau điều chỉnh (Định lý 4.2), chúng ta không chỉ xác định rằng việc điều chỉnh đặt chặn tất cả các đường dẫn cửa sau, nhưng chúng tôi cũng chỉ định rằng không chứa bất kỳ hậu duệ nào của . Tại sao? Có hai loại sự vật điều đó có thể sai nếu chúng ta đặt điều kiện vào hậu duệ của :

1. Chúng tôi chặn luồng nhân quả từ đến .
2. Chúng ta tạo ra mối liên hệ phi nhân quả giữa và .

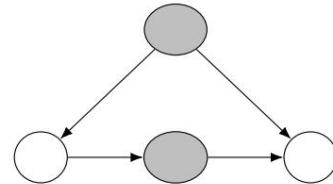
Như chúng ta sẽ thấy, lý do khá trực quan là tại sao chúng ta muốn tránh loại đầu tiên. Loại thứ hai phức tạp hơn chút và chúng tôi sẽ chia nó thành hai phần khác nhau, mỗi phần có đoạn văn riêng. Phần phức tạp hơn này thực ra đó là lý do tại sao chúng tôi trì hoãn việc giải thích này sau khi giới thiệu SCM, thay vì quay lại khi chúng tôi giới thiệu tiêu chí điều chỉnh cửa sau trong Phần 4.4.

Nếu chúng ta điều kiện hóa một nút nằm trên đường có hướng từ đến , thì chúng ta chặn dòng nhân quả dọc theo con đường nhân quả đó. Chúng ta sẽ đề cập đến một nút trên đường dẫn có hướng từ đến với vai trò là trung gian, vì nó làm trung gian tác động của trên . Ví dụ, trong Hình 4.11, tất cả dòng nhân quả bị chặn bởi . Điều này có nghĩa là chúng tôi sẽ đe lường mối liên hệ bằng không giữa và (vì chặn tất cả các đường dẫn cửa sau). Trong Hình 4.12, chỉ một phần của dòng nhân quả bị chặn bởi . Đó là vì nhân quả vẫn có thể chảy dọc theo cạnh . Trong trường hợp này, chúng ta sẽ nhận được ước tính khía cạnh của tác động nhân quả, nhưng nó vẫn sẽ bị sai lệch, do dòng nhân quả mà khói.

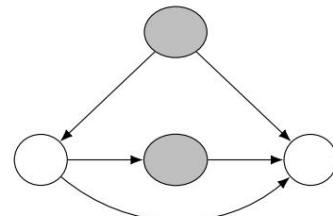
Nếu chúng ta đặt điều kiện vào hậu duệ của không phải là người hòa giải, nó có thể bỏ chặn một đường đi từ đến đã bị chặn bởi một máy va chạm. Ví dụ, đây là trường hợp có điều kiện trên trong Hình 4.13. Điều này gây ra sự vô nhân quả mối liên hệ giữa và , làm sai lệch ước tính nguyên nhân tác dụng. Hãy xem xét loại đường dẫn chung sau đây, trong đó , . Điều hòa trên , hoặc bất kỳ hậu duệ nào của trong đường đi như thế này, sẽ gây ra sai lệch máy va chạm. Cái đó là, ước tính tác động nhân quả sẽ bị sai lệch bởi mối liên hệ phi nhân quả mà chúng ta gây ra khi đặt điều kiện vào hoặc bất kỳ hậu duệ nào của nó (xem Mục 3.6).

"Bài tập đọc tích cực: Bạn có thể nhớ lại nguyên tắc/giả định quan trọng nào khác ?

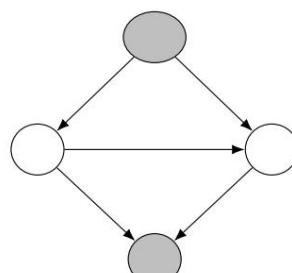
Bài tập đọc tích cực: Hãy lấy những gì bạn bấy giờ đã biết về phương trình cấu trúc, và liên hệ nó với các phần khác của chương này. Ví dụ, làm thế nào để can thiệp vào phương trình cấu trúc liên quan đến giả định mô đun? Giả định tính mô đun cho SCM (Giả định 4.2) liên quan như thế nào đến giả định tính mô đun trong mạng Bayesian nhân quả (Giả định 4.1)? Giả định về tính môđun này cho SCM có còn mang lại cho chúng ta cửa sau không? điều chỉnh?



Hình 4.11: Đồ thị nhân quả trong đó tất cả nhân quả bị chặn bởi điều kiện hóa trên .



Hình 4.12: Biểu đồ nhân quả trong đó một phần của nhân quả bị chặn bởi điều hòa trên .



Hình 4.13: Đồ thị nhân quả trong đó điều kiện hòa trên máy va chạm gây ra sai lệch.

Còn điều hòa trên trong Hình 4.14 thì sao? Điều đó có gây ra sự thiên vị không? Hãy nhớ lại rằng đồ thị thường được vẽ mà không vẽ rõ ràng các biến nhiễu. Nếu chúng ta phóng to một phần của biểu đồ, làm rõ biến nhiễu của , chúng ta sẽ có Hình 4.15. Bây giờ, chúng ta thấy rằng tạo thành một sự vô đạo đức. Do đó, việc điều hòa gây ra sự liên kết giữa và . Sự liên kết phi nhân quả gây ra này là một dạng khác của sai lệch máy va chạm. Bạn có thể thấy điều này không thỏa mãn vì không phải là một trong những bậc cha mẹ vô đạo đức ở đây; đúng hơn và là những người có lối sống vô đạo đức. Vậy tại sao điều này lại làm thay đổi mối liên hệ giữa và ? Một cách để hiểu trực giác về điều này là hiện nay có sự liên kết cảm ứng giữa và thông qua cạnh , đây cũng là một cạnh mà mối liên hệ nhân quả đang diễn ra. Bạn có thể nghĩ hai loại liên kết này bị vướng vào nhau dọc theo cạnh , khiến cho mối liên hệ được quan sát giữa và không thuận túy là quan hệ nhân quả. Xem Pearl [18, Phần 11.3.1 và 11.3.3] để biết thêm thông tin về chủ đề này.

Lưu ý rằng chúng ta thực sự có thể điều kiện hóa một số hậu duệ của mà không tạo ra các mối liên hệ phi nhân quả giữa và . Ví dụ: quy định các hậu duệ của không nằm trên bất kỳ con đường nhân quả nào dẫn đến sẽ không gây ra sự thiên vị. Tuy nhiên, như bạn có thể thấy trong đoạn văn trên, việc này có thể hơi phức tạp, vì vậy an toàn nhất là không đặt điều kiện vào bất kỳ hậu duệ nào của , như tiêu chí cửa sau quy định. Ngay cả bên ngoài các mô hình nhân quả đồ họa (ví dụ như trong tài liệu về kết quả tiềm năng), quy tắc này vẫn thường được áp dụng; nó thường được mô tả là không điều chỉnh bất kỳ biến só sau điều trị nào.

M-Bias Thật không may, ngay cả khi chúng ta chỉ dựa vào các đồng biến tiềm xử lý, chúng ta vẫn có thể gây ra sai lệch máy va chạm. Hãy xem điều gì sẽ xảy ra nếu chúng ta điều kiện hóa máy va chạm 2 trên Hình 4.16. Việc làm này sẽ mở ra một đường dẫn cửa sau, dọc theo đó sự liên kết phi nhân quả có thể diễn ra. Điều này được gọi là sai lệch M do hình chữ M mà mối liên hệ phi nhân quả này diễn ra khi biểu đồ được vẽ với những đứa trẻ đứng dưới cha mẹ chúng.

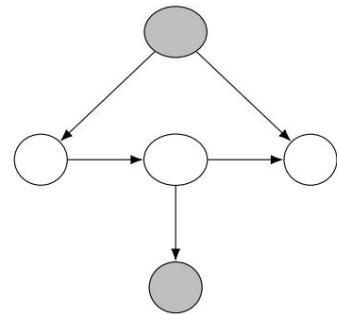
Để biết nhiều ví dụ về độ lệch của máy va chạm, xem Elwert và Winship [19].

## 4.6 Các ứng dụng ví dụ về Điều chỉnh cửa sau

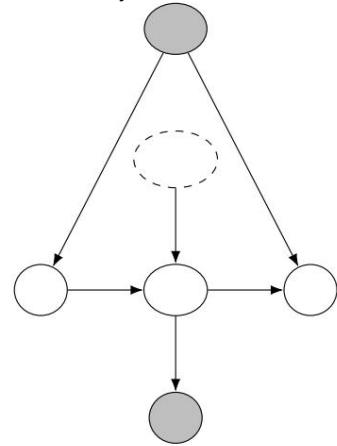
### 4.6.1 Mối liên hệ và mối quan hệ nhân quả trong một ví dụ về đồ chơi

Trong phần này, chúng tôi thừa nhận một quy trình tạo đồ chơi và rút ra độ lệch của đại lượng liên kết [ | ]. Chúng ta so sánh điều này với đại lượng nhân quả [ | do( )], mang đến cho chúng ta chính xác những gì chúng ta muốn. Lưu ý rằng cả hai đại lượng này thực tế đều là hàm của . Nếu cách xử lý là nhị phân thì chúng ta sẽ chỉ xem xét sự khác biệt giữa các đại lượng với = 1 và với = 0. Tuy nhiên, vì quá trình tạo của chúng ta sẽ là [ | do( )] tóm tắt tính và thực sự cung cấp cho chúng tôi tất cả thông tin về hiệu quả điều trị, bắt kè việc điều trị là liên tục, nhị phân hay đa giá trị. Chúng tôi sẽ sử dụng dữ liệu vô hạn để chúng tôi có thể làm việc với mong đợi. Điều này có nghĩa là phần này không liên quan gì đến việc ước tính; để ước tính, xem phần tiếp theo

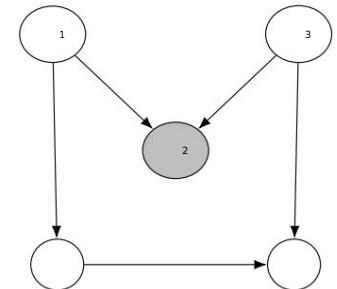
Quá trình sinh sản mà chúng tôi xem xét có biểu đồ nhân quả trong Hình 4.17



Hình 4.14: Biểu đồ nhân quả trong đó con của người hòa giải bị điều hòa.



Hình 4.15: Biểu đồ nhân quả phóng to trong đó con của người hòa giải được điều chỉnh.



Hình 4.16: Đồ thị nhân quả mô tả M-Thiến kiền.

và các phương trình cấu trúc sau:

$$:= 1 \quad (4.28)$$

$$:= + 2 . \quad (4.29)$$

Lưu ý rằng trong phương trình cấu trúc của , là hệ số đứng trước . Điều này có nghĩa là tác động nhân quả của lên là . Hãy ghi nhớ điều này khi chúng tôi trải qua những tính toán này.

Từ đồ thị nhân quả ở tập điều chỉnh Hình , chúng ta có thể thấy rằng là đủ 4.17 . Vì vậy, [ | làm( )] = [ | , ]. Hãy tính toán giá trị của đại lượng này trong ví dụ của chúng tôi.

$$[ | , ] = [ + 2 | = , ] \quad (4.30)$$

$$= + 2 \quad (4.31)$$

$$= + 2 [ ] \quad (4.32)$$

Điều quan trọng là chúng tôi đã sử dụng đẳng thức mà phương trình cấu trúc của (Phương trình 4.29) cho ta phương trình 4.30. Bây giờ chúng ta chỉ cần lấy đạo hàm để tìm hệ quả nhân quả:

$$\frac{[ | , ]}{=} . \quad (4.33)$$

Chúng tôi đã có chính xác những gì chúng tôi đang tìm kiếm. Bây giờ chúng ta hãy chuyển sang đại lượng kết hợp:

$$[ | = ] = [ + 2 | = ] \quad (4.34)$$

$$= + 2 [ | = ] \quad (4.35)$$

$$= + \frac{2}{1} \quad (4.36)$$

Trong phương trình 4.36, chúng ta đã sử dụng đẳng thức mà phương trình cấu trúc với (Phương trình 4.28) cho ta. Nếu sau đó chúng ta lấy đạo hàm, chúng ta thấy rằng có sự thiên vị gây nhiễu:

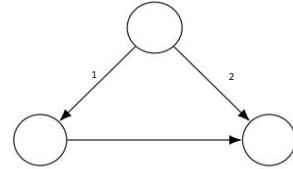
$$\frac{[ | ]}{1} = + \frac{2}{1} . \quad (4.37)$$

Tóm lại, [ | , ] đã mang lại cho chúng tôi tác động nhân quả mà chúng tôi đang tìm kiếm (Phương trình 4.33), trong khi đại lượng liên kết [ | ] đã không (Phương trình 4.37). Bây giờ, chúng ta hãy xem qua một ví dụ cũng đề cập đến ước tính tài khoản.

#### 4.6.2 Một ví dụ hoàn chỉnh với ước tính

Hãy nhớ lại rằng chúng tôi đã ước tính một giá trị cụ thể về tác động nhân quả của natri lượng huyết áp ở Phần 2.5. Ở đó, chúng tôi đã sử dụng tiềm năng khuôn khổ kết quả. Ở đây, chúng ta sẽ làm điều tương tự, nhưng sử dụng nguyên nhân đồ thị. Điều tiết lộ là lỗi 19% mà chúng ta thấy ở Phần 2.5 là do sự điều hòa trên máy va chạm.

Đầu tiên, chúng ta cần viết ra các giả định nhân quả của mình dưới dạng một quan hệ nhân quả đồ thị. Hãy nhớ rằng trong Luque-Fernandez et al. Ví dụ của [8] từ dịch tễ học, việc điều trị là bổ sung natri và kết quả là



Hình 4.17: Đồ thị nhân quả cho ví dụ về đồ chơi

[8]: Luque-Fernandez và cộng sự. (2018), 'Ghi chú giáo dục: Hiệu ứng máy va chạm nghịch lý trong phân tích dữ liệu dịch tễ học về bệnh không lây nhiễm: một phương pháp có thể tái sản xuất minh họa và ứng dụng web'

huyết áp. Các dòng biến là tuổi và lượng protein trong nước tiểu (protein niệu). Tuổi tác là nguyên nhân phổ biến gây ra cả huyết áp và khả năng tự điều chỉnh nồng độ natri của cơ thể. Ngược lại, lượng protein trong nước tiểu cao là do huyết áp cao và lượng natri cao. Điều này có nghĩa là protein niệu là một yếu tố va chạm. Chúng tôi mô tả đồ thị nhân quả trong Hình 4.18.

Bởi vì là một máy va chạm nên việc điều hòa nó gây ra sai lệch. Bởi vì và được nhóm lại với nhau thành "đồng biến" trong Phần 2.5, nên chúng tôi đã điều chỉnh tất cả chúng. Đây là lý do tại sao chúng tôi thấy rằng ước tính của chúng tôi giảm 19% so với hiệu ứng nhân quả thực sự 1,05. Bây giờ chúng ta đã làm rõ mối quan hệ nhân quả bằng biểu đồ nhân quả, tiêu chí cửa sau (Định nghĩa 4.1) yêu cầu chúng ta chỉ điều chỉnh cho và không điều chỉnh cho . Chính xác hơn, chúng tôi đã thực hiện điều chỉnh sau trong Phần 2.5:

$$\text{[ } | \text{ } , \text{ } . \text{ } ] \quad (4.38)$$

Và bây giờ, chúng ta sẽ sử dụng điều chỉnh cửa sau (Định lý 4.2) để thay đổi ước tính thống kê của chúng ta thành như sau:

$$\text{[ } | \text{ } , \text{ } . \text{ } ] \quad (4.39)$$

Chúng tôi vừa loại bỏ máy va chạm khỏi các biến mà chúng tôi điều chỉnh. Để ước tính, giống như chúng tôi đã làm trong Phần 2.5, chúng tôi sử dụng công cụ ước tính có sự hỗ trợ của mô hình. Chúng tôi thay thế kỳ vọng bên ngoài trên bằng giá trị trung bình thực nghiệm trên và thay thế kỳ vọng có điều kiện [ | , ] với mô hình học máy (trong trường hợp này là hồi quy tuyến tính).

Cũng giống như việc viết ra biểu đồ đã khiến chúng ta đơn giản không đặt điều kiện vào trong Công thức 4.39, mà ước tính cũng hầu như không thay đổi. Chúng ta chỉ cần thay đổi một dòng mã trong chương trình trước đó (Liệt kê 2.1). Chúng ta hiển thị toàn bộ chương trình với dòng code cố định bên dưới:

```

1 nhập numpy dưới dạng np
2 nhập giao trúc dưới dạng pd
3 từ sklearn.Tuyến_model nhập LinearRegression 4

5 Xt = df[['sodium', 'age']] 6 y =
df['blood_Pressure'] 7 model =
LinearRegression() 8 model.fit(Xt,
y) 9

10 Xt1 = pd.DataFrame.copy(Xt)
11 Xt1['natri'] = 1 12 Xt0
= pd.DataFrame.copy(Xt)
13 Xt0['sodium'] = 0 14
eat_est = np.mean(model.predict(Xt1) - model.predict(Xt0)) 15 print('ATE ước
tính:', eat_est)

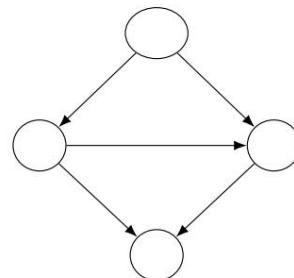
```

Cụ thể, chúng tôi đã thay đổi dòng 5 từ

```

5 Xt = df[['natri', 'tuổi', 'protein niệu']]
trong Liệt kê 2.1 tới
5 Xt = df[['natri', 'tuổi']]
trong Liệt kê 4.1. Khi chạy mã sửa đổi này, chúng tôi nhận được ước tính ATE là 1,0502, tương ứng với sai số 0,02% (giá trị thực là 1,05) khi sử dụng

```



Hình 4.18: Biểu đồ nhân quả cho ví dụ về huyết áp. là lượng natri tiêu thụ. là huyết áp. là tuổi. Và điều quan trọng là lượng protein bài tiết trong nước tiểu là một máy va chạm.

Liệt kê 4.1: Mã Python để ước tính ATE mà không cần điều chỉnh cho máy va chạm

Mã đầy đủ, hoàn chỉnh với mô phỏng, có sẵn tại [https://github.com/brydenneal/causal-book-code/blob/master/sodium\\_example.py](https://github.com/brydenneal/causal-book-code/blob/master/sodium_example.py).

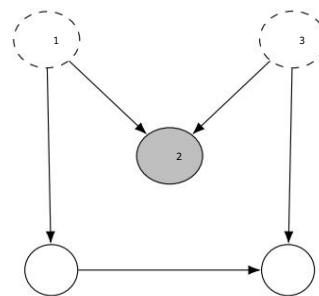
một mẫu khá lớn.<sup>9</sup>

Sự tiến triển của việc giảm độ lệch Khi xem xét mối liên hệ tổng thể giữa và bằng cách đơn giản hồi quy trên, chúng tôi đã có được ước tính giảm đáng kinh ngạc 407% so với hiệu ứng nhân quả thực sự, phần lớn là do sai lệch gây nhiễu (xem Phần 2.5). Khi chúng tôi điều chỉnh tất cả các biến số trong Phần 2.5, chúng tôi đã giảm sai số phần trăm xuống còn 19%. Trong phần này, chúng ta thấy lỗi còn lại này là do độ lệch của máy và chạm. Khi chúng tôi loại bỏ độ lệch của máy và chạm, bằng cách không điều chỉnh máy và chạm, lỗi đã trở thành không tồn tại.

Kết quả tiềm năng và M-Bias Để công bằng với văn hóa chung xung quanh khung kết quả tiềm năng, thông thường chỉ dựa vào các đồng biến tiềm xử lý. Điều này sẽ ngăn cản người thực hành tuân thủ quy tắc này điều hòa máy và chạm trong Hình 4.18.

Tuy nhiên, không có lý do gì mà không thể có các máy và chạm tiềm xử lý gây ra sai lệch M (Phần 4.5.3). Trong Hình 4.19, chúng tôi mô tả một ví dụ về độ lệch M được tạo ra bằng cách điều hòa trên 2. Chúng ta có thể khắc phục điều này bằng cách điều hòa thêm 1 và/hoặc 3, nhưng trong ví dụ này, chúng không được quan sát (được biểu thị bằng các đường đứt nét). Điều này có nghĩa là cách duy nhất để tránh M-bias trong Hình 4.19 là không đặt điều kiện vào các hiệp phương sai 2.

<sup>9</sup> **Bài tập đọc tích cực:** Cho rằng được tạo ra dưới dạng hàm tuyến tính của và , liệu chúng ta có thể sử dụng hệ số ở phía trước trong hồi quy tuyến tính để ước tính tác động nhân quả không?



Hình 4.19: Đồ thị nhân quả mô tả M- Bias chỉ có thể tránh được bằng cách không điều hòa trên máy và chạm 2. Điều này là do thực tế là các nút nét đứt 1 và 3 không được quan sát.

## 4.7 Xem lại các giả định

Nhóm giả định chính đầu tiên được mã hóa bằng biểu đồ nhân quả mà chúng tôi viết ra. Ý nghĩa chính xác của biểu đồ nhân quả này được xác định bởi hai giả định chính, mỗi giả định có thể có nhiều dạng khác nhau:

### 1. Giả định tính mõi đun

Các hình thức khác nhau:

Giả định tính mõi đun cho Mạng Bayes nhân quả (Giả định 4.1)

Giả định tính mõi đun cho SCM (Giả định 4.2)

Quy luật phản thực tế (Định nghĩa 4.3)

### 2. Giả định Markov Các dạng

tương đương khác nhau: Giả

định Markov cục bộ (Giả định 3.1)

Hệ số hóa mạng Bayes (Định nghĩa 3.1)

Giả định Markov toàn cục (Định lý 3.1)

Cho trước, hai giả định này (và tính tích cực), nếu tiêu chí cửa sau (Định nghĩa 4.1) được thỏa mãn trong biểu đồ nhân quả giả định của chúng ta thì chúng ta có sự nhận dạng. Lưu ý rằng mặc dù tiêu chí cửa sau là điều kiện đủ để nhận dạng nhưng nó không phải là điều kiện cần. Chúng ta sẽ thấy điều này nhiều hơn ở Chương 6.

Trang trọng hơn Nếu bạn thực sự yêu thích chủ nghĩa hình thức cầu kỳ, có một số nguồn có liên quan để bạn tham khảo. Bạn có thể thấy những tiên đề cơ bản rằng làm nền tảng cho Quy luật phản thực tế trong [20, 21], hoặc nếu bạn muốn có sách giáo khoa, bạn có thể tìm thấy chúng trong [18, Chương 7.3]. Để xem bằng chứng về sự tương đương của cả ba dạng giả định Markov, hãy xem [13, Chương 3] chẳng hạn .

Bây giờ bạn đã quen thuộc với các mô hình đồ họa nhân quả và SCM, có thể bạn nên quay lại và đọc lại Chương 2 trong khi cố gắng tạo mối liên hệ với những gì bạn đã học về mô hình đồ họa nhân quả trong hai chương vừa qua.

[20]: Galles và Pearl (1998), 'Một tiên đề Đặc điểm của phản tác dụng nhân quả cũng'

[21]: Halpern (1998), 'Tiên đề hóa nguyên nhân Lý luận'

[18]: Ngọc Trai (2009), Nhân Quả

[13]: Koller và Friedman (2009), Mô hình đồ họa xác suất: Nguyên tắc và kỹ thuật

Mối liên hệ với Không can thiệp, Nhất quán và Tích cực Giả định không có can thiệp (Giả định 2.4) thường ẩn trong biểu đồ nhân quả, vì kết quả (nghĩ ) thường chỉ có một nút duy nhất (nghĩ ) để điều trị với tư cách là cha mẹ, thay vì hơn là có nhiều nút xử lý , 1, +1, v.v. làm nút cha. Tuy nhiên, DAG nguyên nhân có thể được mở rộng đến các cài đặt có nhiều [22]. Tính nhất quán (Giả định 2.5) tuân theo các tiên đề của SCM (xem [18, Hé quả 7.3.2] và [23]).

Tính tích cực (Giả định 2.3) vẫn là một giả định rất quan trọng mà chúng ta phải đưa ra, mặc dù đôi khi nó bị bỏ qua trong tài liệu về mô hình đồ họa.

[22]: Ogburn và VanderWeele (2014),

'Sơ đồ nguyên nhân gác nhiều'

[18]: Ngọc Trai (2009), Nhân Quả

[23]: Pearl (2010), 'Về quy tắc nhất quán trong suy luận nhân quả: tiên đề, định nghĩa, giả định hay định lý?'

# 5

## Thí nghiệm ngẫu nhiên

Các thí nghiệm ngẫu nhiên khác biệt đáng kể so với quan sát học. Trong các thí nghiệm ngẫu nhiên, người thực nghiệm có toàn quyền kiểm soát cơ chế phân công điều trị (cách phân công điều trị). Ví dụ, trong loại thử nghiệm ngẫu nhiên đơn giản nhất, Người thí nghiệm chỉ định ngẫu nhiên (ví dụ bằng cách tung đồng xu) cho mỗi người tham gia nhóm điều trị hoặc nhóm đối chứng. Sự kiểm soát hoàn toàn này về cách lựa chọn phương pháp điều trị là điểm khác biệt giữa các thử nghiệm ngẫu nhiên với các nghiên cứu quan sát. Trong thiết lập thử nghiệm đơn giản này, điều trị hoàn toàn không phải là một chức năng của đồng biển! Ngược lại, trong quan sát nghiên cứu, việc xử lý hầu như luôn luôn là một chức năng của một số đồng biển. Như chúng ta sẽ thấy, sự khác biệt này là chìa khóa để xác định liệu nhiều có hiện diện trong dữ liệu của chúng tôi.

Trong các thí nghiệm ngẫu nhiên, sự liên kết là quan hệ nhân quả. Điều này là do các thí nghiệm ngẫu nhiên đặc biệt ở chỗ chúng đảm bảo rằng không có gây bối rối. Kết quả là, điều này cho phép chúng ta đo lường tác động nhân quả  $[(-1)] \quad [(-0)]$  thông qua sự khác biệt liên kết  $[ \quad | \quad = 1] \quad [ \quad | \quad = 0]$ .

Trong các phần sau đây, chúng tôi giải thích lý do tại sao điều này xảy ra từ nhiều của những quan điểm khác nhau. Nếu bất kỳ lời giải thích nào trong số này phù hợp với bạn, điều đó có thể đủ tốt. Chắc chắn bám sát theo cách trực quan nhất giải thích hấp dẫn ở Mục 5.3.

### 5.1 Khả năng so sánh và cân bằng hiệp phương sai

Lý tưởng nhất là nhóm can thiệp và đối chứng sẽ giống nhau về mọi mặt, các khía cạnh, ngoại trừ việc điều trị. Điều này có nghĩa là chúng chỉ khác nhau ở sự đối xử mà họ nhận được (tức là chúng có thể so sánh được). Điều này sẽ cho phép chúng tôi quy định bất kỳ sự khác biệt nào trong kết quả điều trị và kiểm soát các nhóm vào điều trị. Nói rằng các nhóm điều trị này đều giống nhau trong mọi thử khác ngoài việc điều trị và kết quả của họ đều giống như nói rằng họ có cùng sự phân bổ các yếu tố gây nhiễu. Bởi vì mọi người thường kiểm tra đặc tính này trên các biến quan sát (thường là những gì mọi người có nghĩa là "đồng biển"), khái niệm này được gọi là cân bằng đồng biển.

**Định nghĩa 5.1 (Cân bằng hiệp phương sai)** Chúng ta có cân bằng hiệp phương sai nếu sự phân bổ của hiệp phương sai là giống nhau giữa các nhóm can thiệp. Chính thức hơn,

$$(-1) = (-0) \quad (5.1)$$

Việc ngẫu nhiên hóa ngữ ý sự cân bằng đồng biển, trên tất cả các biến đồng thời, thậm chí những cái không được quan sát. Theo trực giác, điều này là do việc điều trị được chọn ở ngẫu nhiên, bắt kè, do đó nhóm can thiệp và đối chứng nên trông rất giống nhau. Bằng chứng rất đơn giản. Bởi vì hoàn toàn không xác định bởi (chỉ bằng việc lật đồng xu), độc lập với . Điều này có nghĩa rằng

5.1 Khả năng so sánh và cân bằng đồng biển . . . . .	49
5.2 Khả năng trao đổi . . . . .	50
5.3 Không có đường dẫn cửa sau . . . . .	51

Ký hiệu = có nghĩa là "phân phối bình đẳng" sự."

$(\quad | \quad = 1) = (\quad)$ . Tương tự, nó có nghĩa là  $(\quad | \quad = 0) = (\quad)$ . Vì thế,  
ta có  $(\quad | \quad = 1) = (\quad | \quad = 0)$ .

Mặc dù chúng tôi đã chứng minh rằng sự ngẫu nhiên hóa ngũ ý sự cân bằng đồng biến, chúng tôi chưa chứng minh rằng sự cân bằng hiệp phương sai đó ngũ ý rằng mối liên hệ là quan hệ nhân quả.<sup>1</sup> Bây giờ chúng tôi sẽ chứng minh điều đó bằng cách chỉ ra rằng  $(\quad | \text{do}(\quad)) = (\quad | \quad)$ . Để chứng minh, tính chất chính mà chúng ta sử dụng là cân bằng đồng biến hàm ý và là độc lập.

Bằng chứng. Đầu tiên, giả sử là một tập điều chỉnh đủ có khả năng chứa các biến không được quan sát (sự ngẫu nhiên hóa cũng cân bằng các hiệp phương sai không được quan sát).

Tập hợp điều chỉnh như vậy phải tồn tại vì chúng tôi cho phép nó chứa bất kỳ biến nào, được quan sát hoặc không được quan sát. Khi đó, chúng ta có kết quả sau từ việc điều chỉnh của sau (Định lý 4.2):

$$(\quad | \text{làm}(\quad)) = (\quad | \quad, \quad) \quad (5.2)$$

Bằng cách nhân với chúng ta thu được phân bố chung của tử số:  $(\quad | \quad)$ ,

$$= \frac{(\quad | \quad, \quad)}{(\quad | \quad)} \quad (5.3)$$

$$= \frac{(\quad | \quad, \quad)}{(\quad | \quad)} \quad (5.4)$$

Bây giờ, chúng ta sử dụng tính chất quan trọng là :

$$= \frac{(\quad | \quad, \quad)}{(\quad | \quad)} \quad (5.5)$$

Việc áp dụng quy tắc Bayes và sự gạt ra ngoài lề sẽ cho chúng ta phần còn lại:

$$= (\quad, \quad | \quad) \quad (5.6)$$

$$= (\quad | \quad) \quad (5.7)$$

□

## 5.2 Khả năng trao đổi

Khả năng trao đổi (Giả định 2.1) cho chúng ta một góc nhìn khác về lý do tại sao ngẫu nhiên hóa lại khiến quan hệ nhân quả ngang bằng với mối liên kết. Để biết lý do tại sao, hãy xem xét thí nghiệm tưởng tượng sau đây. Chúng tôi quyết định nhóm can thiệp của một cá nhân bằng cách tung đồng xu ngẫu nhiên như sau: nếu đồng xu là mặt ngửa, chúng tôi chỉ định cá nhân đó vào nhóm can thiệp ( $\quad = 1$ ), và nếu đồng xu là mặt sấp, chúng tôi chỉ định cá nhân đó vào nhóm đối chứng ( $\quad = 0$ ). Nếu các nhóm có thể trao đổi được, chúng ta có thể trao đổi các nhóm này và kết quả trung bình sẽ giữ nguyên. Điều này đúng về mặt trực giác nếu chúng ta chọn các nhóm tung đồng xu. Hãy tưởng tượng đơn giản là hoán đổi ý nghĩa của "đầu" và "đuôi" trong thí nghiệm này. Bạn có mong đợi điều đó sẽ thay đổi kết quả không? Không. Đây là lý do tại sao các thí nghiệm ngẫu nhiên mang lại cho chúng ta khả năng trao đổi.

<sup>1</sup> Hãy nhớ lại rằng trực giác cho thấy sự cân bằng đồng biến có nghĩa là mọi thứ đều giống nhau giữa các nhóm can thiệp, ngoại trừ can thiệp, do đó can thiệp phải là lời giải thích cho sự thay đổi trong .

Nhớ lại Phần 2.3.2 rằng khả năng trao đổi trung bình chính thức như sau:

$$[\quad(1) \mid \quad = 1] = [\quad(1) \mid \quad = \quad] \quad (5.8)$$

$$[\quad(0) \mid \quad = 0] = [\quad(0) \mid \quad = 1] \quad (5.9)$$

Sự "trao đổi" là khi chúng ta đi từ  $(1)$  ở nhóm can thiệp đến  $(1)$  ở nhóm đối chứng (Công thức 5.8) và từ  $(0)$  ở nhóm đối chứng đến  $(0)$  ở nhóm can thiệp (Phương trình 5.9).

Để xem bằng chứng tại sao mỗi liên hệ lại là quan hệ nhân quả trong các thí nghiệm ngẫu nhiên qua lăng kính khả năng trao đổi, hãy nhớ lại bằng chứng ở Phần 2.3.2. Đầu tiên, hãy nhớ lại rằng Công thức 5.8 có nghĩa là cả hai đại lượng trong nó đều bằng kết quả mong đợi cận biên  $[\quad(1)]$  và tương tự, Công thức 5.8 có nghĩa là cả hai đại lượng trong nó đều bằng kết quả mong đợi cận biên  $[\quad(0)]$ . Khi đó, ta có chứng minh sau:

$$\begin{aligned} [\quad(1)] - [\quad(0)] &= [\quad(1) \mid \quad = 1] - [\quad(0) \mid \quad = 0] \quad (\text{xem lại 2.3}) \\ &= [\quad \mid \quad = 1] - [\quad \mid \quad = 0] \quad (\text{xem lại 2.4}) \end{aligned}$$

### 5.3 Không có đường dẫn cửa sau

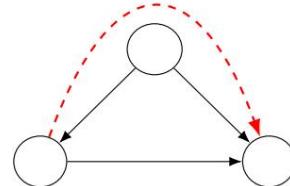
Quan điểm cuối cùng mà chúng ta sẽ xem xét để biết tại sao mỗi liên hệ lại là quan hệ nhân quả trong các thử nghiệm ngẫu nhiên chính là quan điểm của các mô hình nhân quả đồ họa. Trong dữ liệu quan sát thường xuyên, hầu như luôn có sự nhiễu loạn. Ví dụ, trong Hình 5.1, chúng ta thấy rằng là yếu tố gây nhiễu ảnh hưởng của lên Liên kết phi nhân quả chảy dọc theo con đường cửa sau.

Tuy nhiên, nếu chúng ta chọn ngẫu nhiên , điều kỳ diệu sẽ xảy ra: không còn có cha mẹ nhân quả nào nữa, như chúng ta mô tả trong Hình 5.2. Điều này là do hoàn toàn là ngẫu nhiên. Nó không phụ thuộc vào bất cứ điều gì ngoài kết quả của việc tung đồng xu (hoặc bộ tạo số ngẫu nhiên lượng tử, nếu bạn thích loại công cụ này). Bởi vì không có cạnh vào nêu trong quá trình ngẫu nhiên hóa, không có đường dẫn cửa sau. Vậy tập rỗng là tập điều chỉnh đủ. Điều này có nghĩa là tất cả mối liên hệ từ đến đều có tính nhân quả. Chúng ta có thể xác định  $(\quad | \text{do}(\quad = \quad))$  bằng cách áp dụng điều chỉnh cửa sau (Định lý 4.2), điều chỉnh cho tập trống:

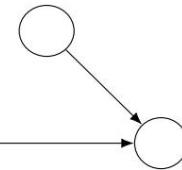
$$(\quad | \text{làm}(\quad = \quad)) = (\quad | \quad = \quad)$$

Với điều đó, chúng tôi kết thúc cuộc thảo luận về lý do tại sao mỗi liên hệ lại là quan hệ nhân quả trong các thí nghiệm ngẫu nhiên. Hy vọng rằng ít nhất một trong ba cách giải thích này dễ hiểu đối với bạn và dễ dàng lưu giữ trong trí nhớ dài hạn.

hiệp hội gây nhiễu



Hình 5.1: Cấu trúc nhân quả của lén ảnh hưởng đến lén .



Hình 5.2: Cấu trúc nhân quả khi chạy thông trại điều trị.

# Nhận dạng phi tham số 6

Trong Phần 4.4, chúng ta thấy rằng việc đáp ứng tiêu chí cửa sau là đủ để cung cấp cho chúng tôi khả năng nhận dạng, nhưng liệu tiêu chí cửa sau có cần thiết không? Nói cách khác, liệu có thể nhận dạng được mà không thể chặn tất cả các đường dẫn cửa sau?

Ví dụ: hãy xem xét rằng chúng tôi có dữ liệu được tạo theo đồ thị trong hình 6.1. Chúng tôi không quan sát thấy trong dữ liệu này nên chúng tôi không thể chặn đường dẫn cửa sau qua và liên kết gây nhiễu chảy dọc theo nó. Nhưng chúng ta vẫn cần xác định được nguyên nhân. Hóa ra là nó có thể xác định tác động nhân quả trong biểu đồ này bằng cách sử dụng cửa trước tiêu chuẩn. Chúng ta sẽ xem tiêu chí frontdoor và điều chỉnh tương ứng trong Phần 6.1. Sau đó, chúng ta sẽ xem xét việc nhận dạng tổng quát hơn nữa trong Phần 6.2 khi chúng tôi giới thiệu phép tính. Chúng ta sẽ kết thúc bằng đồ họa điều kiện nhận dạng tại Mục 6.3.

## 6.1 Điều chỉnh cửa trước

Trực giác cao về lý do tại sao chúng ta có thể xác định được tác động nhân quả của đối với trong biểu đồ ở Hình 6.1 (ngay cả khi chúng ta không thể điều chỉnh yếu tố gây nhiễu vì nó không được quan sát) như sau: một người hòa giải như rất hữu ích; chúng ta có thể cô lập sự liên kết chảy qua bằng cách tập trung phân tích thống kê của chúng tôi về và mỗi liên hệ duy nhất xuyên suốt là mỗi liên hệ nhân quả (sự liên kết đi dọc theo các đường dẫn có hướng từ đến). Chúng tôi minh họa trực giác này trong Hình 6.2, trong đó chúng tôi chỉ mô tả mỗi liên hệ nhân quả. Trong phần này, chúng tôi sẽ tập trung phân tích vào bằng cách sử dụng quy trình ba bước (xem Hình 6.3 để minh họa tương ứng):

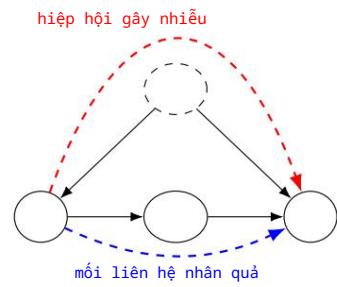
1. Xác định tác động nhân quả của lên .
2. Xác định tác động nhân quả của lên .
3. Kết hợp các bước trên để xác định tác động nhân quả của lên .

Bước 1 Đầu tiên, chúng ta sẽ xác định tác dụng của lên : ( | do( ) ). Bởi vì là một máy va chạm trên đường dẫn xuyên qua , nó chặn đường dẫn cửa sau đó. Vì vậy không có đường dẫn cửa sau nào được bỏ chặn từ đến . Điều này có nghĩa rằng mỗi liên hệ duy nhất chảy từ đến là mỗi liên hệ nhân quả mà chảy dọc theo cạnh nối chúng. Vì vậy, chúng tôi có những điều sau đây nhận dạng thông qua điều chỉnh cửa sau (Định lý 4.2, sử dụng khoảng trống đặt làm bộ điều chỉnh):<sup>1</sup>

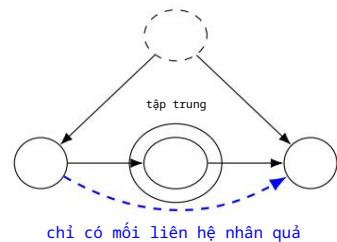
$$( | \text{làm}( )) = ( | ) \quad (6.1)$$

Bước 2 Thứ hai, chúng ta sẽ xác định tác dụng của lên : ( | do( ) ). Vì chặn đường dẫn cửa sau nên chúng ta có thể đơn giản

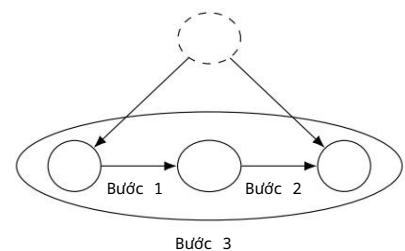
6.1 Điều chỉnh cửa trước . . . . .	52
6.2 thực hiện phép tính . . . . .	55
Ứng dụng: Điều chỉnh cửa trước . . . . .	57
6.3 Xác định khả năng nhận dạng từ đồ thị . . . . .	58



Hình 6.1: Đồ thị nhân quả trong đó không được quan sát nên chúng ta không thể chặn đường đi cửa sau. Chúng tôi mô tả dòng chảy nhân quả sự liên kết và dòng chảy gây nhiễu với các đường đứt nét.



Hình 6.2: Ngược lại với Hình 6.1, khi chúng tôi tập trung phân tích vào , chúng tôi có thể chỉ để cô lập mỗi liên hệ nhân quả .



Hình 6.3: Minh họa các bước thực hiện điều chỉnh cửa trước.

<sup>1</sup> **Bài tập đọc tích cực:** Viết chứng minh cho Phương trình 6.1 không sử dụng cửa sau điều chỉnh. Thay vào đó, hãy bắt đầu từ hệ số rút gọn (Mệnh đề 4.1) như chúng tôi đã làm ở Phần 4.3.1. Gợi ý: Chúng minh có thể khá ngắn. Chúng tôi cung cấp bằng chứng trong Phụ lục A.1, trong trường hợp bạn gặp khó khăn.

điều chỉnh cho . Do đó, sử dụng lại điều chỉnh cửa sau, chúng ta có kết quả sau:

$$( | \text{làm}( )) = ( | , ) ( ) \quad (6.2)$$

Bước 3 Bây giờ chúng ta đã biết cách thay đổi thay đổi (bước 1) và cách thay đổi thay đổi (bước 2), chúng ta có thể kết hợp cả hai điều này để biết cách thay đổi thay đổi (qua ):

$$( | \text{làm}( )) = ( | \text{làm}( )) ( | \text{làm}( )) \quad (6.3)$$

Thứ số đầu tiên ở phía bên phải tương ứng với việc đặt thành và quan sát giá trị kết quả của . Yếu tố thứ hai tương ứng với việc cài đặt thành chính xác giá trị có được từ việc cài đặt và sau đó quan sát giá trị nào của kết quả . Chúng ta phải tính tổng trên vì  $( | do( ))$  có tính xác suất, vì vậy chúng ta phải tính tổng độ hỗ trợ của nó. Nói cách khác, chúng ta phải tính tổng tất cả các cách thể hiện có thể có của các biến ngẫu nhiên có phân phối là  $( | do( ))$ .

Sau đó, nếu chúng ta thay Công thức 6.1 và 6.2 vào Công thức 6.3, chúng ta sẽ có được sự điều chỉnh cửa trước (tiếp tục đọc để xem định nghĩa của tiêu chí cửa trước):

**Định lý 6.1 (Điều chỉnh Frontdoor)** Nếu  $( , , )$  thỏa mãn tiêu chí frontdoor và ta có số dương thì

$$( | \text{làm}( )) = ( | ) ( | , ) ( ) \quad (6.4)$$

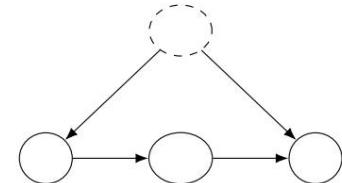
Đồ thị nhân quả mà chúng ta đang sử dụng (Hình 6.4) là một ví dụ về đồ thị đơn giản thỏa mãn tiêu chí cửa trước. Để có được định nghĩa đầy đủ, trước tiên chúng ta phải định nghĩa sự hòa giải đầy đủ/hoàn chỉnh: một tập hợp các biến làm trung gian hoàn toàn tác động của lên nếu tất cả các đường dẫn nhân quả (có hướng) từ đến đều đi qua . Bây giờ chúng tôi đưa ra định nghĩa chung về tiêu chí cửa trước:

**Định nghĩa 6.1 (Tiêu chí Frontdoor)** Một tập hợp các biến thỏa mãn tiêu chí frontdoor liên quan đến và nếu những điều sau đây là đúng:

1. hoàn toàn làm trung gian tác động của lên (tức là tất cả các đường nhân quả từ đến đều đi qua ).
2. Không có đường dẫn cửa sau nào được bỏ chặn từ đến .
3. Tất cả các đường dẫn cửa sau từ đến đều bị chặn bởi .

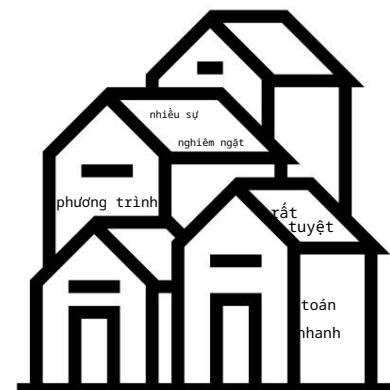
Mặc dù các phương trình 6.1 và 6.2 là những ứng dụng đơn giản của việc điều chỉnh cửa sau, chúng tôi đã tự tay tiến tới phương trình 6.3, đây là chìa khóa cho việc điều chỉnh cửa trước (Định lý 6.1). Bây giờ chúng ta sẽ hướng dẫn cách lấy phương trình 6.3. **Bài tập đọc tích cực:** Hãy dừng đọc ở đây và tự mình làm việc này.

Chúng ta sắp bước vào Equationtown (Hình 6.5), vì vậy nếu bạn hài lòng với trực giác mà chúng tôi đưa ra cho bước 3 và không muốn nhìn thấy nhiều phương trình, vui lòng bỏ qua đến cuối phần chứng minh (ký hiệu bằng ký hiệu) .



Hình 6.4: Đồ thị nhân quả đơn giản thỏa mãn tiêu chí cửa trước

<sup>2</sup> **Bài tập đọc tích cực:** Hãy nghĩ về một biểu đồ khác với Hình 6.4 thỏa mãn tiêu chí cửa trước. Ngoài ra, với mỗi điều kiện, hãy nghĩ về một đồ thị không chỉ thỏa mãn điều kiện đó.



Hình 6.5: Equationtown

Bằng chứng. Như thường lệ, chúng ta bắt đầu với việc phân tích nhân tử rút gọn, sử dụng biểu đồ nhân quả trong Hình 6.4. Từ hệ số hóa mạng Bayes (Định nghĩa 3.1), ta có:

$$( \cdot, \cdot, \cdot, \cdot ) = ( \cdot ) ( \cdot | \cdot ) ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) \quad (6.5)$$

Sau đó, bằng cách sử dụng phép phân tích rút gọn (Mệnh đề 4.1), chúng ta loại bỏ hệ số cho :

$$( \cdot, \cdot, \cdot, \cdot | \text{do}(\cdot) ) = ( \cdot ) ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) \quad (6.6)$$

Tiếp theo, chúng tôi loại trừ và :

$$( \cdot, \cdot, \cdot, \cdot | \text{l\^am}(\cdot) ) = ( \cdot ) ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) \quad (6.7)$$

$$( \cdot | \text{l\^am}(\cdot) ) = ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot ) \quad (6.8)$$

Mặc dù chúng ta đã loại bỏ tất cả các toán tử do, hãy nhớ rằng chúng ta vẫn chưa hoàn thành vì không được quan sát. Vì vậy chúng ta cũng phải xóa khỏi sự biểu lộ. Đây là lúc chúng ta phải sáng tạo một chút.

Chúng ta muốn có thể kết hợp  $( \cdot | \cdot, \cdot )$  và  $( \cdot )$  thành một thừa số chung trên cả và để có thể loại trừ . Để làm điều này, chúng ta cần đặt phía sau thanh điều kiện của hệ số  $( \cdot )$ . Cái này sẽ

sẽ dễ dàng nếu chúng ta có thể hoán đổi  $( \cdot )$  lấy  $( \cdot | \cdot )$  trong phương trình 6.8.<sup>3</sup>

Điều quan trọng cần lưu ý là chúng ta thực sự có thể bao gồm phía sau thanh điều hòa nếu cũng ở đó vì d-tách khỏi trong Hình 6.6. Trong toán học, điều này có nghĩa là đẳng thức sau được giữ:

$$( \cdot | \cdot ) = ( \cdot | \cdot, \cdot ) \quad (6.9)$$

Tuyệt vời, vậy làm cách nào để đưa vào bữa tiệc này? Thủ thuật thông thường là điều hòa nó và gạt nó ra ngoài lề:

$$( \cdot | \text{l\^am}(\cdot) ) = ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot ) \quad (6.8 \text{ xem lại})$$

$$= ( \cdot | \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot | \cdot ) ( \cdot ) \quad (6.10)$$

$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot | \cdot ) ( \cdot ) \quad (6.11)$$

$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot | \cdot ) ( \cdot ) \quad (6.12)$$

Tuyệt, nhưng bây giờ chúng ta không thể kết hợp  $( \cdot | \cdot, \cdot )$  và  $( \cdot | \cdot )$  vì  $( \cdot | \cdot, \cdot )$  thiếu mới được giới thiệu này đằng sau thanh điều hòa. May mắn thay, chúng ta có thể khắc phục điều đó và kết hợp hai yếu tố:

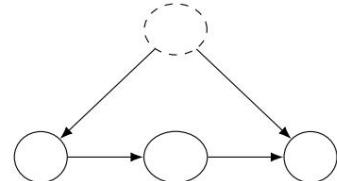
$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot | \cdot ) ( \cdot ) \quad (6.13)$$

$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) ( \cdot | \cdot ) \quad (6.14)$$

$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) \quad (6.15)$$

$$= ( \cdot | \cdot ) ( \cdot ) ( \cdot ) ( \cdot | \cdot, \cdot ) \quad (6.16)$$

<sup>3</sup> **Bài tập đọc tích cực:** Tại sao lại dễ dàng loại bỏ nếu trường hợp  $( \cdot ) = ( \cdot | \cdot )$ ? Và tại sao sự bình đẳng này không được giữ vững?



Hình 6.6: Đồ thị nhân quả đơn giản thỏa mãn đáp ứng tiêu chí cửa trước

<sup>4</sup> **Bài tập đọc tích cực:** Tại sao  $( \cdot | \cdot, \cdot )$  bằng  $( \cdot | \cdot )$ ?

Điều này phù hợp với kết quả được nêu trong Định lý 6.1, vì vậy chúng ta đã hoàn thành việc suy ra phép điều chỉnh cửa trước mà không cần sử dụng phép điều chỉnh cửa sau. Tuy nhiên, chúng ta vẫn cần chứng minh rằng Phương trình 6.3 đúng để chứng minh cho bước 3. Để làm được điều đó, tất cả những gì còn lại là nhận ra rằng các phần này khớp với Phương trình 6.1 và 6.2 và thế chúng vào:

$$\begin{aligned}
 & = ( | \text{làm} ) = ( | ) \\
 & = ( | \text{làm} ) = ( | , ) \quad (6.1) \\
 & = ( | \text{làm} ) = ( | ) \quad (6.2)
 \end{aligned}$$

□

Và chúng ta đã hoàn tất! Chúng tôi chỉ cần thông minh hơn một chút trong việc sử dụng phân tách d và loại trừ. Một phần lý do tại sao chúng tôi thực hiện việc chứng minh đó là vì chúng tôi sẽ chứng minh việc điều chỉnh cửa trước bằng cách sử dụng phép tính trong Phần 6.2. Bằng cách này, bạn có thể dễ dàng so sánh một chứng minh sử dụng hệ số rút gọn với một chứng minh sử dụng phép tính để chứng minh cùng một kết quả.

## 6.2 làm phép tính

Như chúng ta đã thấy trong phần trước, hóa ra việc đáp ứng tiêu chí của sau (Định nghĩa 4.1) là không cần thiết để xác định các tác động nhân quả. Ví dụ: nếu tiêu chí của trước (Định nghĩa 6.1) được thỏa mãn, điều đó cũng mang lại cho chúng ta khả năng nhận dạng. Điều này dẫn đến các câu hỏi sau: liệu chúng ta có thể xác định các ước tính nhân quả khi đồ thị nhân quả liên quan không thỏa mãn tiêu chí của sau cũng như tiêu chí của trước không? Nếu vậy thì làm thế nào?

Phép tính thực của Pearl [24] cho chúng ta câu trả lời cho những câu hỏi này.

[24]: Pearl (1995), 'Sơ đồ nhân quả cho nghiên cứu thực nghiệm'

Như chúng ta sẽ thấy, phép tính thực hiện cung cấp cho chúng ta các công cụ để xác định các tác động nhân quả bằng cách sử dụng các giả định nhân quả được mã hóa trong biểu đồ nhân quả. Nó sẽ cho phép chúng tôi xác định bất kỳ ước tính nhân quả nào có thể xác định được. Cụ thể hơn, hãy xem xét một ước lượng nhân quả tùy ý  $( | \text{do}( = ) )$ ,  $=$ , trong đó  $\text{do}$  là một tập hợp tùy ý các biến kết quả,  $=$  là một tập hợp các biến can thiệp tùy ý, và  $\text{do}$  là một tập hợp tùy ý (có khả năng trống) gồm các hiệp phương sai mà chúng ta muốn chọn mức độ cụ thể của tác động nhân quả mà chúng ta đang xem xét. Lưu ý rằng điều này có nghĩa là chúng ta có thể sử dụng phép tính thực tế để xác định các tác động nhân quả khi có nhiều phương pháp điều trị và/hoặc nhiều kết quả.

Để trình bày các quy tắc tính toán, chúng ta phải xác định một chút ký hiệu cho các dạng mở rộng của đồ thị nhân quả . Đặt  $\text{biểu thị đồ thị}$  mà chúng ta có được nếu chúng ta lấy  $\text{và loại bỏ tất cả các cạnh tới các nút}$  trong tập hợp ; nhớ lại Phần 4.2 rằng đây được gọi là biểu đồ thao tác.

Đặt  $\text{biểu thị đồ thị}$  mà chúng ta có được nếu lấy  $\text{và loại bỏ tất cả các cạnh đi ra khỏi các nút trong tập hợp}$ . Ý nghĩa giúp bạn ghi nhớ điều này là nghĩ về cha mẹ như được vẽ phía trên con cái của họ trong biểu đồ, vì vậy thanh phia trên đang cắt các cạnh vào của nó và thanh bên dưới đang cắt các cạnh ra của nó. Kết hợp cả hai điều này, chúng ta sẽ sử dụng để biểu thị đồ thị có các cạnh vào và các cạnh ra từ bị loại bỏ. Và nhớ lại Phần 3.7 rằng chúng ta sử dụng để biểu thị sự phân tách d trong . Bây giờ chúng tôi đã sẵn sàng; phép tính chỉ bao gồm ba quy tắc:

**Định lý 6.2 (Quy tắc thực hiện phép tính)** Cho đồ thị nhân quả , liên kết

phân phối và các tập hợp rời rạc của các biến , , và , các quy tắc sau đây đúng.

Quy tắc 1:

$$( \quad | \text{làm}( \quad ), \quad , \quad ) = ( \quad | \text{làm}( \quad ), \quad ) \text{ if } \quad - \quad | \quad , \quad (6.18)$$

Quy tắc 2:

$$( \quad | \text{do}( \quad ), \text{do}( \quad ), \quad ) = ( \quad | \text{do}( \quad ), \quad , \quad ) \text{ if } \quad - \quad | \quad , \quad (6.19)$$

Quy tắc 3:

$$( \quad | \text{do}( \quad ), \text{do}( \quad ), \quad ) = ( \quad | \text{do}( \quad ), \quad ) \text{ nếu } \quad - \quad | \quad , \quad (6.20)$$

(6.20) trong đó ( ) biểu thị tập hợp các nút của không phải là tổ tiên của bất kỳ nút nào của trong .

Bây giờ, thay vì tái tạo lại các cách chứng minh các quy tắc này từ Pearl [24], chúng ta sẽ đưa ra trực giác cho từng quy tắc đó dưới dạng các khái niệm mà chúng ta đã thấy trong cuốn sách này.

[24]: Pearl (1995), 'Sơ đồ nhân quả cho nghiên cứu thực nghiệm'

Quy tắc 1 Trực giác Nếu chúng ta áp dụng Quy tắc 1 và chỉ cần loại bỏ can thiệp do( ), chúng ta sẽ nhận được kết quả sau ([Bài tập đọc tích cực](#): đây là khái niệm quen thuộc nào?):

$$( \quad | \quad , \quad ) = ( \quad | \quad ) \text{ nếu } \quad - \quad | \quad , \quad (6.21)$$

Đây chính là điều mà phép tách d mang lại cho chúng ta theo giả định Markov; nhớ lại Định lý 3.1 rằng phép tách d trong đồ thị hàm ý tính độc lập có điều kiện trong . Điều này có nghĩa là Quy tắc 1 chỉ đơn giản là sự khái quát hóa Định lý 3.1 cho phân phối can thiệp.

Trực giác Quy tắc 2 Cũng giống như Quy tắc 1, chúng ta sẽ loại bỏ can thiệp do( ) khỏi Quy tắc 2 và xem điều này nhắc nhở chúng ta điều gì ([Bài tập đọc tích cực](#): điều này làm bạn nhớ đến khái niệm nào?):

$$( \quad | \text{do}( \quad ), \quad ) = ( \quad | \quad , \quad ) \text{ nếu } \quad - \quad | \quad , \quad (6.22)$$

Đây chính xác là những gì chúng ta làm khi chứng minh việc điều chỉnh cửa sau (Định lý 4.2) bằng cách sử dụng tiêu chí cửa sau (Định nghĩa 4.1). Như chúng ta đã thấy ở phần cuối của Phần 3.8 và Phần 4.4. Sự kết hợp là quan hệ nhân quả nếu kết quả và cách xử lý được phân tách d bởi một số tập hợp biến phụ thuộc vào . Vì vậy, quy tắc 2 là sự khái quát hóa việc điều chỉnh cửa sau đối với các phân phối can thiệp.

Quy tắc 3 Trực giác Đây là quy tắc khó hiểu nhất. Cũng giống như hai quy tắc còn lại, trước tiên chúng ta sẽ loại bỏ can thiệp do ( ) để giúp việc suy nghĩ về vấn đề này trở nên đơn giản hơn:

$$( \quad | \text{do}( \quad ), \quad ) = ( \quad | \quad ) \text{ nếu } \quad - \quad | \quad , \quad (6.23)$$

Để có được dargestellung thức trong phương trình này, phải xảy ra trường hợp loại bỏ can thiệp do( ) (giống như lấy đồ thị đã thao tác và đưa lại các cạnh đi vào ) không tạo ra liên kết mới nào có thể ảnh hưởng đến . Bởi vì do( ) loại bỏ các cạnh đến của để cho chúng ta , nên mỗi liên kết chính mà chúng ta cần lo lắng là liên kết chảy từ đến — trong (liên kết nhân quả). Vì vậy, bạn có thể

mong đợi rằng điều kiện mang lại cho chúng ta sự đồng thức trong phương trình 6.23 là

— | . Tuy nhiên, chúng ta phải tính chính điều này một chút để tránh gây ra sự liên kết bằng cách điều kiện hóa con cháu của các máy và chạm (xem lại Phần 3.6). Cụ thể, có thể chứa các máy và chạm trong , và có thể chứa hậu duệ của các máy và chạm này. Do đó, để không tạo ra liên kết mới thông qua các bộ va chạm trong khi chúng ta đưa lại các cạnh vào để có được , chúng ta phải giới hạn tập hợp các nút bị thao tác ở những nút không phải là tổ tiên của các nút trong tập điều hòa : ( ).

Tính đầy đủ của phép tính thực tế Có thể tồn tại các ước tính nhân quả có thể xác định được nhưng không thể xác định được nếu chỉ sử dụng các quy tắc tính toán thực hiện trong Định lý 6.2. May mắn thay, Shpitser và Pearl [25] , Huang và Valtorta [26] đã chứng minh một cách độc lập rằng điều này không xảy ra.

Họ đã chứng minh rằng phép tính thực hiện là hoàn chỉnh, có nghĩa là ba quy tắc này đủ để xác định tất cả các ước tính nhân quả có thể xác định được. Bởi vì những bằng chứng này mang tính xây dựng nên chúng cũng thừa nhận các thuật toán xác định bắt kỳ ước tính nhân quả nào trong thời gian đa thức.

Nhận dạng phi tham số Lưu ý rằng tất cả những điều này là về nhận dạng phi tham số ; nói cách khác, phép tính cho chúng ta biết liệu chúng ta có thể xác định được một ước tính nhân quả nhất định chỉ bằng cách sử dụng các giả định nhân quả được mã hóa trong biểu đồ nhân quả hay không. Nếu chúng tôi đưa ra nhiều giả định hơn về phân phối (ví dụ: tính tuyến tính), chúng tôi có thể xác định được nhiều ước tính nhân quả hơn. Điều đó sẽ được gọi là nhận dạng tham số . Chúng ta không thảo luận về nhận dạng tham số trong chương này, mặc dù chúng ta sẽ thảo luận trong các chương sau.

[25]: Shpitser và Pearl (2006), 'Xác định các phân phối can thiệp chung trong mô hình nguyên nhân bán Markovian đê quayels'

[26]: Huang và Valtorta (2006), 'Ngọc trai' Tính toán can thiệp đã hoàn tất'

### 6.2.1 Ứng dụng: Điều chỉnh cửa trước

Nhớ lại biểu đồ đơn giản mà chúng ta đã sử dụng thỏa mãn tiêu chí cửa trước (Hình 6.7) và nhớ lại việc điều chỉnh cửa trước:

$$( | \text{làm}( )) = ( | ) ( | , ) ( | ) \quad (\text{xem lại 6.4})$$

Ở cuối Phần 6.1, chúng ta đã thấy bằng chứng cho việc điều chỉnh cửa trước chỉ sử dụng hệ số rút gọn. Để có ý tưởng về cách hoạt động của phép tính và trực giác mà chúng tôi sử dụng trong các phép chứng minh sử dụng nó, bây giờ chúng tôi sẽ thực hiện phép chứng minh điều chỉnh cửa trước bằng cách sử dụng các quy tắc của phép tính.

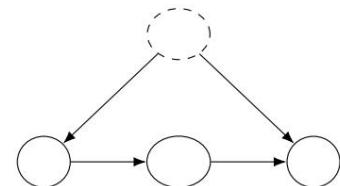
Bằng chứng. Mục tiêu của chúng ta là xác định ( | do( )). Bởi vì chúng ta có trực giác mà chúng ta đã mô tả trong Phần 6.1 rằng bộ hòa giải đầy đủ sẽ giúp chúng ta, điều đầu tiên chúng ta sẽ làm là đưa vào phương trình thông qua thủ thuật biến hóa:

$$( | \text{làm}( )) = ( | \text{làm}( ), ) ( | \text{làm}( )) \quad (6.24)$$

Bởi vì đường dẫn cửa sau từ đến trong Hình 6.7 bị chặn bởi bộ va chạm , nên tất cả các liên kết chày từ đến là nhân quả, vì vậy chúng ta có thể áp dụng Quy tắc 2 để có được kết quả sau:

$$= ( | \text{làm}( ), ) ( | ) \quad (6.25)$$

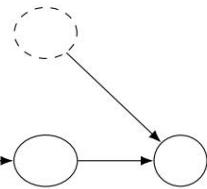
Bây giờ, vì là yếu tố trung gian đầy đủ của tác động nhân quả của lên , nên chúng ta có thể thay thế ( | do( ), ) bằng ( | do( )), nhưng điều này sẽ



Hình 6.7: Đồ thị nhân quả đơn giản thỏa mãn đáp ứng tiêu chí cửa trước

thực hiện hai bước tính toán. Để loại bỏ do( ), chúng ta cần sử dụng Quy tắc 3, quy tắc này yêu cầu không có tác động nhân quả lên trong biểu đồ liên quan. Chúng ta có thể có được điều này bằng cách loại bỏ cạnh từ đến (Hình 6.9); trong phép tính, chúng ta thực hiện điều này bằng cách sử dụng Quy tắc 2 (theo hướng ngược lại như trước) để làm( ). Chúng ta có thể làm điều này vì do( ) hiện tại làm cho nó không có đường dẫn cửa sau từ đến trong (Hình 6.8).

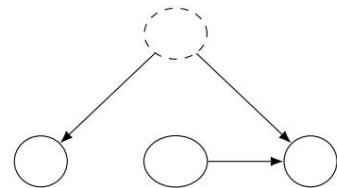
$$= ( \quad | \text{ làm}( \quad ), \text{ làm}( \quad ) ) ( \quad | \quad ) \quad (6.26)$$



Hình 6.8:

Bây giờ, như đã dự định, chúng ta có thể loại bỏ do( ) bằng Quy tắc 3. Chúng ta có thể sử dụng Quy tắc 3 ở đây vì không có mối quan hệ nhân quả nào đi từ đến — (Hình 6.9).

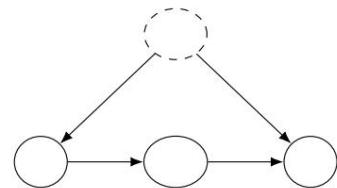
$$= ( \quad | \text{ làm}( \quad ) ) ( \quad | \quad ) \quad (6.27)$$



Hình 6.9:

Tất cả những gì còn lại là loại bỏ toán tử do cuối cùng này. Như chúng ta đã thảo luận trong Phần 6.1, chặn đường dẫn cửa sau duy nhất từ đến trong biểu đồ (Hình 6.10). Điều này có nghĩa là, nếu chúng ta có thể điều kiện hóa trên , thì chúng ta có thể loại bỏ toán tử do cuối cùng này. Như thường lệ, chúng tôi làm điều đó bằng cách điều chỉnh và loại trừ . Sắp xếp lại một chút và sử dụng để loại trừ vì đã có sẵn:

$$= ( \quad | \quad ) ( \quad | \text{ làm}( \quad ), \quad ) ( \quad | \text{ làm}( \quad ) ) \quad (6.28)$$



Hình 6.10:

Bây giờ, chúng ta có thể chỉ cần áp dụng Quy tắc 2, vì chặn đường dẫn cửa sau từ đến :

$$= ( \quad | \quad ) ( \quad | \quad , \quad ) ( \quad | \text{ làm}( \quad ) ) \quad (6.29)$$

Và cuối cùng, chúng ta có thể áp dụng Quy tắc 3 để loại bỏ do( ) cuối cùng vì không có tác động nhân quả nào của lên (tức là không có đường đi có hướng từ đến trong đồ thị trong (Hình 6.10)).

$$= ( \quad | \quad ) ( \quad | \quad , \quad ) ( \quad ) \quad (6.30)$$

□

Điều đó kết thúc bằng chứng của chúng tôi về việc điều chỉnh cửa trước bằng cách sử dụng phép tính. Nó đi theo một con đường khác với cách chứng minh mà chúng tôi đưa ra ở cuối Phần 6.1, trong đó chúng tôi sử dụng phép nhân tử rút gọn, nhưng cả hai cách chứng minh đều dựa chủ yếu vào trực giác mà chúng tôi có được khi nhìn vào biểu đồ.

**Bài tập đọc tích cực:** Giả sử tiêu chí cửa sau, chứng minh điều chỉnh cửa sau bằng quy tắc thực hiện phép tính.

### 6.3 Xác định khả năng nhận dạng từ đồ thị

Thật vui khi biết rằng chúng ta có thể xác định bất kỳ ước tính nhân quả nào có thể xác định được bằng cách sử dụng phép tính thực hiện, nhưng điều này không thỏa mãn bằng việc biết liệu ước tính nhân quả có thể được xác định bằng cách chỉ nhìn vào biểu đồ nhân quả hay không. Ví dụ: tiêu chí cửa sau (Định nghĩa 4.1) và tiêu chí cửa trước (Định nghĩa 6.1) đã cho chúng ta những cách đơn giản để biết chắc chắn rằng ước tính nhân quả có thể được xác định được. Tuy nhiên, có rất nhiều

Ước tính nhân quả có thể xác định được, mặc dù đồ thị nhân quả tương ứng không đáp ứng tiêu chí cửa sau hoặc cửa trước. Tiêu chí đồ họa tổng quát hơn tồn tại sẽ cho chúng ta biết rằng những ước tính này có thể xác định được. Chúng ta sẽ thảo luận về các tiêu chí đồ họa tổng quát hơn để nhận dạng trong phần này.

Can thiệp một biến Khi chúng ta quan tâm đến tác động nhân quả của một can thiệp lên một biến duy nhất, Tian và Pearl [27] cung cấp một tiêu chí đồ họa tương đối đơn giản, đủ để có thể nhận dạng: tiêu chí trẻ em không bị nhầm lẫn.

**Định nghĩa 6.2 (Tiêu chí con không bị nhầm lẫn)** Tiêu chí này được thỏa mãn nếu có thể chặn tất cả các đường dẫn cửa sau từ biến xử lý  $\rightarrow$  tới tất cả các con của nó là tổ tiên của  $\rightarrow$  bằng một bộ điều hòa duy nhất.

Tiêu chí này khái quát hóa tiêu chí cửa sau (Định nghĩa 4.1) và tiêu chí cửa trước (Định nghĩa 6.1). Giống như họ, đó là điều kiện đủ để nhận dạng:

**Định lý 6.3 (Khả năng nhận dạng trẻ em không bị nhầm lẫn)** Giả sử  $G$  là tập hợp các biến kết quả và  $\rightarrow$  là một biến duy nhất. Nếu tiêu chí trẻ em không bị nhầm lẫn và tính tích cực được thỏa mãn thì  $(\rightarrow = \cdot | do(\rightarrow = \cdot))$  có thể xác định được [27].

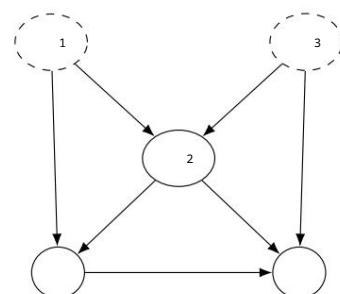
Trực giác đối với tiêu chí trẻ em không bị nhầm lẫn ngũ ý khả năng nhận dạng tương tự như trực quan đối với tiêu chí cửa trước; nếu chúng ta có thể cô lập tất cả mối liên hệ nhân quả thoát ra khỏi việc điều trị đọc theo các đường dẫn đến  $\rightarrow$ , thì chúng ta có khả năng nhận dạng được. Để thấy được trực giác này, trước tiên, hãy xem xét rằng tất cả mối liên hệ nhân quả từ  $\rightarrow$  phải chảy qua các con của nó. Chúng ta có thể cô lập mối liên hệ nhân quả này nếu không có sự nhiễu loạn giữa  $\rightarrow$  và bất kỳ nút con nào của nó.<sup>5</sup> Sự cô lập này đối với tất cả các mối liên hệ nhân quả là điều giúp chúng ta có thể nhận dạng được tác động nhân quả của  $\rightarrow$  trên bất kỳ nút nào khác trong biểu đồ. Trực giác này có thể khiến bạn nghi ngờ rằng tiêu chí này là cần thiết trong trường hợp rất cụ thể khi tập kết quả  $\rightarrow$  là tất cả các biến khác trong biểu đồ ngoài  $\rightarrow$ ; hóa ra điều này là đúng [27]. Nhưng điều kiện này không cần thiết nếu  $\rightarrow$  là tập nhỏ hơn thế.

Để giúp bạn hiểu rõ hơn về trực giác tại sao tiêu chí trẻ em không bị nhầm lẫn lại đủ để nhận dạng, chúng tôi đưa ra một biểu đồ ví dụ trong Hình 6.12. Trong Hình 6.12a, chúng ta hình dung luồng liên kết gây nhiễu và liên kết nhân quả trong biểu đồ này. Sau đó, chúng tôi mô tả sự cô lập của mối liên hệ nhân quả trong biểu đồ đó ở Hình 6.12b.

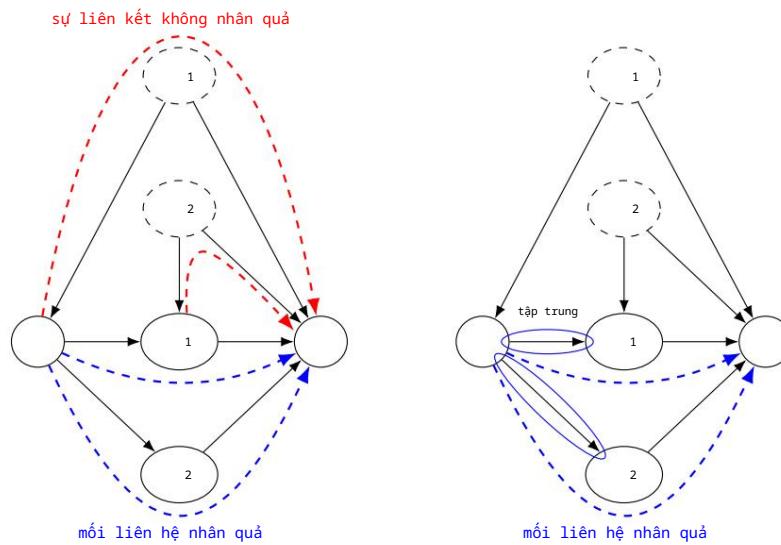
Điều kiện cần thiết Tiêu chí con không bị xáo trộn không cần thiết cho khả năng nhận dạng, nhưng nó có thể hỗ trợ trực quan đồ họa của bạn để có được điều kiện cần thiết trong đầu. Đây là một: Đối với mỗi đường dẫn cửa sau từ  $\rightarrow$  đến bất kỳ con  $\rightarrow$  nào của  $\rightarrow$  là tổ tiên của  $\rightarrow$ , có thể chặn đường dẫn đó [18, tr. 92]. Trực giác cho điều này là vì mối liên hệ nhân quả đi từ  $\rightarrow$  đến  $\rightarrow$  phải thông qua con cái của  $\rightarrow$  là tổ tiên của  $\rightarrow$ , nên để có thể cô lập mối liên kết nhân quả này, tác động của  $\rightarrow$  lên những đứa trẻ trong gian này phải là không có căn cứ. Và điều kiện tiên quyết để các mối quan hệ  $(\rightarrow = \cdot | do(\rightarrow = \cdot))$  này không bị xáo trộn là bất kỳ đường dẫn cửa sau nào từ  $\rightarrow$  đến  $\rightarrow$  đều phải chặn được (điều chúng tôi nêu trong điều kiện này). Thật không may, điều kiện này là không đủ. Để biết lý do tại sao, hãy xem Hình 6.11. Đường dẫn cửa sau

[27]: Tian and Pearl (2002), 'Một vị tướng Điều kiện xác định tác động nhân quả'

<sup>5</sup> Điều này tương tự với những gì chúng ta đã thấy với tiêu chí cửa trước trong Phần 6.1, trong đó chúng ta có thể cô lập mối liên hệ nhân quả đi qua bộ trung gian đầy đủ nếu mối quan hệ không bị ràng buộc (không có đường dẫn cửa sau nào được bỏ chặn).



Hình 6.11: Biểu đồ trong đó việc chặn một đường dẫn cửa sau sẽ bỏ chặn một đường dẫn cửa sau khác



(a) Hình dung dòng chảy của mối liên hệ gây nhiễu và mối liên hệ nhân quả.

(b) Hình dung sự cô lập của mối liên hệ nhân quả chảy từ đến con cái của nó, cho phép những điều vô căn cứ tiêu chí trẻ em để ngũ ý khả năng nhận dạng.

Hình 6.12: Đồ thị ví dụ thỏa mãn tiêu chí trẻ em không bị nhầm lẫn

bị chặn lại bởi máy va chạm 2. Và chúng ta có thể chặn đường dẫn cửa sau mở khóa đường dẫn cửa sau khác trong đó 2 là máy va chạm. Có thể chặn cả hai đường dẫn riêng lẻ không có nghĩa là chúng ta có thể chặn cả hai bằng một bộ điều hòa duy nhất. Tóm lại, tiêu chí trẻ em không bị nhầm lẫn là đủ nhưng không cần thiết, và điều kiện liên quan này là cần thiết nhưng không đủ. Ngoài ra, mọi thứ chúng ta đã thấy trong phần này cho đến nay đều dành cho can thiệp biến đơn.

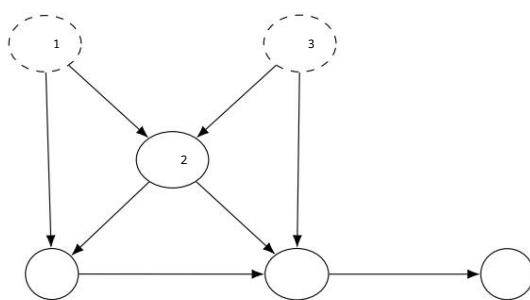
Điều kiện cần và đủ cho các can thiệp đa biến Shpitser và Pearl [25] cung cấp tiêu chí cần và đủ để nhận dạng  $(\text{ } = \text{ } | \text{ do}(\text{ } = \text{ }))$  khi và là tùy ý tập hợp các biến: tiêu chí phòng ngừa. Tuy nhiên, điều này nằm ngoài phạm vi của cuốn sách này, vì nó đòi hỏi những đồ vật phức tạp hơn như hàng rào, cây chũ C và những đồ vật có lá khác. Tiến xa hơn nữa, Shpitser và Pearl [28] cung cấp một tiêu chí cần và đủ cho một cách tổng quát nhất loại ước tính nhân quả: tác động nhân quả có điều kiện, có dạng  $(\text{ } = \text{ } | \text{ do}(\text{ } = \text{ }), \text{ } = \text{ })$ , trong đó , , và đều là các tập hợp tùy ý của biến.

[25]: Shpitser và Pearl (2006), 'Xác định các phân phối can thiệp chung trong Mod nguyên nhân bán Markovian độ quyels'

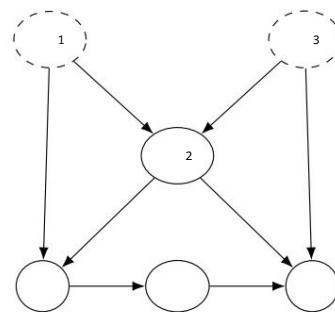
[28]: Shpitser và Pearl (2006), 'Xác định phân phối can thiệp có điều kiện'

#### Bài tập đọc tích cực:

1. Tiêu chí vô căn cứ (Định nghĩa 6.2) có thỏa mãn trong Hình-ure 6.13a?
2. Tiêu chí vô căn cứ có thỏa mãn ở Hình 6.13b không?
3. Liệu chúng ta có thể nhận dạng được trong Hình 6.13b bằng tiêu chí nào đơn giản hơn không? mà chúng ta đã thấy trước đây?



(a)



(b)

Hình 6.13: Biểu đồ cho các câu hỏi về tiêu chí trẻ em không bị nhầm lẫn

## 7

## Ước lượng

Trong chương trước, chúng ta đã đề cập đến việc nhận dạng. Một khi chúng tôi xác định được một số ước tính nhân quả bằng cách giảm nó thành ước tính thống kê, chúng ta vẫn có nhiều công việc phải làm. Chúng ta cần có được ước tính tương ứng. Trong chương này, chúng ta sẽ bao gồm nhiều công cụ ước tính khác nhau mà chúng ta có thể sử dụng để thực hiện việc này. Điều này không có nghĩa là gần như đầy đủ vì có nhiều công cụ ước tính khác nhau những tác động nhân quả, nhưng nó nhằm mục đích cung cấp cho bạn một sự giới thiệu chắc chắn về chúng. Tất cả các công cụ ước tính mà chúng tôi đưa vào các phần đầy đủ đều được hỗ trợ bởi mô hình công cụ ước tính (nhớ lại từ Phần 2.4). Và tất cả đều hoạt động tùy ý các mô hình thống kê chẳng hạn như những mô hình bạn có thể nhận được từ scikit-learn [29].

## 7.1 Vòng sơ loại

Hãy nhớ lại ở Chương 2 rằng chúng ta biểu thị hiệu quả can thiệp riêng lẻ (ITE) với và hiệu quả điều trị trung bình (ATE) với :

$$(1) \quad (0) \quad (7.1)$$

$$[ \quad (1) \quad (0) ] \quad (7.2)$$

ITE là loại tác động nhân quả cụ thể nhất, nhưng chúng khó để ước tính mà không có giả định chắc chắn (trên những giả định được thảo luận trong Chương 2 và 4). Tuy nhiên, chúng ta thường muốn ước tính những tác động nhân quả mà được cá nhân hóa hơn một chút so với ATE.

Ví dụ: giả sử chúng ta đã quan sát hiệp phương sai của một cá nhân ; Chúng ta có thể muốn sử dụng chúng để ước tính tác động cụ thể hơn cho cá nhân đó (và bất kỳ ai khác có đồng biến ). Điều này đưa chúng ta đến mức trung bình có điều kiện hiệu quả điều trị (CATE) ( ):

$$( ) \quad [ \quad (1) \quad (0) | = ] \quad (7.3)$$

được điều kiện hóa không cần phải bao gồm tất cả các điều được quan sát đồng biến, nhưng điều này thường xảy ra khi mọi người đề cập đến CATE. Chúng tôi gọi hiệu quả điều trị trung bình cá nhân hóa (IATE).

ITE và "CATE" (chúng tôi gọi là IATE) đôi khi được kết hợp với nhau, nhưng chúng không giống nhau. Ví dụ, hai cá nhân có thể có cùng đồng biến, nhưng kết quả tiềm năng của chúng có thể khác nhau vì những khác biệt không được quan sát khác giữa những cá nhân này. Nếu chúng ta bao gồm mọi thứ về một cá nhân có liên quan đến kết quả tiềm năng của họ trong vectơ , thì ITE và "CATE" giống nhau nếu  $=$ . Trong một nguyên nhân đồ thị, tương ứng với tất cả các biến ngoại sinh trong đồ thị phỏng đại đồ thị có mối liên hệ nhân quả dẫn tới .

7.1 Sơ bộ . . . . .	62
7.2 Mô hình kết quả có điều kiện-điện tử (COM) . . . . .	63
7.3 Nhóm Out có điều kiện-đến với Mô hình hóa (GCOM) .	64
7.4 Tăng hiệu quả dữ liệu . . . . .	65
TARNET . . . . .	65
Người học X. . . . .	66
7.5 Điểm Xu hướng . . . . .	67
7.6 Trọng số theo xác suất nghịch đảo (IPW) . . . . .	68
7.7 Phương pháp mạnh mẽ gấp đôi . . . . .	70
7.8 Các phương pháp khác . . . . .	70
7.9 Nhận xét kết luận . . . . .	71
Khoảng tin cậy . . . . .	71
So sánh với ngẫu nhiên	
Thí nghiệm . . . . .	72

[29]: Pedregosa và cộng sự. (2011), 'Scikit-tìm hiểu: Học máy bằng Python'

<sup>1</sup> Đoạn văn này chưa rất nhiều thông tin .

Bài tập đọc tích cực:

1) Hãy thuyết phục bản thân rằng ITE và "CATE" (cái mà chúng tôi gọi là IATE) đều giống nhau nếu  $=$  .

2) Hãy thuyết phục bản thân rằng tương ứng với các biến ngoại sinh trong phỏng đại biểu đồ có mối liên hệ nhân quả trời chay tôi .

Tính vô căn cứ Trong suốt chương này, bất cứ khi nào chúng ta ước tính một ATE, chúng ta sẽ giả định rằng là một tập hợp điều chỉnh đủ, và bất cứ khi nào chúng ta ước tính một CATE, chúng ta sẽ giả định rằng là một tập hợp điều chỉnh đủ. Nói cách khác, để ước tính ATE, chúng ta giả sử rằng thỏa mãn tiêu chí cửa sau (Định nghĩa 4.1); tương tự với ước lượng ATE, chúng ta giả định rằng chúng ta có khả năng trao đổi có điều kiện cho trước (Giả định 2.2). Và tương tự đối với ước tính CATE, giả sử là tập điều chỉnh đủ có nghĩa là chúng ta đang giả định rằng thỏa mãn tiêu chí cửa sau / mang lại cho chúng ta sự vô căn cứ.

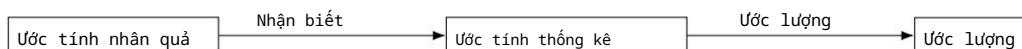
Giả định vô căn cứ này cho chúng ta nhận dạng tham số<sup>2</sup> và cho phép chúng ta tập trung vào ước lượng trong chương này.

## 7.2 Mô hình hóa kết quả có điều kiện (COM)

Chúng tôi quan tâm đến việc ước tính ATE . Chúng ta sẽ bắt đầu bằng việc nhắc lại công thức điều chỉnh (Định lý 2.1), công thức này có thể được suy ra như một hệ quả tất yếu của điều chỉnh cửa sau (Định lý 4.2), như chúng ta đã thấy trong Phần 4.4.1:

$$[(1, \dots, 0)] = [(\dots | = 1, \dots)] - [(\dots | = 0, \dots)] \quad (7.4)$$

Ở vé trái của phương trình 7.4, chúng ta có ước tính nhân quả, và ở vé phải, chúng ta có ước tính thống kê (tức là chúng ta đã xác định được đại lượng nhân quả này). Sau đó, bước tiếp theo trong Sơ đồ nhận dạng-ước tính (xem Hình 7.1 được sao chép từ Phần 2.4) là lấy ước tính về ước tính (thống kê) này.



Hình 7.1: Sơ đồ nhận dạng-Ước tính - sơ đồ minh họa quá trình chuyển từ ước tính nhân quả mục tiêu sang ước tính tương ứng, thông qua nhận dạng và ước tính.

Điều đơn giản nhất cần làm là điều chỉnh mô hình thống kê (mô hình học máy) phù hợp với kỳ vọng có điều kiện  $[(\dots | = 1, \dots)]$  rồi ước chứng với giá trị trung bình thực nghiệm trên điểm dữ liệu ( $\dots$ ). Và đây chính xác là những gì chúng ta đã làm trong các ví dụ ước tính đơn giản ở Phần 2.5 và 4.6.2. Để làm rõ hơn điều này, chúng tôi giới thiệu thay cho kỳ vọng có điều kiện này:

$$(1, \dots, 0, \dots) - [(\dots | = 1, \dots)] + [(\dots | = 0, \dots)] = 0 \quad (7.5)$$

Sau đó, chúng ta có thể điều chỉnh mô hình thống kê phù hợp với . Chúng tôi sẽ biểu thị rằng những mô hình được trang bị này gần đúng với có mũ:  $\hat{\dots}$ . Chúng ta sẽ gọi mô hình  $\hat{\dots}$  là mô hình kết quả có điều kiện. Bây giờ, chúng ta có thể viết rõ ràng công cụ ước tính được mô hình hỗ trợ (cho ATE) mà chúng ta đã mô tả:

$$\hat{y} = \frac{1}{(1, \dots, 0, \dots)} \quad (7.6)$$

Chúng tôi sẽ đề cập đến các công cụ ước tính có dạng này là công cụ ước tính mô hình kết quả có điều kiện (COM). Bởi vì việc giảm thiểu sai số bình phương trung bình (MSE) của việc dự đoán từ các cặp ( $\dots, \dots$ ) tương đương với việc lập mô hình kỳ vọng có điều kiện này [xem, ví dụ, 10, Phần 2.4], nên chúng tôi có nhiều mô hình khác nhau.

**Bài tập đọc chủ động:** Hai phép tính gần đúng khác nhau mà chúng ta thực hiện trong công cụ ước tính này là gì và mỗi ước lượng thống kê trong Công thức 7.4 thay thế những phần nào của ước tính thống kê?

[10]: Hastie và cộng sự. (2001), Các yếu tố của việc học thống kê

có thể sử dụng `for` ^ trong Công thức 7.6 để lấy công cụ ước tính COM (ví dụ, xem scikit-learn [29]).

Đối với ước tính CATE, vì chúng tôi giả định rằng là bộ điều chỉnh đủ, thay vì chỉ ^, nên chúng tôi phải thêm ^3 làm đầu vào cho mô hình kết quả có điều kiện của mình. Chính xác hơn, để ước tính CATE, chúng tôi xác định như sau:

$$(7.7) \quad \hat{y}(x, \cdot) = \hat{y}(1, \cdot) - \hat{y}(0, \cdot)$$

Sau đó, chúng tôi huấn luyện mô hình thống kê ^ để dự đoán từ ( , , ). Và điều này cung cấp cho chúng ta công cụ ước tính COM sau cho CATE ( ):

$$(7.8) \quad \hat{y}(x, \cdot) = \frac{1}{\hat{y}(1, \cdot) - \hat{y}(0, \cdot)}$$

trong đó ^ là số điểm dữ liệu có = . Khi chúng ta quan tâm đến IATE (CATE trong đó ^ là tất cả các biến số được quan sát), thường là 1, điều này giúp đơn giản hóa công cụ ước tính của chúng ta thành sự khác biệt đơn giản giữa các dự đoán:

$$(7.9) \quad \hat{y}(x, \cdot) = \hat{y}(1, \cdot) - \hat{y}(0, \cdot)$$

Mặc dù IATE khác với ITE ( ( ) ), nếu chúng ta thực sự muốn đưa ra ước tính cho ITE, thì việc lấy công cụ ước tính này làm công cụ ước tính của chúng ta vềITE cũng là điều tương đối phổ biến:

$$(7.10) \quad \hat{y}(x, \cdot) = \hat{y}(1, \cdot) - \hat{y}(0, \cdot)$$

Tuy nhiên, điều này có thể sẽ không đáng tin cậy do vi phạm tích cực nghiêm trọng.<sup>4</sup>

Công cụ ước tính COM có nhiều mặt Công cụ ước tính COM có nhiều tên khác nhau trong tài liệu. Ví dụ, chúng thường được gọi là công cụ ước tính tinh toán G, công thức G tham số hoặc tiêu chuẩn hóa trong dịch tễ học và thống kê sinh học. Bởi vì chúng tôi đang điều chỉnh một mô hình thống kê duy nhất cho ở đây, nên "công cụ ước tính COM" đôi khi được gọi là "S-learner", trong đó "S" là viết tắt của "single".

### 7.3 Mô hình hóa kết quả có điều kiện được nhóm (GCOM)

Để có được ước tính trong phương trình 7.6, chúng ta phải huấn luyện một mô hình dự đoán từ ( , ). Tuy nhiên, thường là một chiều, trong khi có thể là nhiều chiều. Nhưng đầu vào ^ cho là thứ duy nhất thay đổi giữa hai số hạng bên trong tổng ^(1, ) + ^(0, ) . Hãy tưởng tượng ghép nối với một vectơ 100 chiều và sau đó truyền dữ liệu đó qua mạng thần kinh mà chúng ta đang sử dụng cho ^ . Có vẻ hợp lý khi mạng có thể bỏ qua trong khi tập trung vào 100 chiều khác của đầu vào. Điều này sẽ dẫn đến ước tính ATE bằng 0. Và, thực sự, có một số bằng chứng về việc các công cụ ước tính COM bị sai lệch về 0 [30].

Vậy làm thế nào chúng ta có thể đảm bảo rằng mô hình ^ không bỏ qua ? Chắc chắn chúng ta chỉ có thể huấn luyện hai mô hình khác nhau ^ 1( ) và ^ 0( ) mô hình đó 1( ) và

[29]: Pedregosa và cộng sự. (2011), 'Scikit-learn: Học máy bằng Python'

<sup>3</sup> [Bài tập đọc tích cực](#): Tại sao chúng ta thêm vào tập điều chỉnh khi quan tâm đến CATE?

[Bài tập đọc tích cực](#): Viết ra ước lượng nhân quả và ước lượng thống kê dẫn chúng ta tới công cụ ước lượng trong phương trình 7.8, và chứng minh rằng chúng bằng nhau dưới điều kiện vô căn cứ và đương. Nói cách khác, xác định CATE.

<sup>4</sup> [Bài tập đọc tích cực](#): Tại sao lại có sự vi phạm tích cực nghiêm trọng ở đây? Điều này chỉ áp dụng trong Công thức 7.10 hay cả Công thức 7.9? Điều gì sẽ xảy ra nếu có nhiều đơn vị với = ?

[30]: Kunzel và cộng sự. (2019), 'Metalearners để ước tính hiệu quả điều trị không đồng nhất bằng cách sử dụng máy học'

$\theta(\cdot)$ , tương ứng, trong đó

$$\hat{1}(\cdot) = [ \cdot | \cdot ] = 1, \quad = [ \cdot | \cdot ] \text{ và } \hat{\theta}(\cdot) = [ \cdot | \cdot ] = \theta, \quad = [ \cdot | \cdot ]. \quad (7.11)$$

Việc sử dụng hai mô hình riêng biệt cho các giá trị xử lý đảm bảo rằng không thể bị bỏ qua. Để huấn luyện các mô hình thống kê này, trước tiên chúng tôi nhóm dữ liệu thành một nhóm trong đó  $= 1$  và một nhóm trong đó  $= 0$ . Sau đó, chúng tôi huấn luyện  $\hat{1}(\cdot)$  để dự đoán từ trong nhóm trong đó  $= 1$ . Và, tương tự, chúng ta huấn luyện  $\hat{\theta}(\cdot)$  để dự đoán từ trong nhóm trong đó  $= 0$ . Điều này mang lại cho chúng ta đạo hàm tự nhiên của các công cụ ước tính COM (Phương trình 7.6), các công cụ ước tính mô hình kết quả có điều kiện (GCOM) được nhóm :

$$\hat{1} = \frac{1}{\hat{1}(\cdot) + \hat{\theta}(\cdot)} \quad (7.12)$$

Và như chúng ta đã thấy, trong Phương trình 7.8, chúng ta có thể thêm làm đầu vào cho **Và**  $\hat{1} = \hat{\theta}$  để có được công cụ ước tính GCOM cho CATE ( $\cdot$ ):

$$\hat{1}(\cdot) = \frac{1}{\hat{1}(\cdot) + \hat{\theta}(\cdot)} \quad (7.13)$$

Mặc dù ước tính GCOM dường như khắc phục được vấn đề mà ước tính COM có thể gặp phải liên quan đến sai lệch về hiệu quả điều trị bằng 0, nhưng nó có một nhược điểm quan trọng. Trong ước tính COM, chúng tôi có thể sử dụng tất cả dữ liệu khi ước tính mô hình đơn  $\hat{1}$ . Tuy nhiên, trong ước tính mô hình kết quả có điều kiện được nhóm, chúng tôi chỉ sử dụng nhóm  $= 1$  để ước tính  $\hat{1}$  và chúng tôi chỉ sử dụng nhóm  $= 0$  để ước tính  $\hat{\theta}$ . Điều quan trọng là chúng tôi đang bỏ lỡ việc tận dụng tối đa dữ liệu của mình bằng cách không sử dụng tất cả dữ liệu để ước tính  $\hat{1}$  và tất cả dữ liệu để ước tính  $\hat{\theta}$ .

## 7.4 Tăng hiệu quả dữ liệu

Trong phần này, chúng tôi sẽ đề cập đến hai cách để giải quyết vấn đề về hiệu quả dữ liệu mà chúng tôi đã đề cập có trong ước tính GCOM ở cuối phần trước: TARNet (Phần 7.4.1) và X-Learner (Phần 7.4.2).

### 7.4.1 TARNet

Hãy xem xét rằng chúng tôi đang sử dụng mạng lưới thần kinh cho các mô hình thống kê của mình; bắt đầu từ đó, chúng ta sẽ so sánh ước tính COM chuẩn, ước tính GCOM và TARNet. Trong ước tính COM chuẩn, mạng nơ-ron được sử dụng để dự đoán từ ( $\cdot, \cdot$ ) (xem Hình 7.2a). Điều này có vấn đề về khả năng mang lại ước tính ATE thiên về 0, vì mạng có thể bỏ qua đại lượng vô hướng  $\cdot$ , đặc biệt khi  $\cdot$  có nhiều chiều. Chúng tôi đảm bảo rằng không thể bị bỏ qua trong ước tính GCOM bằng cách sử dụng hai mạng nơ-ron riêng biệt cho hai nhóm điều trị (Hình 7.2b). Tuy nhiên, điều này không hiệu quả vì chúng tôi chỉ sử dụng dữ liệu nhóm xử lý để huấn luyện một mạng và dữ liệu nhóm kiểm soát để huấn luyện mạng kia.

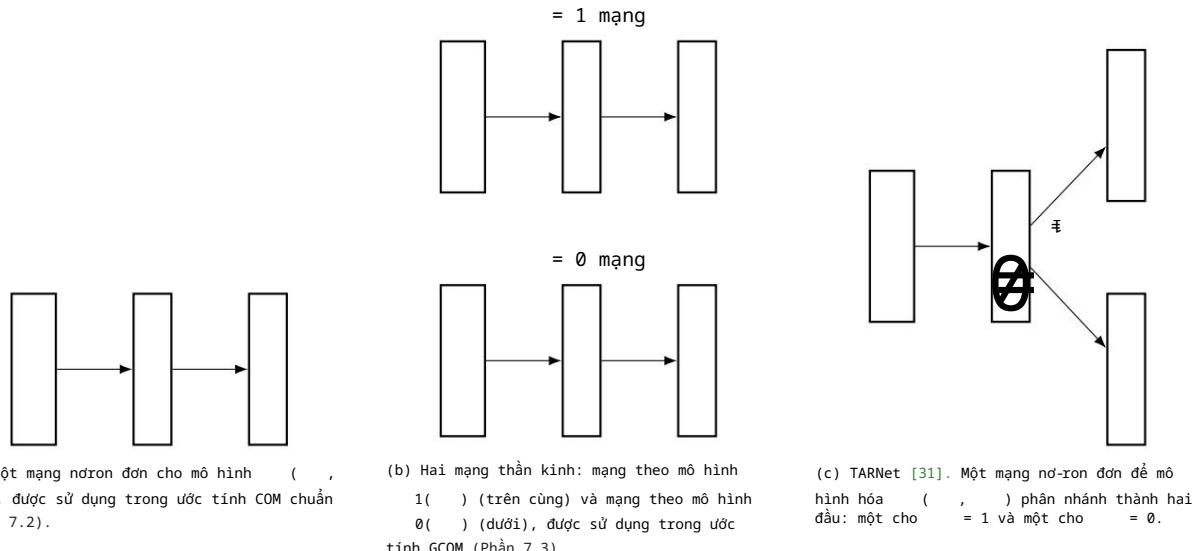
Chúng ta có thể đạt được điểm trung gian giữa ước tính COM vanilla và ước tính GCOM bằng cách sử dụng Shalit et al. TARNet của [31]. Với TARNet, chúng tôi sử dụng một mạng duy nhất chỉ lấy làm đầu vào nhưng sau đó phân nhánh thành

<sup>5</sup> Künzel và cộng sự. [30] gọi công cụ ước tính GCOM là "T-learner" trong đó "T" là "hai" vì nó yêu cầu lắp hai mô hình khác nhau:  $\hat{1}$  và  $\hat{\theta}$ .

[31]: Shalit và cộng sự. (2017), 'Ước tính hiệu quả điều trị trên từng cá nhân: giới hạn khái quát và thuật toán'

hai đầu (mạng lưới con) riêng biệt cho mỗi nhóm điều trị. Sau đó, chúng tôi sử dụng mô hình này cho  $(\cdot, \cdot)$  để lấy công cụ ước tính COM. Điều này có ưu điểm là học cách biểu diễn bắt khả tri về điều trị (TAR) của bằng cách sử dụng tất cả dữ liệu trong khi vẫn buộc mô hình không bỏ qua bằng cách phân nhánh thành hai đầu cho các giá trị khác nhau của  $\cdot$ .

Nói cách khác, TARNet sử dụng kiến thức chúng ta có về  $\cdot$  (như một biến quan trọng duy nhất) trong kiến trúc của nó. Tuy nhiên, các mạng con cho mỗi đầu này chỉ được đào tạo bằng dữ liệu cho nhóm điều trị tương ứng chứ không phải tất cả dữ liệu.<sup>6</sup>



Hình 7.2: Kiến trúc mạng thần kinh thô cho ước tính COM chuẩn (trái), ước tính GCOM (giữa) và TARNet (phải). Trong hình này, chúng tôi sử dụng mỗi mũi tên để biểu thị một mạng con có số lớp tùy ý.

#### 7.4.2 Người học X

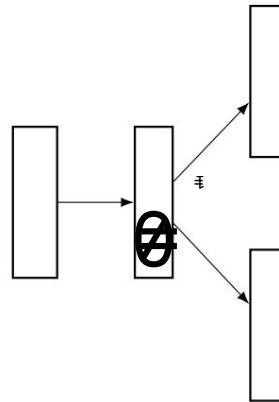
Chúng ta vừa thấy rằng một cách để tăng hiệu quả dữ liệu so với ước tính GCOM là sử dụng TARNet, một công cụ ước tính COM có chung một số đặc tính với các công cụ ước tính GCOM. Tuy nhiên, TARNet vẫn không sử dụng tất cả dữ liệu cho mô hình đầy đủ (mạng thần kinh). Trong phần này, chúng ta sẽ bắt đầu với ước tính GCOM và dựa trên đó để tạo ra một lớp các công cụ ước tính sử dụng tất cả dữ liệu cho cả hai mô hình là một phần của công cụ ước tính. Công cụ ước tính trong lớp này được gọi là X-learner [30]. Không giống như TARNet, X-learner không phải là công cụ ước tính COM hay công cụ ước tính GCOM.

Có ba bước đối với X-learning và bước đầu tiên giống hệt như bước được sử dụng trong ước tính GCOM: ước tính  $\hat{Y}^1(\cdot)$  bằng cách sử dụng dữ liệu nhóm can thiệp và ước tính  $\hat{Y}^0(\cdot)$  bằng cách sử dụng dữ liệu nhóm kiểm soát.<sup>7</sup> As trước đây, điều này có thể được thực hiện với bất kỳ mô hình nào giảm thiểu MSE. Để đơn giản, trong phần này, chúng ta sẽ xem xét IATE ( $\cdot$  là tất cả các biến được quan sát) trong đó  $\cdot$  thỏa mãn tiêu chí của sau ( $\cdot$  chứa  $\cdot$  và không có hậu duệ của  $\cdot$ ).

Bước thứ hai là phần quan trọng nhất vì đây là nơi chúng ta kết thúc việc sử dụng tất cả dữ liệu cho cả hai mô hình và nơi xuất phát của chữ "X". Chúng tôi chỉ định  $\hat{Y}^1(\cdot)$  cho ước tính ITE của nhóm điều trị và  $\hat{Y}^0(\cdot)$  cho nhóm điều trị

<sup>6</sup> **Bài tập đọc tích cực:** Phần nào của TARNet giống Hình 7.2a và phần nào giống Hình 7.2b? Ưu điểm/nhược điểm Hình 7.2a

đến 7.2c có mối liên hệ gì với nhau không?



[30]: Künzel và cộng sự. (2019), 'Metalearners để ước tính hiệu quả điều trị không đồng nhất bằng cách sử dụng máy học'

<sup>7</sup> Hãy nhớ rằng  $\hat{Y}^1(\cdot)$  và  $\hat{Y}^0(\cdot)$  là xấp xỉ của  $[Y^1 | Y^0] = 1, Y^1 = 1]$  và  $[Y^1 | Y^0] = 0, Y^0 = 0]$  tương ứng.

Ước tính nhóm ITE:

$$\hat{1}, = (1) \hat{0}( ) \quad (7.14)$$

$$\hat{0}, = \hat{1}( ) \hat{0} \quad (7.15)$$

Ở đây,  $\hat{1}$ , được ước tính bằng cách sử dụng kết quả của nhóm can thiệp và phản thực giả định mà chúng ta nhận được từ  $\hat{0}$ . Tương tự,  $\hat{0}$ , được ước tính bằng cách sử dụng kết quả của nhóm đối chứng và phản thực giả định mà chúng ta nhận được từ  $\hat{1}$ . Nếu bạn vẽ một đường thẳng giữa các kết quả tiềm năng được quan sát và một đường thẳng giữa các kết quả tiềm năng được quy ước, bạn có thể thấy hình chữ "X". Điều quan trọng là, chữ "X" này cho chúng ta biết rằng mỗi ước tính ITE của mỗi nhóm điều trị  $\hat{1}$ , sử dụng cả dữ liệu của nhóm điều trị (kết quả tiềm năng quan sát được trong quá trình điều trị) và dữ liệu của nhóm đối chứng (trong  $\hat{0}$ ). Tương tự,  $\hat{0}$ , được ước tính bằng dữ liệu từ cả hai nhóm can thiệp.

Tuy nhiên, mỗi ước tính ITE chỉ sử dụng một điểm dữ liệu duy nhất từ nhóm xử lý tương ứng. Chúng ta có thể khắc phục điều này bằng cách điều chỉnh mô hình  $\hat{1}( )$  để dự đoán  $\hat{1}$ , từ nhóm xử lý tương ứng 's. Cuối cùng, chúng ta có mô hình  $\hat{1}( )$  phù hợp khi sử dụng tất cả dữ liệu (dữ liệu nhóm điều trị vừa rồi và dữ liệu nhóm kiểm soát khi  $\hat{0}$  phù hợp ở bước 1). Tương tự, chúng ta có thể điều chỉnh mô hình  $\hat{0}( )$  để dự đoán  $\hat{0}$ , từ nhóm kiểm soát tương ứng 's. Đầu ra của bước 2 là hai công cụ ước tính khác nhau cho IATE:  $\hat{1}( )$  và  $\hat{0}( )$ .

Cuối cùng, ở bước 3, chúng ta kết hợp  $\hat{1}( )$  và  $\hat{0}( )$  với nhau để có được IATE công cụ ước tính:

$$\hat{( )} = ( ) \hat{0}( ) + (1) \quad (7.16)$$

$\hat{( )} = ( )$  trong đó  $( )$  là một hàm trọng số nào đó tạo ra các giá trị từ 0 đến 1. Künzel et al. [30] báo cáo rằng việc ước tính điểm xu hướng (được giới thiệu trong phần tiếp theo) có hiệu quả, nhưng việc chọn hàm hằng số 0 hoặc 1 cũng có ý nghĩa nếu các nhóm can thiệp có quy mô rất khác nhau. Hoặc việc chọn  $( )$  để giảm thiểu phuong sai của  $\hat{( )}$  cũng có thể hấp dẫn.

[30]: Künzel và cộng sự. (2019), 'Metalearners để ước tính hiệu quả điều trị không đồng nhất bằng cách sử dụng máy học'

Bài tập đọc tích cực: Trong phần này, chúng ta đã đề cập đến X-learner để ước tính IATE. Người học X sẽ được gì hơn? Ước tính CATE chung ( $=$  là tùy ý và không nhất thiết phải chứa tất cả các yếu tố gây nhiễu) như thế nào?

## 7.5 Điểm xu hướng

Cho rằng vectơ của các biến thỏa mãn tiêu chí sau (hoặc, tương đương,  $( = 1, = 0) | = 1$ ), chúng ta có thể tự hỏi liệu có thực sự cần thiết phải đặt điều kiện vào toàn bộ vectơ đó để cô lập mối liên hệ nhân quả, đặc biệt khi có nhiều chiều. Hóa ra không phải vậy. Nếu thỏa mãn tính vô căn và tính dương, thì thực tế chúng ta có thể thoát khỏi chỉ với điều kiện hóa trên vô hướng ( $= 1 | = 1$ ). Chúng ta sẽ để  $( )$  biểu thị ( $= 1 | = 1$ ), vì chúng ta sẽ coi  $( )$  là điểm xu hướng vì đây là xu hướng (xác suất) nhận được điều trị vì  $= 1$ . Điều kỳ diệu của việc có thể điều kiện hóa đại lượng vô hướng ( $= 1$ ) thay cho vectơ  $= 1$  là do định lý điểm xu hướng của Rosenbaum và Rubin [32] :

[32]: Rosenbaum và Rubin (1983), 'Vai trò trung tâm của điểm xu hướng trong các nghiên cứu quan sát về tác động nhân quả'

Định lý 7.1 (Định lý Điểm Xu hướng) Cho trước tính tích cực, tính không nhiễu cho trước, ngũ ý tính không vô căn cứ với điểm xu hướng ( $= 1$ ).

Tương đương,

$$( -1, \quad 0) \quad | \quad = \quad ( -1, \quad 0) \quad | \quad ( ) . \quad (7.17)$$

Chúng tôi cung cấp một chứng minh toán học truyền thống hơn ở Phụ lục A.2 và chúng minh bằng đồ thị ở đây. Hãy xem xét biểu đồ trong Hình 7.3. Bởi vì cạnh từ đến là biểu tượng cho cơ chế  $( \quad | \quad )$  và vì điểm xu hướng mô tả hoàn toàn sự phân bố đó  $( \quad = 1 | \quad ) = ( \quad )$ , nên chúng ta có thể nghĩ về xu hướng cho điểm như một yếu tố trung gian đầy đủ về tác động của lên . Điều này có nghĩa là chúng ta có thể vẽ lại cái này

đồ thị có  $( \quad )$  nằm giữa và . Và trong biểu đồ được vẽ lại ở Hình 7.4 này, chúng ta có thể thấy rằng  $( \quad )$  chặn tất cả các đường dẫn cửa sau mà chặn, vì vậy  $( \quad )$  phải là một bộ điều chỉnh đủ nếu là như vậy. Do đó, chúng ta có một chứng minh đồ họa về định lý điểm xu hướng bằng cách sử dụng điều chỉnh cửa sau (Định lý 4.2).

Điều quan trọng là định lý này có nghĩa là chúng ta có thể hoán đổi  $( \quad )$  thay cho bất cứ nơi nào chúng ta điều chỉnh cho trong một công cụ ước tính nhất định trong chương này. Ví dụ, điều này có vẻ rất hữu ích khi có nhiều chiều.

Hãy nhớ lại Sự đánh đổi giữa Tích cực và Vô căn cứ ở Phần 2.3.4. Khi chúng tôi dựa vào nhiều biến số không gây ra sai lệch và chậm hơn, chúng tôi sẽ giảm thiểu nhiễu. Tuy nhiên, điều này phải trả giá bằng việc giảm sự chồng chéo vì trong  $( \quad = 1 | \quad )$  ngày càng có nhiều chiều hơn. Điểm xu hướng dường như cho phép chúng tôi khắc phục vấn đề đó một cách kỳ diệu vì  $( \quad )$  vẫn là đại lượng vô hướng, ngay cả khi phát triển về chiều. Tuyệt vời phải không?

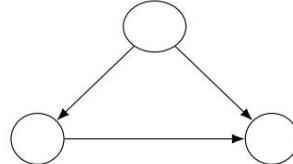
Thật không may, chúng ta thường không có quyền truy cập vào  $( \quad )$ . Đúng hơn, điều tốt nhất chúng ta có thể làm là mô hình hóa nó. Chúng tôi thực hiện điều này bằng cách đào tạo một mô hình để dự đoán từ . Ví dụ, hồi quy logistic (mô hình logit) được sử dụng xác phỏ biến để thực hiện việc này. Và bởi vì mô hình này phù hợp với nhiều chiều, nên theo một nghĩa nào đó, chúng ta vừa chuyển bài toán dương sang mô hình của chúng ta cho  $( \quad )$ .

## 7.6 Trọng số xác suất nghịch đảo (IPW)

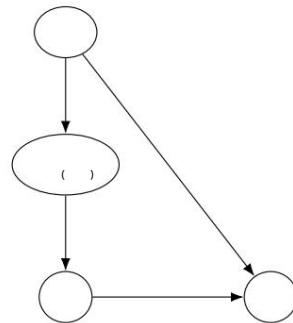
Điều gì sẽ xảy ra nếu chúng ta có thể lấy mẫu lại dữ liệu theo cách khiến nó trở thành mối liên hệ nhân quả? Đây là động lực đằng sau việc tạo ra “quần thể giả” được tạo thành từ các phiên bản được điều chỉnh lại trọng số của quần thể được quan sát. Để hiểu được điều này, chúng ta hãy nhớ lại tại sao sự liên tưởng nói chung không phải là quan hệ nhân quả.

Sự liên kết không phải là quan hệ nhân quả trong biểu đồ ở Hình 7.5 vì là nguyên nhân chung của và . Nói cách khác, cơ chế tạo ra phụ thuộc vào , và cơ chế tạo ra phụ thuộc vào . Tập trung vào cơ chế tạo ra , chúng ta có thể viết biểu thức này dưới dạng toán học là  $( \quad | \quad ) \quad ( \quad )$ . Hóa ra là chúng ta có thể cân nhắc lại dữ liệu để có được một quần thể giả trong đó  $( \quad | \quad ) = ( \quad )$  hoặc  $( \quad | \quad )$  bằng một hằng số nào đó; phần quan trọng là chúng ta làm cho độc lập với . Đồ thị tương ứng cho một quần thể giả như vậy không có cạnh từ đến vì không phụ thuộc vào ; chúng tôi mô tả điều này trong Hình 7.6.

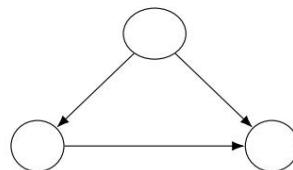
Hóa ra điểm xu hướng là chìa khóa cho việc tái cân nhắc này. Tất cả những gì chúng ta phải làm là cân nhắc lại từng điểm dữ liệu bằng cách xử lý và các yếu tố gây nhiễu



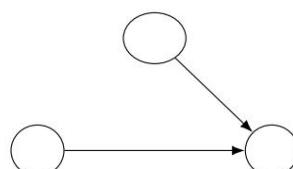
Hình 7.3: Đồ thị đơn giản trong đó thỏa mãn tiêu chí cửa sau



Hình 7.4: Biểu đồ minh họa rằng  $( \quad )$  chặn (các) đường dẫn cửa sau mà chặn.



Hình 7.5: Đồ thị đơn giản trong đó làm ảnh hưởng đến trên



Hình 7.6: Biểu đồ hiệu quả cho quần thể giả mà chúng ta có được bằng cách tính lại trọng số dữ liệu được tạo theo biểu đồ trong Hình 7.5 bằng cách sử dụng trọng số xác suất nghịch đảo .

bởi xác suất nghịch đảo của nó để nhận được giá trị xử lý khi nó có giá trị . Đây là lý do tại sao kỹ thuật này được gọi là trọng số xác suất nghịch đảo (IPW). Đối với những cá nhân được điều trị 1, trọng số này là  $\frac{1}{1 - \hat{\pi}}$  và đối với những cá nhân được điều trị 0, trọng số này là  $\frac{1}{\hat{\pi}}$ , số sẽ là  $(\cdot)$ . Nếu  $\hat{\pi}$  là xác suất điều trị liên tục, thì trọng số xu hướng

**Bài tập đọc tích cực:** Tại sao mẫu số là 1  $(\cdot)$  khi  $\hat{\pi} = 0$ . Gợi ý: nhớ lại định nghĩa chính xác của  $(\cdot)$ .

ghi điểm để điều trị liên tục.

Tại sao những gì chúng tôi mô tả trong đoạn trên lại có tác dụng? Vâng, hãy nhớ lại rằng mục tiêu của chúng ta là loại bỏ nhiều bằng cách “loại bỏ” cạnh đi từ đến (tức là chuyển từ Hình 7.5 sang Hình 7.6). Và cơ chế 1 mà cạnh mô tả là  $(\cdot | \cdot)$ . Bằng cách tính trọng số các điểm dữ liệu theo  $(\cdot | \cdot)$ , chúng tôi đang loại bỏ nó một cách hiệu quả. Đó là trực giác. Về mặt hình thức, chúng ta có phương trình nhận dạng sau:

$$[\cdot(\cdot)] = \frac{(\cdot = 1)}{(\cdot | \cdot)} \quad (7.18)$$

trong đó  $(\cdot = 1)$  là biến ngẫu nhiên chỉ báo nhận giá trị 1 nếu  $= 1$  và 0 nếu ngược lại. Chúng ta chứng minh phương trình 7.18 bằng cách sử dụng công thức điều chỉnh quen thuộc  $[\cdot(\cdot)] = [\cdot[\cdot | \cdot, \cdot]]$  (Định lý 2.1) tại Phụ lục A.3.

Giả sử xử lý nhị phân, phương trình nhận dạng sau đây cho ATE được rút ra từ Công thức 7.18:

$$[\cdot(1) - \cdot(0)] = \frac{(\cdot = 1)}{(\cdot)} - \frac{(\cdot = 0)}{1 - (\cdot)} \quad (7.19)$$

Bây giờ chúng ta đã có ước tính thống kê ở dạng IPW, chúng ta có thể có được công cụ ước tính IPW. Thay thế kỳ vọng bằng phương pháp thực nghiệm và  $(\cdot)$  bằng mô hình điểm xu hướng  $\hat{\pi}(\cdot)$ , chúng ta thu được các công thức tương đương sau đây của công cụ ước tính IPW cơ bản<sup>9</sup> cho ATE:

$$\hat{\pi} = \frac{1}{\frac{1}{1 - \hat{\pi}} + \frac{1}{\hat{\pi}}} = \frac{(\cdot = 1)}{\hat{\pi}(\cdot)} + \frac{(\cdot = 0)}{1 - \hat{\pi}(\cdot)} \quad (7.20)$$

$$= \frac{1}{\frac{1}{1 - \hat{\pi}} + \frac{1}{\hat{\pi}}} = \frac{1}{\frac{1}{1 - \hat{\pi}} + \frac{1}{\hat{\pi}}} = \frac{1}{1 - \hat{\pi}(\cdot)} \quad (7.21)$$

trong đó 1 và 0 lần lượt là số đơn vị nhóm can thiệp và đơn vị nhóm đối chứng.

Giảm bớt trọng số Như bạn có thể thấy trong các phương trình 7.20 và 7.21, nếu điểm xu hướng rất gần với 0 hoặc 1 thì các ước tính sẽ tăng vọt. Để ngăn chặn điều này, không có gì lạ khi cắt bớt điểm xu hướng nhỏ hơn đến và những điểm lớn hơn 1 đến 1 (cắt bớt các trọng số một cách hiệu quả để không lớn hơn ), mặc dù điều này dẫn đến các vấn đề riêng của nó như sự thiên vị.

Ước tính CATE Chúng ta có thể mở rộng công cụ ước tính ATE trong Công thức 7.20 để có được công cụ ước tính IPW cho CATE  $(\cdot)$  bằng cách chỉ giới hạn ở các điểm dữ liệu trong đó  $= :$

$$\hat{\pi}(\cdot) = \frac{1}{\frac{1}{1 - \hat{\pi}} + \frac{1}{\hat{\pi}}} = \frac{(\cdot = 1)}{\hat{\pi}(\cdot)} + \frac{(\cdot = 0)}{1 - \hat{\pi}(\cdot)} \quad (7.22)$$

<sup>9</sup> Công cụ ước tính này có nguồn gốc từ Horvitz và Thompson [33]

[33]: Horvitz và Thompson (1952), 'A Khái quát hóa việc lấy mẫu mà không cần thay thế từ một vũ trụ hữu hạn'

**Bài tập đọc tích cực:** Công thức tương ứng của công cụ ước tính IPW cơ bản cho  $[\cdot(\cdot)]$  là gì ?

trong đó là số điểm dữ liệu có = . Tuy nhiên, công cụ ước tính trong chương trình 7.22 có thể nhanh chóng gặp phải vấn đề sử dụng lượng dữ liệu rất nhỏ, dẫn đến phương sai cao. Ước tính CATE tổng quát hơn bằng công cụ ước tính IPW phức tạp hơn và nằm ngoài phạm vi của cuốn sách này.

Xem, ví dụ, Abrevaya et al. [34] và các tài liệu tham khảo trong đó.

[34]: Abrevaya và cộng sự. (2015), 'Ước tính hiệu quả điều trị trung bình có điều kiện'

## 7.7 Phương pháp mạnh mẽ gấp đôi

Chúng ta đã thấy rằng có thể ước tính tác động nhân quả bằng cách lập mô hình ( , ) [ | , ] (Mục 7.2 đến 7.4) hoặc bằng mô hình ( ) ( = 1 | ) (Mục 7.6). Điều gì sẽ xảy ra nếu chúng ta lập mô hình cho cả ( , ) và ( )? Chà, chúng ta có thể và các công cụ ước tính thực hiện điều này đôi khi mạnh mẽ gấp đôi. Công cụ ước tính mạnh gấp đôi có đặc tính là nó là công cụ ước tính nhất quán<sup>10</sup> của nếu ^ là công cụ ước tính nhất quán của hoặc ^ là ước tính nhất quán của . Nói cách khác, chỉ một trong ^ và ^ cần được xác định rõ ràng. Ngoài ra, tốc độ tại đó một công cụ ước lượng mạnh gấp đôi hội tụ về ^ là tích của tốc độ ^ hội tụ về và tốc độ ^ hội tụ về . Điều này làm cho độ bền gấp đôi trở nên rất hữu ích khi chúng ta sử dụng các mô hình học máy linh hoạt ở chiều cao bởi vì, trong cài đặt này, mỗi mô hình riêng lẻ ( ^ và ^ ) của chúng ta hội tụ chậm hơn tốc độ lý tưởng 1/2 của .

Tuy nhiên, có một số tranh cãi về việc các phương pháp mạnh mẽ gấp đôi hoạt động tốt như thế nào trong thực tế nếu không có ít nhất một trong số ^ hoặc ^ được chỉ định rõ ràng [35]. Tuy nhiên, điều này có thể gây tranh cãi vì chúng tôi ngày càng sử dụng tốt hơn các công cụ ước tính robust gấp đôi với các mô hình học máy linh hoạt (xem, ví dụ: [36]). Trong khi đó, các công cụ ước tính hiện có về hoạt động tốt nhất đều là mô hình linh hoạt (không giống như các công cụ ước tính IPW thuần túy) [37]. Đây là lý do tại sao chúng tôi bắt đầu chương này với các công cụ ước tính lập mô hình và dành một số phần cho các công cụ ước tính đó.

Các phương pháp mạnh mẽ gấp đôi phần lớn nằm ngoài phạm vi của cuốn sách này, vì vậy chúng tôi giới thiệu cho người đọc phần giới thiệu của Seaman và Vansteelandt [38], cùng với các tác phẩm có ảnh hưởng khác về chủ đề này: [39-41]. Ngoài ra, có rất nhiều công trình mạnh mẽ gấp đôi về các phương pháp đã hoạt động khá tốt trong các cuộc thi [37]; danh mục này được gọi là ước tính khả năng tối đa được nhắm mục tiêu (TMLE). [42-44].

## 7.8 Các phương pháp khác

Vì chương này chỉ là phần giới thiệu về ước lượng trong suy luận nhân quả nên có một số phương pháp mà chúng tôi đã hoàn toàn bỏ qua. Chúng tôi sẽ mô tả ngắn gọn một số cái phỏ biến nhất trong phần này.

So khớp Trong các phương pháp so sánh, chúng tôi cố gắng so khớp các đơn vị trong nhóm can thiệp với các đơn vị trong nhóm đối chứng và loại bỏ những đơn vị không khớp để tạo ra các nhóm có thể so sánh được. Chúng ta có thể so khớp trong không gian hiệp phương sai thô, không gian hiệp phương sai thô hoặc không gian điểm xu hướng. Có các hàm khoảng cách khác nhau để quyết định mức độ gần nhau của hai đơn vị. Hơn nữa, có nhiều tiêu chí khác nhau để quyết định xem một khoảng cách nhất định có đủ gần để được tính là trùng khớp hay không (một tiêu chí yêu cầu phải khớp chính xác), mỗi đơn vị nhóm xử lý có thể có bao nhiêu trận đấu, bao nhiêu trận đấu

<sup>10</sup>Một công cụ ước tính là nhất quán nếu nó hội tụ xác suất về ước tính của nó khi số lượng mẫu tăng lên.

[35]: Kang và Schafer (2007), 'Làm sáng tỏ tính mạnh mẽ kép: So sánh các chiến lược thay thế để ước tính một Trung bình dân số từ dữ liệu không đầy đủ'

[36]: Zavich và Breskin (2020), Học máy để suy luận nhân quả: về việc sử dụng các công cụ ước tính phù

hợp chéo [37]: Dorie et al. (2019), 'Tự động so với Các phương pháp tự làm để suy luận nhân quả ence: Bài học rút ra từ cuộc thi phân tích dữ liệu'

[38]: Seaman và Vansteelandt (2018), 'Introduzione đến Phương pháp mạnh mẽ gấp đôi cho Dữ liệu không đầy đủ'

[39]: Tsiatis (2007), Lý thuyết bán tham số và dữ liệu còn

thiếu [40]: Robins et al. (1994), 'Ước tính các hệ số hồi quy khi một số biến hồi quy không phải lúc nào cũng được quan sát' [41]: Bang và Robins (2005), 'Ước tính mạnh mẽ gấp đôi trong các mô hình suy luận nhân quả và dữ liệu bị thiếu'

[42]: Van Der Laan và Rubin (2006), 'Có mục tiêu tối đa hóa khả năng học tập'

[43]: Schuler và Rose (2017), 'Có mục tiêu Ước tính khả năng tối đa cho Suy luận nhân quả trong nghiên cứu quan sát'

[44]: Van der Laan và Rose (2011), Học tập có mục tiêu: suy luận nhân quả cho dữ liệu quan sát và thực nghiệm

mỗi đơn vị nhóm kiểm soát có thể có, v.v. Ví dụ, hãy xem Stuart [45] để biết ôn tập.

Học máy kép Trong học máy kép, chúng tôi đưa ba mô hình vào hai giai đoạn: hai mô hình ở giai đoạn đầu tiên và một mô hình cuối cùng ở giai đoạn thứ hai. Giai đoạn đầu tiên: 1.

Điều chỉnh mô hình để dự đoán từ để có được ^ dự đoán.  
11 2. Lắp mô hình để dự đoán từ để có được ^ dự đoán.

Sau đó, ở giai đoạn thứ hai, chúng ta "tách ra" bằng cách xem xét ^ và ^ . Theo một nghĩa nào đó, chúng tôi đã giải mã được tác động của việc điều trị đối với kết quả bằng việc loại bỏ một phần này. Sau đó, chúng ta điều chỉnh một mô hình để dự đoán ^ từ ^ . Điều này cho chúng ta những ước tính về tác động nhân quả. Để biết thêm về chủ đề này, hãy xem ví dụ [46-49].

Cây nhân quả và rừng Một phương pháp ước tính phổ biến khác là phân vùng để quy dữ liệu thành các tập hợp con có cùng tác dụng xử lý [50]. Điều này tạo thành một cây nhân quả trong đó các lá là tập hợp con của quần thể có tác động nhân quả tương tự. Vì các khu rừng ngẫu nhiên thường hoạt động tốt hơn cây quyết định nên sẽ thật tuyệt nếu loại chiến lược này có thể được mở rộng sang các khu rừng ngẫu nhiên. Và nó có thể. Phần mở rộng này được gọi là rừng nhân quả [51], là một phần của lớp tổng quát hơn được gọi là rừng ngẫu nhiên tổng quát [52]. Điều quan trọng là các phương pháp này được phát triển với mục tiêu mang lại khoảng tin cậy hợp lệ cho các ước tính.

[45]: Stuart (2010), 'Phương pháp so khớp cho Suy luận nhân quả: Đánh giá và nhìn Phía trước'

<sup>11</sup> Bài tập đọc tích cực: Mô hình này khác với ^ như thế nào?

[46]: Chernozhukov và cộng sự. (2018), 'Học máy Double-ble/debiased cho các thông số xử lý và cấu trúc' [47]: Felton (2018), Chernozhukov et al. về Học máy đôi / Debiased [48]: Syrgkanis (2019), Học máy trực giao/kép [49]: Foster và Syrgkanis (2019), Học thống kê trực giao [50]: Athey và Imbens (2016), 'Phân vùng để quy cho những tác động nhân quả không đồng nhất'

[51]: Wager và Athey (2018), 'Ước tính và suy luận về hiệu quả xử lý không đồng nhất bằng cách sử dụng rừng ngẫu nhiên'

[52]: Athey và cộng sự. (2019), 'Rừng ngẫu nhiên tổng quát'

## 7.9 Nhận xét kết luận

### 7.9.1 Khoảng tin cậy

Cho đến nay, trong chương này, chúng ta chỉ thảo luận về các ước tính điểm cho các tác động nhân quả. Chúng tôi chưa thảo luận về cách chúng tôi có thể đánh giá mức độ không chắc chắn của mình do lấy mẫu dữ liệu. Chúng ta chưa thảo luận về cách tính khoảng tin cậy trên những ước tính này. Xét cho cùng, đây là một quan điểm học máy ; ai quan tâm đến khoảng tin cậy... Bỏ chuyện đùa sang một bên, bởi vì chúng tôi đang cho phép các mô hình học máy tùy ý trong tất cả các công cụ ước tính mà chúng tôi thảo luận, nên thực sự khá khó để có được khoảng tin cậy hợp lệ.

Bootstrapping Một cách để có được khoảng tin cậy là sử dụng bootstrap- ping. Với bootstrapping, chúng tôi lặp lại quá trình ước tính tác động nhân quả nhiều lần, mỗi lần với một mẫu khác (có thay thế) từ dữ liệu của chúng tôi. Điều này cho phép chúng ta xây dựng một phân phối thực nghiệm cho ước tính. Sau đó, chúng ta có thể tính toán bất kỳ khoảng tin cậy nào mà chúng ta muốn từ phân bố thực nghiệm đó. Thật không may, khoảng tin cậy được khởi động không phải lúc nào cũng hợp lệ. Ví dụ: nếu chúng tôi lấy khoảng tin cậy 95% được khởi động, nó có thể không chứa giá trị thực (ước tính) trong 95% thời gian.

Các mô hình chuyên biệt Một cách khác để có được khoảng tin cậy là phân tích các mô hình rất cụ thể, thay vì cho phép các mô hình tùy ý. Mô hình tuyến tính là ví dụ đơn giản nhất về điều này; thật dễ dàng để có được khoảng tin cậy trong các mô hình tuyến tính. Tương tự, nếu chúng ta sử dụng mô hình tuyến tính làm mô hình giai đoạn thứ hai trong học máy kép, chúng ta có thể nhận được khoảng tin cậy. Đáng chú ý, cây nhân quả và rừng nhân quả được phát triển với mục tiêu đạt được khoảng tin cậy.

### 7.9.2 So sánh với các thí nghiệm ngẫu nhiên

Bạn có thể đọc ở đâu đó rằng một số kỹ thuật điều chỉnh này đảm bảo rằng chúng tôi đã giải quyết được yếu tố gây nhiễu và tách biệt tác động nhân quả.

Tất nhiên, điều này không đúng khi có sự nhiễu loạn không quan sát được. Những phương pháp này chỉ giải quyết được vấn đề gây nhiễu được quan sát. Nếu có bất kỳ yếu tố gây nhiễu nào không được quan sát, thì các phương pháp này sẽ không khắc phục được điều đó giống như phương pháp ngẫu nhiên hóa (Chương 5). Những phương pháp điều chỉnh này không phải là phép thuật. Và thật khó để biết khi nào là hợp lý để cho rằng chúng ta đã quan sát thấy tất cả các yếu tố gây nhiễu.

Đó là lý do tại sao điều quan trọng là phải tiến hành phân tích độ nhạy trong đó chúng tôi đánh giá mức độ chắc chắn của ước tính tác động nhân quả của chúng tôi đối với yếu tố gây nhiễu không quan sát được.

Đây là chủ đề của chương tiếp theo.

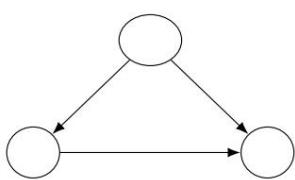
**Bài tập đọc tích cực:** Chúng ta đã sử dụng loại công cụ ước tính nào trong các ví dụ ước tính ở Phần 2.5 và 4.6.2?

# Sự gây nhiễu không được quan sát:

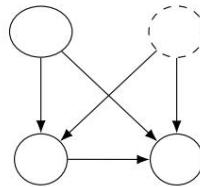
## Phân tích giới hạn và độ nhạy

số 8

Tất cả các phương pháp trong Chương 7 đều giả định rằng chúng ta không có bất kỳ dữ liệu nào không quan sát được. Tuy nhiên, tính vô căn cứ là một giả định không thể kiểm chứng được. Trong các nghiên cứu quan sát, cũng có thể có một số yếu tố gây nhiễu không được quan sát. Do đó, chúng tôi muốn biết ước tính của chúng tôi chắc chắn đến mức nào đến sự nhiễu loạn không quan sát được. Cách đầu tiên chúng ta có thể làm là lấy một giới hạn trên và giới hạn dưới của tác động nhân quả bằng cách sử dụng các giả định đáng tin cậy (Mục 8.1). Một cách khác chúng ta có thể làm điều này là mô phỏng mức độ mạnh mẽ của tác động của yếu tố gây nhiễu lên việc điều trị và tác động của yếu tố gây nhiễu lên kết quả điều trị kết quả cần phải làm cho tác động nhân quả thực sự khác biệt đáng kể từ ước tính của chúng tôi (Phần 8.2).



(a) Không có nhiễu loạn không quan sát được



(b) Nhiễu không quan sát được ( )

Hình 8.1: Ở bên trái, chúng ta có cài đặt mà chúng ta đã xem xét cho đến nay, nơi chúng ta có tính vô căn cứ/tiêu chí cửa sau. Ở bên phải, chúng ta có một biểu đồ đơn giản trong đó yếu tố gây nhiễu không được quan sát làm cho tác động nhân quả của lên không thể xác định được.

### 8.1 Giới hạn

Có sự cân bằng giữa mức độ thực tế và độ tin cậy của các giả định của chúng ta là gì và chúng tôi có thể nhận được kết quả nhận dạng chính xác đến mức nào. Manski [53] gọi đây là "Quy luật giảm độ tin cậy: độ tin cậy của suy luận giảm dần theo độ mạnh của các giả định được duy trì."

Tùy thuộc vào những giả định mà chúng ta sẵn sàng đưa ra, chúng ta có thể rút ra các giới hạn phi tham số khác nhau về tác động nhân quả. Chúng ta đã thấy rằng nếu chúng tôi sẵn sàng thừa nhận tính vô căn cứ (hoặc một số biểu đồ nhân quả trong mà tác động nhân quả có thể xác định được) và tính tích cực, chúng ta có thể xác định một điểm duy nhất cho tác động nhân quả. Tuy nhiên, điều này có thể không thực tế. Vì ví dụ, luôn có thể có sự nhiễu loạn không quan sát được trong quan sát học.

Đây chính là động lực thúc đẩy công việc của Charles Manski về giới hạn các tác động nhân quả [53-60]. Điều này cho chúng ta một khoảng thời gian mà hiệu ứng nhân quả phải nằm trong đó, đúng hơn là cho chúng ta biết chính xác điểm nào trong khoảng đó mà tác động nhân quả phải là. Trong phần này, chúng tôi sẽ giới thiệu về các phương pháp phi tham số này giới hạn và cách tìm ra chúng.

Các giả định mà chúng tôi cho là yếu hơn sự vô căn cứ,

vì vậy chúng cho chúng ta những khoảng thời gian mà hiệu ứng nhân quả phải rơi vào (theo những khoảng thời gian này

8.1 Giới hạn . . . . .	73
Không có ràng buộc giả định . . .	74
Điều trị lại đơn điệu	
tài trợ . . . . .	76
Lựa chọn điều trị đơn điệu-	
chuyên . . . . .	78
Lựa chọn điều trị tối ưu	79
8.2 Phân tích độ nhạy . . . . .	82
Khái niệm cơ bản về độ nhạy trong tuyến tính	
Cài đặt . . . . .	82
Thêm Cài đặt chung . . .	85

[53]: Manski (2003), Xác định một phần Phân phối xác suất: Chuỗi Springer trong Sô liệu thống kê

[54]: Manski (1989), 'Giải phẫu vấn đề lựa chọn'

[55]: Manski (1990), 'Phi tham số Giới hạn về hiệu quả điều trị'

[56]: Manski (1993), 'Vấn đề nhận dạng-lems trong Khoa học xã hội'

[57]: Manski (1994), 'Vấn đề lựa chọn '

[58]: Manski (1997), 'Điều trị đơn điệu Phản ứng'

[59]: Manski và Pepper (2000), ' Các biến thể của nhạc cụ đơn điệu: Với ứng dụng cho việc quay trở lại trường học'

[53]: Manski (2003), Xác định một phần Phân phối xác suất: Chuỗi Springer trong Sô liệu thống kê

[60]: Manski (2013), Chính sách công trong một thế giới không chắc chắn

giả định). Nếu chúng ta giả định giả định mạnh mẽ hơn về tính vô căn cứ thì các khoảng này sẽ thu gọn về một điểm duy nhất. Điều này minh họa quy luật giảm độ tin cậy.

### 8.1.1 Ràng buộc không có giả định

Nói tắt cả những gì chúng ta biết về kết quả tiềm năng  $(0)$  và  $(1)$  là chúng nằm trong khoảng từ  $0$  đến  $1$ . Khi đó, giá trị tối đa của ITE  $(1)$  là  $1$  ( $0 - 0$ ) và tối thiểu là  $-1$  ( $0 - 1$ ):

$$1 \leq (1) \quad (0) \leq 1 \quad \text{nếu } , \quad 0 \leq (-) \leq 1 \quad (8.1)$$

Vì vậy, chúng ta biết rằng tất cả các ITE phải nằm trong khoảng độ dài 2. Bởi vì tất cả các ITE phải nằm trong khoảng độ dài 2 này, nên ATE cũng phải nằm trong khoảng độ dài 2 này. Điều thú vị là đối với ATE, hóa ra là chúng ta có thể cắt giảm một nửa độ dài của khoảng thời gian này mà không đưa ra bất kỳ giả định nào (vượt quá giá trị tối thiểu/tối đa của kết quả); khoảng mà ATE phải rơi vào chỉ có độ dài 1.

Chúng ta sẽ trình bày kết quả này của Manski [55] trong kích thước tổng quát hơn trong đó kết quả bị chặn giữa và :

[55]: Manski (1990), 'Phi tham số giới hạn về hiệu quả điều trị'

**Giả định 8.1 (Kết quả tiềm năng bị ràng buộc)**

$$, \quad \leq (-) \leq \quad (8.2)$$

Bằng lý do tương tự như trên, điều này dẫn đến các giới hạn sau đối với ITE và ATE:

**Bài tập đọc tích cực:** Đảm bảo bạn làm theo cách chúng tôi đạt được những giới hạn này.

$$\text{????} - - - - - \quad (8.3)$$

$$(8.4)$$

Đây là các khoảng có độ dài  $(1)(0) = 2(0)$ . Và giới hạn của ITE không thể chặt chẽ hơn nếu không có thêm giả định. Tuy nhiên, dường như thật kỳ diệu, chúng ta có thể giảm một nửa độ dài của khoảng thời gian dành cho ATE.

Để thấy điều này, chúng ta viết lại ATE như sau:

$$[(1)(0)] = [(1)] + [(0)] \quad (8.5)$$

$$= ( = 1) [ (1) | = 1] + ( = 0) [ (1) | = 0]$$

$$( = 1) [ (0) | = 1] + ( = 0) [ (0) | = 0] \quad 0] \quad (8.6)$$

Chúng tôi ngay lập tức nhận ra số hạng đầu tiên và số hạng cuối cùng là những kỳ vọng có điều kiện thân thiện mà chúng tôi có thể ước tính từ dữ liệu quan sát:

**Bài tập đọc tích cực:** Chúng ta đang sử dụng giả định nào ở đây?

$$= ( = 1) [ | = 1] + ( = 0) [ (1) | = 0] \\ ( = 1) [ (0) | = 1] + ( = 0) [ | = 0] \quad 0] \quad (8.7)$$

Bởi vì đây là một phép phân tách quan trọng nên chúng ta sẽ đặt tên và hộp cho nó trước khi tiếp tục với đạo hàm giới hạn. Chúng ta sẽ gọi đây là sự phân rã quan sát-phản thực tế (của ATE). Ngoài ra, để có

ký hiệu ngắn gọn hơn một chút, chúng ta sẽ sử dụng  $(\cdot = 1)$  trong tương lai.

#### Mệnh đề 8.1 (Phân rã quan sát-phản thực tế)

$$\begin{aligned} [(\cdot - 1) \cdot (\cdot - 0)] &= [(\cdot - 1) \cdot (\cdot - 1)] + (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \\ &= [(\cdot - 0) \cdot (\cdot - 1)] + (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \quad (8.8) \end{aligned}$$

Thật không may,  $[(\cdot - 1) \cdot (\cdot - 0) = 0]$  và  $[(\cdot - 0) \cdot (\cdot - 1) = 1]$  là phản thực.

Tuy nhiên, chúng ta biết rằng chúng bị giới hạn giữa  $\cdot$  và  $\cdot$ . Do đó, chúng ta đạt được giới hạn trên của biểu thức đầy đủ bằng cách để cho số lượng được thêm vào  $([\cdot - 1) \cdot (\cdot - 0) = 0])$  bằng  $\cdot$  và để cho số lượng bị trừ  $([(\cdot - 0) \cdot (\cdot - 1) = 1])$  bằng  $\cdot$ . Tương tự, chúng ta có thể đạt được giới hạn dưới bằng cách để số hạng được cộng bằng  $\cdot$  và số hạng bị trừ bằng  $\cdot$ .

Mệnh đề 8.2 (Ràng buộc Không giả định) Giả sử biểu thị  $(\cdot = 1)$ , trong đó  $\cdot$  là biến ngắn nhiên nhi phân. Cho rằng kết quả  $\cdot$  bị chặn giữa  $\cdot$  và  $\cdot$  (Giả định 8.1), chúng ta có các giới hạn trên và dưới sau đây về ATE:

$$\begin{aligned} [(\cdot - 1) \cdot (\cdot - 0)] &\leq [(\cdot - 1) \cdot (\cdot - 1)] + (1 \cdot (\cdot - 1)) \quad (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \\ (8.9) \quad [(\cdot - 1) \cdot (\cdot - 0)] &\geq [(\cdot - 1) \cdot (\cdot - 1)] + (1 \cdot (\cdot - 1)) \quad (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \quad (8.10) \end{aligned}$$

Điều quan trọng là độ dài của khoảng này là  $\cdot$ , bằng một nửa độ dài của khoảng ban đầu mà chúng ta đã thấy trong Công thức 8.4. Chúng ta có thể thấy điều này bằng cách trừ giới hạn dưới khỏi giới hạn trên:

$$\begin{aligned} [(\cdot - 1) \cdot (\cdot - 0) = 1] + (1 \cdot (\cdot - 1)) &\quad (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \\ 0 \cdot (1 \cdot (\cdot - 1)) &= (1 \cdot (\cdot - 1)) + (1 \cdot (\cdot - 1)) \quad (1 \cdot (\cdot - 1)) \quad [(\cdot - 1) \cdot (\cdot - 0) = 0] \\ &= \cdot \end{aligned} \quad (8.11) \quad (8.12)$$

Điều này đôi khi được gọi là “không có giả định bị ràng buộc” bởi vì chúng tôi không đưa ra giả định nào khác ngoài việc kết quả là bị ràng buộc. Nếu kết quả không bị giới hạn thì ATE và ITE có thể nằm trong khoảng từ  $\cdot$  đến  $\cdot$ .

#### Ví dụ đang chạy

Hãy xem xét rằng chúng tôi biết rằng các kết quả được giới hạn trong khoảng  $0$  và  $1$  (ví dụ: vì chúng tôi đang ở cài đặt kết quả nhị phân). Điều này có nghĩa là ITE và phải được giới hạn giữa  $-1$  ( $0 - 1$ ) và  $1$  ( $1 - 0$ ), nghĩa là ATE cũng phải được giới hạn giữa  $-1$  và  $1$ . Đối với ví dụ này, cũng xét rằng  $\cdot = 0,3$ ,  $[(\cdot - 1) \cdot (\cdot - 0) = 1] = 0,9$ , và  $[(\cdot - 1) \cdot (\cdot - 0) = 0] = 0,2$ . và  $\cdot^1$  Sau đó, bằng cách thay chúng vào phương trình 8.9 8.10, chúng tôi nhận được các giới hạn sau trên ATE:

$$\begin{aligned} [(\cdot - 1) \cdot (\cdot - 0)] &\leq (.3)(.9) + (1 - .3)(1) - (.3)(0) - (1 - .3)(.2) \quad (8.13) \quad [(\cdot - 1) \cdot (\cdot - 0)] \\ &\geq (.3)(.9) + (1 - .3)(0) - (.3)(1) - (1 - .3)(.2) \quad (8.14) \end{aligned}$$

<sup>1</sup> **Bài tập đọc tách cúc:** Chúng ta ước tính những kỳ vọng có điều kiện này như thế nào?

$$0,17 \leq [ (1) - (0) ] \leq 0,83 \quad (8.15)$$

Lưu ý rằng khoảng này có độ dài 1 ( $= 1$ ), một nửa độ dài của khoảng ban đầu  $1 \leq [ (1) - (0) ] \leq 1$  (Phương trình 8.4). Chúng ta sẽ sử dụng ví dụ đang chạy này trong suốt Phần 8.1.

#### Bài tập đọc tích cực:

- Chúng ta có thể nhận được loại giới hạn nào cho CATE  $[ (1) - (0) | ]$ , giả sử chúng ta có sự tích cực? Điều gì sẽ xảy ra nếu chúng ta không có sự tích cực?
- Giả sử các kết quả tiềm năng bị giới hạn theo nhiều cách khác nhau:  $1 \leq (1) \leq 1$  và  $0 \leq (0) \leq 0$ . Rút ra các giới hạn không có giả định tương ứng trong bối cảnh tổng quát hơn này.

Các giới hạn trong Dự luật 8.2 chặt chẽ đến mức chúng ta có thể đạt được mà không cần giả định thêm. Thật không may, khoảng tương ứng luôn chứa 0,2, điều đó có nghĩa là chúng ta không thể sử dụng giới hạn này để phân biệt "không có tác động nhân quả" với "hiệu ứng nhân quả". Chúng ta có thể có được giới hạn chặt chẽ hơn?

Để giới hạn ATE, chúng ta phải có một số thông tin về phần phản thực của sự phân tách này. Chúng ta có thể dễ dàng ước tính phần quan sát từ dữ liệu. Trong trường hợp không có giả định nào bị ràng buộc (Mệnh đề 8.2), tất cả những gì chúng tôi giả định là các kết quả bị giới hạn bởi và . Nếu chúng ta đưa ra nhiều giả định hơn, chúng ta có thể có được những khoảng thời gian nhỏ hơn. Trong một số phần tiếp theo, chúng tôi sẽ đề cập đến một số giả định đôi khi khá hợp lý, tùy thuộc vào bối cảnh và những giới hạn chặt chẽ hơn mà những giả định này đưa ra cho chúng tôi. Chiến lược chung mà chúng tôi sẽ sử dụng cho tất cả chúng là bắt đầu bằng việc phân rã quan sát-phản thực tế của ATE (Dự luật 8.1),

$$\begin{aligned} [ (1) - (0) ] &= [ | ] = 1] + (1 - ) [ (1) | ] = \\ 0] &\quad [ (0) | ] = 1] (1 - ) [ | ] = 0] , \text{ (xem lại 8.8)} \end{aligned}$$

và có được các khoảng nhỏ hơn bằng cách giới hạn các phần phản thực bằng cách sử dụng các giả định khác nhau mà chúng tôi đưa ra.

Các khoảng mà chúng ta sẽ thấy trong vài phần phụ tiếp theo đều sẽ chứa số 0. Chúng ta sẽ không thấy khoảng hoàn toàn dương hoặc hoàn toàn âm cho đến Phần 8.1.4, vì vậy, hãy bỏ qua phần đó nếu bạn chỉ muốn xem những khoảng đó.

#### 8.1.2 Đáp ứng điều trị đơn điệu

Đối với giả định đầu tiên của chúng tôi ngoài việc giả định các kết quả bị giới hạn, hãy cân nhắc rằng chúng tôi thấy mình đang ở trong một bối cảnh mà việc điều trị chỉ có thể giúp ích là khai thi; nó không thể đau đớn được. Đây là bối cảnh mà Manski [58] xem xét trong bối cảnh. Trong bối cảnh này, chúng ta có thể biến minh cho giả định về đáp ứng điều trị đơn điệu (MTR) :

Giả định 8.2 (Đáp ứng điều trị đơn điệu không âm tính)

$$(1) \geq (0) \quad (8.16)$$

<sup>2</sup> Để biết tại sao giới hạn không giả định luôn chứa số 0, hãy xem xét những gì chúng ta cần để nó không chứa số 0: chúng ta cần giới hạn trên nhỏ hơn 0 hoặc giới hạn dưới lớn hơn 0. Tuy nhiên, điều này không thể xảy ra được. Để biết lý do, hãy lưu ý rằng giới hạn trên tối thiểu đạt được khi  $[ | ] = 1$  và  $[ | ] = 0$  = , cho chúng ta một giới hạn trên (bao gồm) của 0. Tương tự với giới hạn dưới.

**Bài tập đọc tích cực:** Cho thấy giới hạn dưới tối đa là 0.

[58]: Manski (1997), 'Điều trị đơn điệu Phản ứng'

Điều này có nghĩa là mọi ITE đều không âm, vì vậy chúng ta có thể nâng giới hạn dưới của ITE lên từ (Công thức 8.3) lên 0. Vì vậy, theo trực giác, điều này có nghĩa là giới hạn dưới của chúng ta trên ATE sẽ di chuyển lên 0. Và bây giờ chúng ta sẽ thấy rằng đây là trường hợp.

Bây giờ, thay vì giới hạn dưới  $[ (1) | \dots = 0]$  với  $\dots$  và  $[ (0) | \dots = 1]$  với  $\dots$ , chúng ta có thể làm tốt hơn. Bởi vì việc điều trị chỉ giúp ích,  $[ (1) | \dots = 0] \geq [ (0) | \dots = 0] = [ \dots | \dots = 0]$ , nên ta có thể hạ giới hạn  $[ (1) | \dots = 0]$  với  $[ \dots | \dots = 0]$ . Tương tự,  $[ (0) | \dots = 1] \geq [ (1) | \dots = 1] = [ \dots | \dots = 1]$  (vì nhân với số âm sẽ lật bất đẳng thức), nên chúng ta có thể giới hạn dưới  $[ (0) | \dots = 1]$  với  $[ \dots | \dots = 1]$ .

Do đó, chúng ta có thể cải thiện giới hạn dưới không có giả định để đạt được 0, như trực giác của chúng ta gợi ý:

$$\begin{aligned} [ (1) & - (0) ] = [ \dots | \dots = 1] + (1 \dots) [ (1) | \dots = 0] \\ & [ (0) | \dots = 1] (1 \dots) [ \dots | \dots = 0] \\ & \quad (\text{xem lại 8.8}) \\ & \geq [ \dots | \dots = 1] + (1 \dots) [ \dots | \dots = 0] \\ & [ \dots | \dots = 1] (1 \dots) [ \dots | \dots = 0] \quad (8.17) \\ & = 0 \end{aligned}$$

<sup>3</sup> Hãy nhớ lại rằng bằng cách chỉ giả định rằng kết quả đầu ra bị giới hạn giữa và , chúng ta sẽ có được giới hạn dưới không có giả định (Dự luật 8.2):

$$\begin{aligned} [ (1) & - (0) ] \\ & \geq [ \dots | \dots = 1] + (1 \dots) \\ & \quad (1 \dots) [ \dots | \dots = 0] \\ & \quad (\text{xem lại 8.10}) \end{aligned}$$

Mệnh đề 8.3 (Giới hạn dưới MTR không âm) Theo giả định MTR không âm, ATE được giới hạn từ bên dưới bằng 0. Về mặt toán học,

$$[ (1) - (0) ] \geq 0 \quad (8.19)$$

Ví dụ chạy Giới hạn trên không có giả định<sup>4</sup> vẫn được áp dụng ở đây, vì vậy, trong ví dụ chạy của chúng tôi từ Phần 8.1.1 trong đó  $= .3$ ,  $[ \dots | \dots = 1] = .9$ , và  $[ \dots | \dots = 0] = .02$ , khoảng ATE của chúng tôi cải thiện từ  $[ .017, .083 ]$

(Phương trình 8.15) đến  $[ 0, .083 ]$ .

Ngoài ra, hãy nói rằng việc điều trị chỉ có thể làm tổn thương con người; nó không thể giúp họ (ví dụ như một vết thương do đạn bắn chỉ làm mất đi cơ hội sống sót). Trong những trường hợp đó, chúng ta sẽ có giả định về đáp ứng điều trị đơn điệu không dương tính và giới hạn trên MTR không dương tính :

Giả định 8.3 (Đáp ứng điều trị đơn điệu không tích cực)

$$(1) \leq (0) \quad (8.20)$$

<sup>4</sup> Hãy nhớ lại giới hạn trên của việc không có giả định (Dự luật 8.2):

$$\begin{aligned} [ (1) & - (0) ] \\ & \leq [ \dots | \dots = 1] + (1 \dots) \\ & \quad (1 \dots) [ \dots | \dots = 0] \\ & \quad (\text{xem lại 8.9}) \end{aligned}$$

Mệnh đề 8.4 (Giới hạn trên MTR không dương) Theo giả định MTR không dương, ATE được giới hạn từ phía trên bằng 0. Về mặt toán học,

$$[ (1) - (0) ] \leq 0 \quad (8.21)$$

Bài tập đọc tích cực: Chứng minh命题 8.4.

Ví dụ đang chạy Và trong cài đặt này, giới hạn dưới<sup>5</sup> không có giả định vẫn được áp dụng. Điều đó có nghĩa là khoảng ATE trong ví dụ của chúng tôi cải thiện từ  $[ .017, .083 ]$  (Công thức 8.15) thành  $[ .017, 0 ]$ .

**Bài tập đọc chủ động:** Khoảng ATE là bao nhiêu nếu chúng ta giả sử cả MTR không âm và MTR không dương? Điều này có hợp lý không, bằng trực giác?

<sup>5</sup> Hãy nhớ lại giới hạn của việc không có giả định (Dự luật 8.2):

$$\begin{aligned} [ (1) & - (0) ] \\ & \geq [ \dots | \dots = 1] + (1 \dots) \\ & \quad (1 \dots) [ \dots | \dots = 0] \\ & \quad (\text{xem lại 8.10}) \end{aligned}$$

### 8.1.3 Lựa chọn xử lý đơn điệu

Giả định tiếp theo mà chúng ta sẽ xem xét là giả định rằng những người lựa chọn phương pháp điều trị sẽ có kết quả tốt hơn những người không chọn phương pháp điều trị, trong cả hai kịch bản điều trị.

Manski và Pepper [59] đã giới thiệu đây là giả định lựa chọn điều trị đơn điệu (MTS) .

**Giả định 8.4 (Lựa chọn phương pháp điều trị đơn điệu)**

$$[\quad(1)\quad|\quad=1] \geq [\quad(1)\quad|\quad=0] \quad (8.22)$$

$$[\quad(0)\quad|\quad=1] \geq [\quad(0)\quad|\quad=0] \quad (8.23)$$

Như Morgan và Winship [12, Phần 12.2.2] đã chỉ ra, bạn có thể coi đây là sự tự lựa chọn tích cực. Những người nhìn chung có kết quả tốt hơn sẽ tự chọn vào nhóm điều trị. Một lần nữa, chúng ta bắt đầu với phân rã quan sát-phản thực tế và bây giờ chúng ta đạt được giới hạn trên bằng cách sử dụng giả định MTS (Giả định 8.4):

**Dự luật 8.5 (Giới hạn trên của lựa chọn điều trị đơn điệu)** Theo giả định MTS, ATE được giới hạn từ phía trên bởi sự khác biệt liên kết. Về mặt toán học,

$$[\quad(1)\quad|\quad(0)] \leq [\quad|\quad=1] - [\quad|\quad=0] \quad (8.24)$$

Bằng chứng.

$$\begin{aligned} [\quad(1)\quad|\quad(0)] &= [\quad|\quad=1] + (1\quad) - [\quad(1)\quad|\quad=0] \\ &= [\quad(0)\quad|\quad=1] - (1\quad) - [\quad|\quad=0] \end{aligned} \quad (\text{xem lại 8.8})$$

$$\begin{aligned} &[\quad|\quad=1] + (1\quad) - [\quad|\quad=1] \\ &[\quad|\quad=0] - (1\quad) - [\quad|\quad=0] \end{aligned} \quad (8.25)$$

$$0] = [\quad|\quad=1] - [\quad|\quad=0] \quad (8.26)$$

trong đó Phương trình 8.25 được suy ra từ thực tế là (a) Phương trình 8.22 của giả định MTS cho phép chúng ta đạt tới giới hạn trên  $[\quad(1)\quad|\quad=0]$  bởi  $[\quad(1)\quad|\quad=1] = [\quad(1)\quad|\quad=1]$  và (b) Phương trình 8.23 của giả định MTS cho phép chúng ta đạt tới giới hạn trên  $[\quad(0)\quad|\quad=1]$  bởi  $[\quad|\quad=0]$ .  $\square$

Ví dụ chạy Hãy nhớ lại ví dụ chạy của chúng tôi từ Phần 8.1.1 trong đó  $= .3$ ,  $[\quad|\quad=1] = .9$ , và  $[\quad|\quad=0] = .2$ . Giả định MTS cho chúng ta một giới hạn trên và chúng ta vẫn có giới hạn dưới không có giả định.6 Điều đó có nghĩa là khoảng ATE trong ví dụ của chúng ta cải thiện từ  $[.0,17, .0,83]$  (Công thức 8.15) lên  $[.0,17, .0,7]$ .

Cả MTR và MTS Sau đó, chúng ta có thể kết hợp giả định MTR không âm (Giả định 8.2) với giả định MTS (Giả định 8.4) để có được giới hạn dưới trong Mệnh đề 8.3 và giới hạn trên trong Mệnh đề 8.5, tương ứng. Trong ví dụ đang chạy của chúng tôi, điều này mang lại khoảng thời gian sau cho ATE:  $[0, .0,7]$ .

[59]: Manski và Pepper (2000), ' Các biến thể của nhạc cụ đơn điệu: Või ứng dụng cho việc quay trở lại trường học'

[12]: Morgan và Winship (2014), Suy luận phản thực tế và nhân quả: Phương pháp và Nguyên tắc nghiên cứu xã hội

<sup>6</sup> Hãy nhớ lại giới hạn dưới của việc không có giả định (Dự luật 8.2):

$$\begin{aligned} &[\quad(1)\quad|\quad(0)] \\ &\geq [\quad|\quad=1] + (1\quad) \\ &(1\quad) - [\quad|\quad=0] \end{aligned} \quad (\text{xem lại 8.10})$$

Các khoảng chứa 0 Mặc dù các giới hạn từ các giả định MTR và MTS có thể hữu ích để loại trừ các tác động nhân quả rất lớn hoặc rất nhỏ, các khoảng tương ứng vẫn chứa 0. Điều này có nghĩa là những giả định này không đủ để xác định liệu có ảnh hưởng hay không.

#### 8.1.4 Lựa chọn phương pháp xử lý tối ưu

Bây giờ chúng tôi xem xét cái mà chúng tôi sẽ gọi là giả định lựa chọn điều trị tối ưu (OTS) từ Manski [55]. Giả định này có nghĩa là các cá nhân luôn nhận được phương pháp điều trị tốt nhất cho họ (ví dụ: nếu một bác sĩ chuyên môn đang quyết định áp dụng phương pháp điều trị nào cho mọi người). Chúng tôi viết điều này một cách toán học như sau:

**Giả định 8.5 (Lựa chọn phương pháp điều trị tối ưu)**

$$= 1 = \quad (1) \geq \quad (0) , \quad = 0 = \quad (0) > \quad (1) \quad (8.27)$$

Từ giả định OTS, chúng ta biết rằng

$$[ \quad (1) | \quad = 0 ] \leq [ \quad (0) | \quad = 0 ] = [ \quad | \quad = 0 ] . \quad (8.28)$$

Do đó, chúng ta có thể đưa ra giới hạn trên, bằng giới hạn trên  $[ \quad (1) | \quad = 0 ]$  với  $[ \quad | \quad = 0 ]$  và giới hạn trên  $[ \quad (0) | \quad = 1 ]$  với (giống như ở giới hạn trên không có giả định7):

$$\begin{aligned} [ \quad (1) \quad (0) ] &= [ \quad | \quad = 1 ] + (1 \quad ) \quad [ \quad (1) | \quad = 0 ] \\ &\quad [ \quad (0) | \quad = 1 ] \quad (1 \quad ) \quad [ \quad | \quad = 0 ] \\ &\quad (\text{xem lại 8.8}) \\ &[ \quad | \quad = 1 ] + (1 \quad ) \quad [ \quad | \quad = 0 ] \\ &\quad (1 \quad ) \quad [ \quad | \quad = 0 ] \quad (8.29) \\ &= [ \quad | \quad = 1 ] \quad (8.30) \end{aligned}$$

Giả định OTS cũng cho chúng ta biết rằng

$$[ \quad (0) | \quad = 1 ] \leq [ \quad (1) | \quad = 1 ] = [ \quad | \quad = 1 ] , \quad (8.31)$$

tương đương với câu nói  $[ \quad (0) | \quad = 1 ] \geq [ \quad | \quad = 1 ]$ . Vì vậy, chúng ta có thể giới hạn dưới  $[ \quad (0) | \quad = 1 ]$  với  $[ \quad | \quad = 1 ]$ , và chúng ta có thể hạ giới hạn  $[ \quad (1) | \quad = 0 ]$  với (giống như chúng ta đã làm ở giới hạn dưới không có giả định8) để có giới hạn dưới sau:

$$\begin{aligned} [ \quad (1) \quad (0) ] &= [ \quad | \quad = 1 ] + (1 \quad ) \quad [ \quad (1) | \quad = 0 ] \\ &\quad [ \quad (0) | \quad = 1 ] \quad (1 \quad ) \quad [ \quad | \quad = 0 ] \\ &\quad (\text{xem lại 8.8}) \\ &\geq [ \quad | \quad = 1 ] + (1 \quad ) \\ &\quad [ \quad | \quad = 1 ] \quad (1 \quad ) \quad [ \quad | \quad = 0 ] \quad (8.32) \\ &= (1 \quad ) \quad (1 \quad ) \quad [ \quad | \quad = 0 ] \quad (8.33) \end{aligned}$$

[55]: Manski (1990), 'Phi tham số  
Giới hạn về hiệu quả điều trị'

<sup>7</sup> Hãy nhớ lại giới hạn trên của việc không có giả định (Dự luật 8.2):

$$\begin{aligned} [ \quad (1) \quad (0) ] &= [ \quad | \quad = 1 ] + (1 \quad ) \\ &\quad (1 \quad ) \quad [ \quad | \quad = 0 ] \\ &\quad 0] \quad (\text{xem lại 8.9}) \end{aligned}$$

<sup>8</sup> Hãy nhớ lại giới hạn dưới của việc không có giả định (Dự luật 8.2):

$$\begin{aligned} [ \quad (1) \quad (0) ] &\geq [ \quad | \quad = 1 ] + (1 \quad ) \\ &\quad (1 \quad ) \quad [ \quad | \quad = 0 ] \\ &\quad (\text{xem lại 8.10}) \end{aligned}$$

**Mệnh đề 8.6 (Giới hạn lựa chọn xử lý tối ưu 1)** Gọi biểu thị  $(\cdot = 1)$ , trong đó  $\cdot$  là biến ngẫu nhiên nhị phân. Cho rằng kết quả  $\cdot$  là được giới hạn từ bên dưới bởi (Giả định 8.1) và phương pháp xử lý tối ưu là luôn luôn lựa chọn (Giả thiết 8.5), ta có các hàm trên và dưới sau giới hạn trên ATE:

$$[\cdot(1) - \cdot(0)] < [\cdot| = 1] \quad (8.34)$$

$$[\cdot(1) - \cdot(0)] \geq (1 - \cdot) \cdot (1 - \cdot) [\cdot| = 0] \quad (8.35)$$

$$\text{Độ dài khoảng} = [\cdot| = 1] + (1 - \cdot) [\cdot| = 0] \quad (8.36)$$

Thật không may, khoảng này cũng luôn chứa số 0! Điều này có nghĩa là Mệnh đề 8.6 không cho chúng ta biết liệu tác động nhân quả là khác 0 hay không.

Ví dụ chạy Hãy nhớ lại ví dụ chạy của chúng tôi từ Phần 8.1.1 trong đó  $\cdot = 0$ ,  $\cdot = 1$ ,  $\cdot = .3$ ,  $[\cdot| = 1] = .9$ , và  $[\cdot| = 0] = .2$ . Đang cầm những điều này trong Dự luật 8.6 cho chúng ta những điều sau:

$$[\cdot(1) - \cdot(0)] \leq (.3)(.9) - (.3)(.0) \quad (8.37)$$

$$[\cdot(1) - \cdot(0)] \geq (1 - .3)(.0) - (1 - .3)(.2) = 0,14 \quad (8.38)$$

$$\leq [\cdot(1) - \cdot(0)] \leq 0,27 \text{ Độ dài khoảng} \quad (8.39)$$

$$= 0,41 \quad (8.40)$$

Bây giờ chúng ta sẽ đưa ra một khoảng có thể hoàn toàn dương hoặc hoàn toàn âm, có khả năng xác định ATE khác 0.

Một giới hạn có thể xác định dấu hiệu của ATE

Hóa ra là, mặc dù chúng ta lấy giả định OTS từ Manski [55], giới hạn mà chúng tôi đưa ra trong Dự luật 8.6 thực tế không phải là giới hạn mà Manski [55] rút ra giả định đó. Ví dụ: nơi chúng tôi sử dụng  $[\cdot(1)| = 0] \leq [\cdot| = 0]$ , Manski sử dụng  $[\cdot(1)| = 0] \leq [\cdot| = 1]$ . Chúng ta sẽ nhanh chóng chứng minh bất đẳng thức mà Manski sử dụng từ giả định OTS:10

Chúng ta bắt đầu bằng cách áp dụng Công thức 8.42:

$$[\cdot(1)| = 0] = [\cdot(1)| = 0] > \cdot(1) \quad (8.45)$$

Bởi vì biến ngẫu nhiên mà chúng ta đang lấy kỳ vọng là  $\cdot(1)$ , nếu chúng ta lật  $\cdot(0) > \cdot(1)$  thành  $\cdot(0) \leq \cdot(1)$ , thì chúng ta có giới hạn trên:

$$[\cdot(1)| = 0] \leq \cdot(1) \quad (8.46)$$

Cuối cùng, áp dụng phương trình 8.44, ta có kết quả:

$$= [\cdot(1)| = 1] \quad (8.47)$$

$$= [\cdot| = 1] \quad (8.48)$$

Bây giờ chúng ta có cái đó  $[\cdot(1)| = 0] \leq [\cdot| = 1]$ , chúng ta có thể chứng minh giới hạn trên của Manski [55], trong đó chúng ta sử dụng bất đẳng thức then chốt này trong

<sup>9</sup> [Bài tập đọc tích cực](#): Cho thấy rằng khoảng luôn chứa số không.

[55]: Manski (1990), 'Phi tham số Giới hạn về hiệu quả điều trị'

<sup>10</sup> Nhớ lại giả định OTS (Giả định 8.5):

$$= 1 = \cdot(1) \geq \cdot(0) \quad (8.41)$$

$$= 0 = \cdot(0) > \cdot(1) \quad (8.42)$$

Bởi vì chỉ có hai giá trị có thể đếm nhận, điều này tương đương với sau đây (mâu thuẫn):

$$= 0 = \cdot(1) < \cdot(0) \quad (8.43)$$

$$= 1 = \cdot(0) > \cdot(1) \quad (8.44)$$

[55]: Manski (1990), 'Phi tham số Giới hạn về hiệu quả điều trị'

Phương trình 8.49:

$$\begin{aligned} [(-1)(0)] &= [(-1)| = 1] + (1|) \quad [(-1)| = 0] \\ &= [(-1)| = 1] - (1|) \quad [(-1)| = 0] \\ &\quad (8.8 \text{ xem lại}) \\ &= [(-1)| = 1] + (1|) \quad [(-1)| = 1] \\ &\quad (1|) \quad [(-1)| = 0] \quad (8.49) \\ 0] &= [(-1)| = 1] + (1|) \quad [(-1)| = 1] \\ &\quad (1|) \quad [(-1)| = 0] \quad (8.50) \\ &= [(-1)| = 1] \quad (1|) \quad [(-1)| = 0] \quad (8.51) \end{aligned}$$

Tương tự, chúng ta có thể thực hiện một phép đạo hàm tương tự<sup>11</sup> để có được mức thấp hơn ràng buộc:

$$[(-1)(0)] \geq [(-1)| = 1] + (1|) \quad [(-1)| = 0] \quad (8.52)$$

<sup>11</sup> [Bài tập đọc tích cực: Tự suy ra phương trình 8.52](#).

Mệnh đề 8.7 (Giới hạn lựa chọn xử lý tối ưu 2) Gọi biểu thị

( $= 1$ ), trong đó  $\alpha$  là biến ngẫu nhiên nhị phân. Cho rằng kết quả  $\alpha$  là được giới hạn từ bên dưới bởi  $\beta$  (Giả định 8.1) và phương pháp xử lý tối ưu là luôn luôn lựa chọn (Giả thiết 8.5), ta có các hàm trên và dưới sau giới hạn trên ATE:

$$[(-1)(0)] \leq [(-1)| = 1] \quad (1|) \quad [(-1)| = 0] \quad (8.53)$$

$$[(-1)(0)] \geq [(-1)| = 1] + (1|) \quad [(-1)| = 0] \quad (8.54)$$

$$\text{Độ dài khoảng} = (1|) \quad [(-1)| = 1] + [(-1)| = 0] \quad (8.55)$$

Khoảng này cũng có thể bao gồm số 0, nhưng không nhất thiết phải như vậy. Ví dụ, trong ví dụ đang chạy của chúng tôi thì không.

Ví dụ chạy Hãy nhớ lại ví dụ chạy của chúng tôi từ Phần 8.1.1 trong đó

$\alpha = 0$ ,  $\beta = 1$ ,  $\gamma = .3$ ,  $[(-1)| = 1] = .9$ , và  $[(-1)| = 0] = .2$ . Đang cầm những điều này trong Dự luật 8.7 cung cấp cho chúng ta những điều sau đây đối với giới hạn OTS 2:

$$[(-1)(0)] \leq (.9)(.3)(0) \quad (1|.3) \quad (8.56)$$

$$(.2)[(-1)(0)] \geq (-.3)(.9) + (1|.3) \quad (8.57)$$

$$(0) \quad (.2) \quad 0.07 \leq [(-1)| = 1] \quad (8.58)$$

$$(0) \leq 0.76 \quad \text{Độ dài khoảng} = 0.69 \quad (8.59)$$

Vì vậy, trong khi OTS giới hạn 2 từ Manski [55] xác định dấu của ATE trong ví dụ đang chạy của chúng tôi, không giống như giới hạn OTS 1, giới hạn OTS 2 cho chúng tôi một khoảng thời gian lớn hơn 68%. Bạn có thể thấy điều này bằng cách so sánh phương trình 8.40 (trong lề trên) với phương trình 8.59.

Điều này minh họa một số bài học quan trọng:

1. Các giới hạn khác nhau sẽ tốt hơn trong các trường hợp khác nhau.<sup>12</sup>
2. Các giới hạn khác nhau có thể tốt hơn theo những cách khác nhau (ví dụ: xác định dấu hiệu so với nhận được một khoảng nhỏ hơn).

Trên giới hạn May mắn thay vì cả OTS bị ràng buộc 1 và OTS ràng buộc 2 xuất phát từ cùng một giả định (Giả định 8.5), chúng ta có thể lấy giới hạn dưới của OTS giới hạn 2 và giới hạn trên của OTS

Áp dụng ràng buộc OTS 1 (Dự luật 8.6) vào ví dụ đang chạy của chúng tôi:

$$0.14 \leq [(-1)(0)] \leq 0.27 \quad (\text{xem lại 8.39})$$

$$\text{Độ dài khoảng thời gian} = 0.41 \quad (\text{xem lại 8.40})$$

[55]: Manski (1990), 'Phi tham số Giới hạn về hiệu quả điều trị'

<sup>12</sup> [Bài tập đọc tích cực: Sử dụng các phương trình 8.40 và 8.59, rút ra điều kiện theo đó OTS giới hạn 1 mang lại giá trị nhỏ hơn khoảng thời gian và các điều kiện theo đó OTS giới hạn 2 mang lại khoảng thời gian nhỏ hơn.](#)

ràng buộc 1 để có được khoảng thời gian chật chẽ hơn sau đây mà vẫn xác định được dấu hiệu:

$$0,07 \leq [ (1) (0) ] \leq 0,27 \quad (8,60)$$

Tương tự, chúng ta có thể trộn giới hạn dưới từ giới hạnOTS 1 và giới hạn trên củaOTS giới hạn 2, nhưng điều đó sẽ mang lại điều tồi tệ nhất khoảng trong tiêu mục này cho ví dụ cụ thể này. Nó có thể là tốt nhất Tuy nhiên, trong một ví dụ khác.

Trong phần này, chúng tôi đã cho bạn biết loại kết quả mà chúng tôi có thể nhận được từ các giới hạn không tham số, nhưng tất nhiên đây chỉ là phần giới thiệu. Để biết thêm tài liệu về vấn đề này, xem, ví dụ, [53-60].

## 8.2 Phân tích độ nhạy

### 8.2.1 Cơ bản về độ nhạy trong cài đặt tuyến tính

Trước chương này, chúng tôi chỉ làm việc trong môi trường nơi có thể xác định được các tác động nhân quả. Chúng tôi minh họa ví dụ phổ biến của các yếu tố gây nhiễu là nguyên nhân phổ biến của và trong Hình 8.2. TRONG Trong ví dụ này, tác động nhân quả của lên là có thể xác định được. Tuy nhiên, điều gì sẽ xảy ra nếu có một yếu tố gây nhiễu duy nhất không được quan sát, như chúng tôi minh họa trong Hình 8.3. Sau đó, tác động nhân quả là không thể xác định được.

Chúng ta sẽ quan sát thấy độ lệch nào nếu chúng ta chỉ điều chỉnh cho các giá trị được quan sát những điều gì? Để minh họa điều này một cách đơn giản, chúng ta sẽ bắt đầu với một mô hình không ổn định quá trình tạo dữ liệu tuyến tính. Vì vậy, hãy xem xét dữ liệu được tạo bởi các phương trình cấu trúc sau:

$$:= + \quad (8,61)$$

$$:= + + \quad (8,62)$$

Vì vậy đại lượng thích hợp mô tả tác động nhân quả của lên là vì nó là hệ số đứng trước trong phương trình cấu trúc của . Từ điều chỉnh cửa sau (Định lý 4.2) /công thức điều chỉnh (Định lý 2.1), Chúng ta biết rằng

$$[ (1) (0) ] = , [ [ | ] = 1, , ] [ | ] = 0, , ] = \quad (8,63)$$

Nhưng vì không được quan sát nên điều tốt nhất chúng ta có thể làm là chỉ điều chỉnh cho . Ở đây, điều này dẫn đến một sai lệch gây —. Chúng ta sẽ tập trung vào việc nhận dạng, nhiều là không ước tính, vì vậy chúng ta sẽ coi như chúng ta có dữ liệu vô hạn. Cái này có nghĩa là chúng ta có quyền truy cập vào ( , , ). Sau đó chúng ta sẽ viết ra và Chúng minh mệnh đề sau đây về độ lệch nhiễu:

Mệnh đề 8.8 Khi và được tạo ra bởi quá trình tuyến tính không nhiễu trong các phương trình 8.61 và 8.62, độ lệch gây nhiễu của việc điều chỉnh chỉ cho (và

[54]: Manski (1989), 'Giải phẫu vấn đề lựa chọn'

[55]: Manski (1990), 'Phi tham số Giới hạn về hiệu quả điều trị'

[56]: Manski (1993), 'Các vấn đề về nhận dạng trong khoa học xã hội'

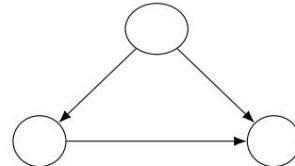
[57]: Manski (1994), 'Vấn đề lựa chọn '

[58]: Manski (1997), 'Điều trị đơn điệu Phản ứng'

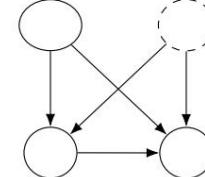
[59]: Manski và Pepper (2000), ' Các biến thể của nhạc cụ đơn điệu: Với ứng dụng cho việc quay trở lại trường học'

[53]: Manski (2003), Xác định một phần Phân phối xác suất: Chuỗi Springer trong Số liệu thống kê

[60]: Manski (2013), Chính sách công trong một thế giới không chắc chắn



Hình 8.2: Cấu trúc nhân quả đơn giản trong đó làm xáo trộn tác động của lên và trong đó là yếu tố gây nhiễu duy nhất.



Hình 8.3: Cấu trúc nhân quả đơn giản trong đó là yếu tố gây nhiễu được quan sát và là những yếu tố gây nhiễu không được quan sát.

<sup>13</sup> **Bài tập đọc tích cực:** Giả định nào bị vi phạm khi dữ liệu được tạo ra bởi một quy trình không có tiếng ồn?

không phải ) là . Về mặt toán học:

$$[ \quad [ \quad | \quad ] = 1, \quad ] \quad [ \quad | \quad ] = 0, \quad ] ] \\ , \quad [ \quad [ \quad | \quad ] = 1, \quad , \quad ] \quad [ \quad | \quad ] = 0, \quad , \quad ] ] = \boxed{0} \quad (8.64)$$

Bằng chứng. Chúng ta sẽ chứng minh Mệnh đề 8.8 theo 3 bước:

1. Nhận biểu thức dạng đóng cho  $[ \quad [ \quad | \quad = \quad , \quad ] ]$  theo  $,$   
 $,$   $,$  và  $.$

2. Sử dụng bước 1 để lấy biểu thức dạng đóng cho hiệu  $[ \quad [ \quad | \quad ] ]$  theo  $,$   
 $= 1,$   $\quad ]$   $[ \quad | \quad = 0,$   $\quad ] ]$ .

3. Trừ đi  $,$   $[ \quad [ \quad | \quad = 1,$   $, \quad ] ]$   $[ \quad | \quad = 0,$   $, \quad ] ] =$   $.$

Đầu tiên, chúng ta sử dụng phương trình cấu trúc của  $\hat{A}$  (Phương trình 8.62)

14 Bài tập đọc tích cực: Chứng tỏ rằng

$$[ \quad [ \quad | \quad = \quad , \quad ] ] = \quad [ \quad \quad \quad + \quad \quad \quad + \quad \quad | \quad = \quad , \quad ] (8,65) (8,66)$$

Đây là nơi chúng tôi sử dụng phương trình cấu trúc cho (Phương trình 8.61)  
Sắp xếp lại nó cho ta = kỳ ————— . Sau đó chúng ta có thể sử dụng nó cho  
vòng có điều kiện còn lại:

(8, 67)

(8.68)

$$[ ]^+ \quad - \quad - \quad [ ]^+ \quad (8,69)$$

Sau đó, sắp xếp lại một chút, chúng ta có như sau:

$\vdash + \_$        $+ \_$       [ ]      (8, 70)

Phản ứng nhất quan trọng của vân đề này là những phản ứng thuộc vào vì chúng ta muốn biết tác dụng của lên . Ví dụ: hãy xem xét ước tính ATE dự kiến mà chúng ta sẽ nhận được nếu chỉ điều chỉnh cho :

$$[ \quad [ \quad | \quad ] = 1, \quad ] \quad [ \quad | \quad ] = 0, \quad ] ] \quad (8.71)$$

(1) + \_\_\_\_\_ [

$$+ \quad - \quad (\theta) + \quad - \quad [ \quad ] \quad (8, 72)$$

(8,73)

Cuối cùng trừ đi , , [ [ | ] = 1, , , ] [ | ] = 0, , , ]]:

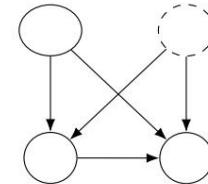
$$\text{Xu hướng} = [ [ | ] = 1, , ] [ | ] = 0, , ] \\ , [ [ | ] = 1, , , ] [ | ] = 0, , , ] \quad (8.74)$$

$$= + \frac{\dots}{\dots} \quad (8.75)$$

$$= \frac{\dots}{\dots} \quad (8.76)$$

□

Tổng quát hóa cho các đồ thị/ước tính tùy ý. Ở đây, chúng tôi đã thực hiện phân tích độ nhạy cho ATE đối với cấu trúc đồ thị đơn giản trong Hình 8.4. Đối với các ước tính tùy ý trong đồ thị tùy ý, trong đó cấu trúc phương trình là tuyến tính, xem Cinelli et al. [61].



Hình 8.4: Cấu trúc nhân quả đơn giản trong đó là yếu tố gây nhiễu được quan sát và là những yếu tố gây nhiễu không được quan sát.

Đồ thị đường viền nhạy cảm

Bởi vì Mệnh đề 8.8 cho chúng ta một biểu thức dạng đóng cho độ lệch trong dựa trên các tham số gây nhiễu không quan sát được và , chúng ta có thể vẽ đồ thị mức độ sai lệch trong đồ thị đường viền. Chúng tôi thể hiện điều này trong Hình 8.5a, trong đó chúng tôi có trên trục x và trên trục y.

Nếu chúng ta sắp xếp lại phương trình 8.7315 để giải cho , chúng ta thu được kết quả sau:

$$= [ [ | ] = 1, , ] [ | ] = 0, , ] \quad (8.77)$$

Vì vậy, với các giá trị cho trước của và , chúng ta có thể tính ATE thực ,

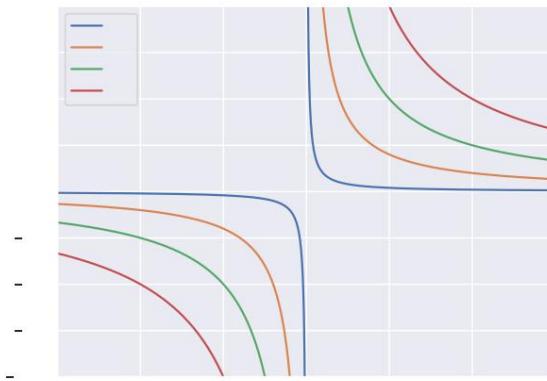
từ đại lượng quan sát [ [ | ] = 1, , ] [ | ] = 0, , ]. Điều này cho phép chúng ta có được các đường cong độ nhạy cho phép chúng ta biết cách

những kết luận chắc chắn như “ [ [ | ] = 1, , ] [ | ] = 0, , ] = 25 là tích cực, vì vậy có khả năng tích cực” là yếu tố gây nhiễu không quan sát được. Chúng tôi âm mưu các đường viền liên quan như vậy của trong Hình 8.5b.

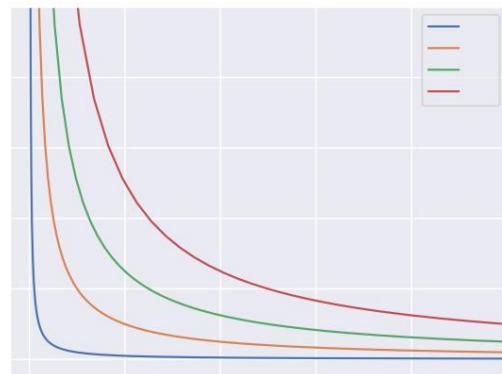
15 Nhớ lại phương trình 8.73:

$$[ [ | ] = 1, , ] [ | ] = 0, , ]$$

$$= + \frac{\dots}{\dots} \quad (8.73 \text{ xem lại})$$



(a) Đường viền của độ lệch nhiễu



(b) Đường nét của ATE thực , cho rằng

$$[ [ | ] = 1, , ] [ | ] = 0, , ] = 25$$

Hình 8.5: Đồ thị đường viền cho độ nhạy trong đó trục x của cả hai là các đường  $\frac{1}{y}$  và trục y là . Có sự tương ứng về mã màu giữa

cong ở phía trên bên phải của Hình 8.5b và các đường cong trong Hình 8.5

Trong ví dụ mà chúng tôi mô tả ở Hình 8.5, hình vẽ cho chúng ta biết rằng đường cong màu xanh lá cây (thứ ba từ dưới lên/trái) cho biết mức độ nhiễu cần phải mạnh đến mức nào để giải thích hoàn toàn mối liên hệ được quan sát.

Nói cách khác,  $(\frac{1}{| | })$  cần đủ lớn để rời trên mặt cỏ

$$\text{đường cong trở lên để ATE thực sự bằng } 0 \text{ hoặc ngược dấu với } [ | ] = 1, [ | ] = 0, [ | ] = 25.$$

### 8.2.2 Cài đặt chung khác

Chúng tôi xem xét cài đặt tuyến tính đơn giản trong Phần 8.2.1 để dễ dàng truyền đạt các khái niệm quan trọng trong phân tích độ nhạy. Tuy nhiên, hiện có cho phép chúng tôi thực hiện phân tích độ nhạy trong các cài đặt tổng quát hơn.

Giả sử chúng ta đang ở trong cài đặt chung trong đó là nhị phân. Đây không phải là trường hợp trong phần trước (xem phương trình 8.61).

và Imbens [63] <sup>16</sup> Rosenbaum và Rubin [62] xem xét một cài đặt xử lý nhị phân đơn giản với nhị phân bằng cách chỉ đặt hàm sigmoid logistic xung quanh về phái của Công thức 8.61 và sử dụng hàm đó cho xác suất xử lý thay vì giá trị thực tế của việc xử lý:

$$( | , ) := \frac{1}{1 + \exp(( + ))} \quad (8,78)$$

Không có giả định nào về hoặc May mắn thay, chúng ta có thể loại bỏ rất nhiều giả định mà chúng ta đã thấy cho đến nay. Không giống như dạng tuyến tính mà chúng tôi giả định cho trong Phần 8.2.1 và dạng tuyến tính mà Rosenbaum, Rubin [62] và Imbens [63] giả định, Cinelli và Hazlett [64] phát triển một phương pháp phân tích độ nhạy không thể xác định được đối với hàm số. dạng của . Phương pháp của họ cũng cho phép là không nhị phân và là một vectơ, thay vì chỉ là một yếu tố gây nhiễu duy nhất không được quan sát.

Các mô hình học máy tùy ý để tham số hóa và Hãy nhớ lại rằng tất cả các công cụ ước tính mà chúng ta đã xem xét trong Chương 7 đều cho phép chúng ta kết hợp các mô hình học máy tùy ý để có được các công cụ ước tính được hỗ trợ bởi mô hình. Có thể sẽ hấp dẫn nếu có một lựa chọn tương tự trong phân tích độ nhạy, có khả năng sử dụng các mô hình giống hệt nhau cho mô hình kết quả có điều kiện và điểm xu hướng mà chúng tôi đã sử dụng để ước tính.

Và đây chính xác là những gì Veitch và Zaveri [65] mang lại cho chúng ta. Và họ thậm chí còn có thể rút ra một biểu thức dạng đóng cho sai lệch gây nhiễu, giả sử các mô hình chúng tôi sử dụng cho và được xác định rõ ràng, điều mà Rosenbaum và Rubin [62] và Imbens [63] đã không làm được trong mô hình đơn giản của họ. cài đặt.

Chú ý; Có rất nhiều lựa chọn Mặc dù chúng tôi chỉ nhấn mạnh một vài lựa chọn ở trên, nhưng có nhiều cách tiếp cận khác nhau để phân tích độ nhạy và mọi người không đồng ý về cách nào là tốt nhất. Điều này có nghĩa là phân tích độ nhạy là một lĩnh vực hoạt động tích cực của nghiên cứu hiện nay.

Xem Liu và cộng sự. [66] để xem xét các phương pháp trước năm 2013.

Rosenbaum là một nhân vật quan trọng khác trong phân tích độ nhạy với một số cách tiếp cận khác nhau của ông [67-69]. Dưới đây là danh sách chưa đầy đủ về một số phương pháp phân tích độ nhạy linh hoạt khác mà bạn có thể muốn xem xét: Franks et al. [70], Yadlowsky và cộng sự. [71], Vanderweele và Arah [72], và Ding và VanderWeele [73].

$$:= + \quad (\text{xem lại 8.61})$$

[62]: Rosenbaum và Rubin (1983), 'Đánh giá độ nhạy đối với một hệ nhị phân không được quan sát

Hiệp phương sai trong một nghiên cứu quan sát với Kết quả nhị phân'

[63]: Imbens (2003), 'Độ nhạy cảm với các giả định ngoại sinh trong đánh giá chương trình-Sự'

<sup>16</sup> Imbens [63] là người đầu tiên giới thiệu đồ thị đường viền giống như đồ thị trong Hình 8.5 của chúng tôi.

[64]: Cinelli và Hazlett (2020), 'Tạo cảm giác nhạy cảm: mở rộng độ lệch biến bị bỏ qua'

[65]: Veitch và Zaveri (2020), Ý thức và Phân tích độ nhạy: Phân tích hậu nghiệm đơn giản về sai lệch do nhiễu không quan sát được

[66]: Liu và cộng sự. (2013), 'Giới thiệu về phân tích độ nhạy đối với yếu tố gây nhiễu không quan sát được trong nghiên cứu phòng ngừa phi thực nghiệm'

[67]: Rosenbaum (2002), Quan sát Học

[68]: Rosenbaum (2010), Thiết kế nghiên cứu quan sát

[69]: Rosenbaum (2017), Quan sát và Cuộc thi nghiên

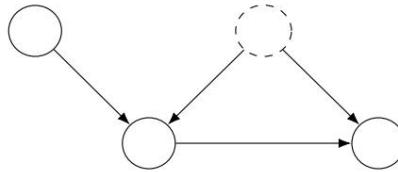
[70]: Franks và cộng sự. (2019), 'Phân tích độ nhạy linh hoạt cho các nghiên cứu quan sát không có hàm ý có thể quan sát được' [71]: Yadlowsky et al. (2020), Giới hạn về hiệu quả điều trị có điều kiện và trung bình với các yếu tố gây nhiễu

không được quan sát [72]: Vanderweele và Arah (2011), 'Công thức sai lệch để phân tích độ nhạy của yếu tố gây nhiễu không do lưỡng được đối với kết quả chung, phương pháp điều trị và các yếu tố gây nhiễu' [73] : Ding và VanderWeele (2016), 'Phân tích độ nhạy không có giả định'

# Biến công cụ | 9

Làm thế nào chúng ta có thể xác định được những tác động nhân quả khi chúng ta có mặt gây nhiễu không quan sát được? Một cách phổ biến là tìm và sử dụng công cụ biến. Một nhạc cụ (biến công cụ) có ba đặc tính chính.

Nó ảnh hưởng đến việc điều trị, nó chỉ ảnh hưởng đến thông qua, và ảnh hưởng của trên là vô căn cứ. Chúng tôi mô tả những phẩm chất này trong Hình 9.1. Những cái này phẩm chất cho phép chúng ta sử dụng để tách biệt mỗi liên hệ nhân quả xuất phát từ đến. Trực giác cho thấy những thay đổi trong sẽ được phản ánh trong và dẫn đến với những thay đổi tương ứng trong. Và những thay đổi đặc biệt tập trung vào này là không có căn cứ (không giống như những thay đổi của gây ra bởi những yếu tố không quan sát được gây nhiễu), vì vậy chúng cho phép chúng ta tách biệt mỗi liên hệ nhân quả chảy từ đến.



Hình 9.1: Đồ thị trong đó là yếu tố gây nhiễu không quan sát được về tác động của lên và là một biến công cụ.

## 9.1 Dụng cụ là gì?

Có ba giả định chính phải được thỏa mãn đối với một biến được coi là một công cụ. Đầu tiên là phải phù hợp trong cảm giác rằng nó phải ảnh hưởng đến.

**Giả định 9.1 (Mức độ liên quan)** có tác động nhân quả lên

Về mặt đồ họa, giả định liên quan tương ứng với sự tồn tại của một cạnh tích cực từ đến trong đồ thị nhân quả. Giả định thứ hai là được gọi là hạn chế loại trừ.

**Giả định 9.2 (Hạn chế loại trừ)** tác động nhân quả lên là hoàn toàn được trung gian bởi

Giả định này được gọi là hạn chế loại trừ vì nó loại trừ từ phương trình cấu trúc của và từ bất kỳ cấu trúc nào khác các phương trình sẽ làm cho mỗi liên hệ nhân quả đi từ đến mà không có đi qua. Về mặt đồ họa, điều này có nghĩa là chúng tôi đã loại trừ đủ các cạnh tiềm năng giữa các biến trong biểu đồ nhân quả sao cho tất cả các nhân quả đường đi từ đến đi qua. Cuối cùng, chúng tôi giả định rằng tác động nhân quả của trên là vô căn cứ:

9.1 Dụng cụ là gì? . . . . .	86
9.2 Không nhận dạng phi tham số của ATE . . . . .	87
9.3 Khởi động: Thiết lập tuyến tính nhẹ	
phân . . . . .	87
9.4 Cài đặt tuyến tính liên tục	88
9.5 Nhận dạng không tham số	
của ATE cục bộ . . . . .	90
Kỹ hiệu tiềm năng mới với	
Nhạc cụ . . . . .	90
Sự phân tầng chính . . . . .	90
ATE địa phương . . . . .	91
9.6 Cài đặt chung khác cho	
Nhận dạng ATE . . . . .	94

**Giả định 9.3 (Tính vô căn cứ của công cụ)** Không có đường dẫn cửa sau từ đến .

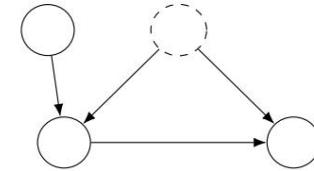
Các công cụ có điều kiện Chúng tôi coi Giả định 9.3 là tính vô căn cứ vô điều kiện, nhưng tất cả các phép toán cho các biến công cụ vẫn hoạt động nếu chúng ta cũng có tính vô căn cứ có điều kiện đối với các biến quan sát . Chúng ta chỉ cần đảm bảo rằng chúng ta có điều kiện dựa trên những biến số có liên quan đó. Trong trường hợp này, bạn có thể thấy được gọi là công cụ điều kiện.

## 9.2 Không nhận dạng phi tham số của ATE

Bạn có thể thấy märk “nếu các biến công cụ cho phép chúng ta xác định các tác động nhân quả thì tại sao chúng ta không thấy chúng trong Chương 6 Nhận dạng phi tham số?” Câu trả lời là các biến công cụ không xác định được tác động nhân quả một cách phi tham số . Chúng ta có nhận dạng phi tham số khi không phải đưa ra bất kỳ giả định nào về dạng tham số. Với các biến công cụ, chúng ta phải đưa ra các giả định về dạng tham số (ví dụ tuyến tính) để xác định các tác động nhân quả.

Chúng ta đã thấy điều kiện cần thiết hữu ích sau đây để nhận dạng phi tham số trong Phần 6.3: Đối với mỗi đường dẫn cửa sau từ đến bất kỳ con nào là tổ tiên của , đều có thể chặn đường dẫn đó [18, tr. 92]. Và chúng ta có thể thấy trong Hình 9.2 có một đường dẫn cửa sau từ đến không thể bị chặn:

. Vì vậy, điều kiện cần thiết này cho chúng ta biết rằng chúng ta không thể sử dụng công cụ để xác định một cách phi tham số tác động của lên .



Hình 9.2: Đồ thị trong đó là yếu tố gây nhiễu không quan sát được về tác động của lên và là biến công cụ.

[18]: Ngọc Trai (2009), Nhân Quả

## 9.3 Khởi động: Thiết lập tuyến tính nhị phân

Để khởi động, chúng ta sẽ bắt đầu trong cài đặt trong đó và là nhị phân và trong đó chúng ta đưa ra giả định tham số rằng là hàm tuyến tính của và :

**Giả định 9.4 (Kết quả tuyến tính)**

$$:= + \quad (9.1)$$

Việc không xuất hiện trong phương trình 9.1 là hệ quả của hạn chế loại trừ (Giả định 9.2).

Sau đó, với giả định này, chúng tôi sẽ cố gắng xác định tác động nhân quả . Bởi vì chúng ta có trực giác rằng sẽ hữu ích trong việc xác định tác động của lên , nên chúng ta sẽ bắt đầu với sai phân liên kết cho mỗi quan hệ - :  $[ | = 1] [ | = 0]$ . Áp dụng ngay Giả thiết 9.4, ta có:

$$[ | = 1] [ | = 0] \quad (9.2)$$

$$= [ | + | = 1] [ | + | = 0] \quad (9.3)$$

Sử dụng tuyến tính của kỳ vọng và sắp xếp lại một chút:

$$= (\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}) + (\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}) \quad (9.4)$$

Bây giờ, chúng tôi sử dụng giả định về tính vô căn cứ của công cụ (Giả định 9.3). Điều này có nghĩa là và độc lập, cho phép chúng ta loại bỏ số hạng :

$$= (\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}) + (\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}) \quad (9.5)$$

$$= (\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}) \quad (9.6)$$

Sau đó, chúng ta có thể giải để có được ước lượng Wald:

#### Mệnh đề 9.1

$$= \frac{\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}}{\underset{[ \quad | \quad = 1]}{\text{---}} + \underset{[ \quad | \quad = 0]}{\text{---}}} \quad (9.7)$$

Do Giả định 9.1, chúng ta biết rằng mẫu số khác 0, nên về phải không phải là không xác định. Sau đó, chúng tôi chỉ cần đưa các phương tiện thực nghiệm vào vị trí của những kỳ vọng có điều kiện này để có được công cụ ước tính Wald [74]:

$$\begin{array}{c} \frac{1}{1} & : & +1 & \frac{1}{0} \\ \hline \frac{1}{1} & : & +1 & \frac{1}{0} \end{array} \quad (9.8)$$

trong đó 1 là số lượng mẫu trong đó = 1 và 0 là số lượng mẫu trong đó = 0.

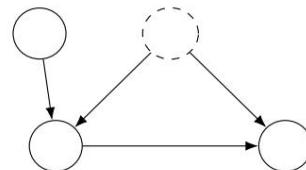
Hiệu ứng Nhân quả dưới dạng Hệ số Đường nhân Khi các phương trình cấu trúc là tuyến tính, bạn có thể coi mỗi liên hệ nhân quả đi từ biến đến biến là tích của các hệ số đọc theo đường có hướng từ đến . Nếu có nhiều đường dẫn, bạn chỉ cần tính tổng các mối liên hệ nhân quả đọc theo tất cả các đường dẫn đó. Tuy nhiên, chúng ta không có quyền truy cập trực tiếp vào mỗi liên hệ nhân quả. Thay vào đó, chúng tôi có thể do lường tổng liên kết và các đường dẫn cửa sau không bị chặn cũng góp phần vào tổng liên kết, đó là lý do tại sao  $[ \quad | \quad = 1] \quad [ \quad | \quad = 0]$  . Vậy làm thế nào chúng ta có thể xác định được tác động của

lên trong Hình 9.3? Bởi vì không có đường dẫn cửa sau từ công cụ đến , nên chúng ta có thể xác định một cách tầm thường tác động của lên :  $[ \quad | \quad = 1] \quad [ \quad | \quad = 0] =$  . Tương tự, chúng ta có thể xác định tác động của công cụ lên :  $[ \quad | \quad = 1] \quad [ \quad | \quad = 0] =$  .

Khi đó, chúng ta có thể chia tác dụng của trên cho tác dụng của — . Trên để xác định Và thường này chính xác là ước lượng Wald trong Mệnh đề 9.1.

[74]: Wald (1940), 'Sự phù hợp của đường thẳng Các dòng nếu cả hai biến đều có lỗi'

**Bài tập đọc tích cực:** Chúng ta đã sử dụng từng Giả định 9.1 đến 9.4 ở đâu trong đoạn hàm trên của Phương trình 9.7.



Hình 9.3: Đồ thị trong đó là yếu tố gây nhiễu không quan sát được về tác động của lên và là biến công cụ.

## 9.4 Cài đặt tuyến tính liên tục

Bây giờ chúng ta sẽ xem xét cài đặt trong đó và liên tục, thay vì nhị phân. Chúng ta vẫn sử dụng tuyến tính của (Giả định 9.4), có nghĩa là tác động nhân quả của lên là . Trong sự liên tục cài đặt, chúng ta có được sự tương tự liên tục tự nhiên của ước tính Wald:

## Mệnh đề 9.2

$$= \frac{\text{Cov}(\quad, \quad)}{\text{Cov}(\quad, \quad)} \quad (9.9)$$

Bằng chứng. Giống như chúng ta đã bắt đầu với  $[ \quad | \quad = 1] = [ \quad | \quad = 0]$  ở phần trước, ở đây, chúng ta sẽ bắt đầu với  $\text{Cov}(\quad, \quad)$  tương tự liên tục. Chúng ta bắt đầu với một đồng nhất thức hiệp phương sai cỗ điển:

$$\text{Cov}(\quad, \quad) = [ \quad ] - [ \quad ] [ \quad ] \quad (9.10)$$

Sau đó, áp dụng giả định kết quả tuyến tính (Giả định 9.4):

$$= [(\quad + \quad)] - [\quad + \quad] [ \quad ] \quad (9.11)$$

Phân phối và sắp xếp lại:

$$= \dots [ \quad ] - [ \quad ] [ \quad ] + (\dots [ \quad ] - [ \quad ] [ \quad ]) \quad (9.13)$$

Bây giờ, chúng ta thấy rằng chúng ta có thể áp dụng lại cùng một nhận dạng hiệp phương sai:

$$= \text{Cov}(\quad, \quad) + \text{Cov}(\quad, \quad) \quad (9.14)$$

Và  $\text{Cov}(\quad, \quad) = 0$  theo giả định về tính vô căn cứ của công cụ (Giả định 9.3):

$$= \text{Cov}(\quad, \quad) \quad (9.15)$$

Cuối cùng, chúng tôi giải quyết cho :

$$= \frac{\text{Cov}(\quad, \quad)}{\text{Cov}(\quad, \quad)} \quad (9.16)$$

trong đó giả định liên quan (Giả định 9.1) cho chúng ta biết rằng mẫu số khác 0.  $\square$

Điều này dẫn chúng ta đến công cụ ước tính tự nhiên sau, tương tự như công cụ ước tính Wald:

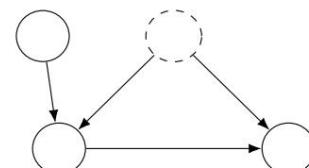
$$\hat{=} \frac{\text{Cov}(\quad, \quad)}{\text{Cov}(\quad, \quad)} \quad (9.17)$$

Một công cụ ước tính tương đương khác được gọi là công cụ ước tính bình phương nhỏ nhất hai giai đoạn (2SLS). Hai giai đoạn như sau:

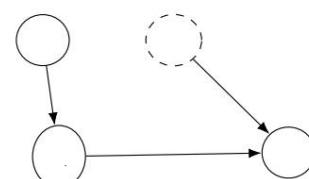
1. Hồi quy tuyến tính trên để ước tính  $[ \quad | \quad ]$ . Điều này cho chúng ta phép chiếu của lên :
2. Hồi quy tuyến tính  $\hat{}$  để ước tính  $[ \quad | \quad ]$ . Lấy ước tính của chúng trên dưới dạng hệ số phù hợp ở phía trước  $\hat{}$ .

Có trực giác hữu ích đi kèm với công cụ ước tính 2SLS. Để thấy điều này, hãy bắt đầu với biểu đồ biến công cụ chuẩn mà chúng ta đang sử dụng (Hình 9.4). Ở giai đoạn một, chúng ta chiếu lên để có được  $\hat{}$  chỉ là hàm của :  $\hat{=} \hat{[ \quad | \quad ]}$ . Sau đó, hãy tưởng tượng một đồ thị trong đó  $\hat{}$  được thay thế bằng  $\hat{}$  (Hình 9.5). Vì  $\hat{}$  không phải là hàm của  $\hat{}$  nên chúng ta có thể nghĩ đến việc loại bỏ cạnh  $\hat{}$  trong biểu đồ này. Bây giờ, vì không có đường dẫn của sau

**Bài tập đọc tích cực:** Chúng ta đã sử dụng giả định hạn chế loại trừ (Giả định 9.2) ở đâu trong chứng minh này?



Hình 9.4: Đồ thị trong đó  $\hat{}$  là không quan sát là yếu tố gây nhiễu về tác động của  $\hat{}$  lên và  $\hat{}$  là một biến công cụ.



Hình 9.5: Phiên bản mở rộng của Hình 9.4, trong đó  $\hat{}$  được thay thế bằng  $\hat{}$  =  $\hat{[ \quad | \quad ]}$ , không thuộc vào  $\hat{}$ , do đó nó không còn có cạnh đi vào từ  $\hat{}$  nữa.

từ ^ đến , chúng ta có thể nhận được mối liên hệ đó là quan hệ nhân quả ở giai đoạn hai, có thể đơn giản hỏi quy trong đó ^ để ước tính hệ quả nhân quả. Lưu ý: Chúng ta cũng khi sử dụng 2SLS trong cài đặt nhị phân mà chúng ta đã thảo luận trong Phần 9.3.

## 9.5 Nhận dạng không tham số của ATE cục bộ

Vấn đề với hai phần trước là chúng ta đã đưa ra giả định tham số mạnh về tính tuyến tính (Giả định 9.4). Ví dụ, giả định này đòi hỏi tính đồng nhất (hiệu quả can thiệp là như nhau đối với mọi đơn vị). Có những biến thể khác mã hóa giả định về tính đồng nhất (ví dụ, xem Hernán và Robins [7, Phần 16.3]), và chúng đều là những giả định mạnh. Lý tưởng nhất là chúng ta có thể sử dụng các biến công cụ để nhận dạng mà không đưa ra bất kỳ giả định tham số nào như tính tuyến tính hoặc tính đồng nhất. Và chúng tôi có thể. Chúng ta chỉ cần giải quyết một ước tính nhân quả cụ thể hơn ATE và hoán đổi giả định tuyến tính bằng một giả định mới. Chúng tôi sẽ thực hiện việc này trong cài đặt nhị phân, vì vậy cả và đều là nhị phân. Trước khi có thể làm điều đó, chúng ta phải định nghĩa một số ký hiệu mới trong Phần 9.5.1 và giới thiệu sự phân tầng chính trong Phần 9.5.2.

[7]: Hernán và Robins (2020), Suy luận nhân quả :  
Chuyện gì xảy ra nếu

### 9.5.1 Ký hiệu tiềm năng mới với các dụng cụ

Giống như chúng ta sử dụng  $(1) (\cdot = 1)$  để biểu thị kết quả tiềm năng mà chúng ta sẽ quan sát được nếu chúng ta thực hiện điều trị và  $(0) (\cdot = 0)$  để biểu thị kết quả tiềm năng mà chúng ta sẽ quan sát được nếu chúng ta thực hiện điều trị. để không xử lý, chúng tôi sẽ xác định ký hiệu tiềm năng tương tự bằng các công cụ.

Chúng ta sẽ coi công cụ là sự khuyến khích cho việc điều trị, vì vậy nếu chúng ta có  $= 1$ , chúng ta được khuyến khích thực hiện điều trị và nếu chúng ta có  $= 0$ , chúng ta được khuyến khích không thực hiện điều trị. Đặt  $(1) (\cdot = 1)$  biểu thị cách xử lý mà chúng ta sẽ thực hiện nếu nhận được giá trị công cụ 1. Tương tự, hãy để  $(0) (\cdot = 0)$  biểu thị cách xử lý mà chúng ta sẽ thực hiện nếu chúng ta nhận được giá trị công cụ .

Sau đó, chúng ta cũng có kết quả tương tự đối với các kết quả tiềm năng khi chúng ta can thiệp vào công cụ chứ không phải vào phương pháp điều trị:  $(\cdot = 1)$  biểu thị kết quả mà chúng ta sẽ quan sát được nếu chúng ta được khuyến khích thực hiện phương pháp điều trị và  $(\cdot = 0)$  biểu thị kết quả mà chúng ta sẽ quan sát được nếu chúng ta được khuyến khích không thực hiện phương pháp điều trị.

### 9.5.2 Sự phân tầng chính

Chúng tôi sẽ phân chia dân số thành bốn tầng chính, dựa trên mối quan hệ giữa sự khuyến khích và cách điều trị được thực hiện . Có bốn tầng vì có một tầng cho mỗi sự kết hợp của đánh giá các biến nhị phân và có thể đảm nhận.

#### Định nghĩa 9.1 (Tầng chính)

- Những người tuân thủ - luôn thực hiện cách đối xử mà họ được khuyến khích thực hiện. Cụ thể là,  $(1) = 1$  và  $(0) = 0$ .

2. Luôn chấp nhận - luôn chấp nhận sự điều trị, bất kể sự khuyến khích.  
Cụ thể là,  $(1) = 1$  và  $(0) = 1$ .
3. Không bao giờ chấp nhận - không bao giờ thực hiện điều trị, bất kể sự khuyến khích.  
Cụ thể là,  $(1) = 0$  và  $(0) = 0$ .
4. Những người thách thức - luôn áp dụng phương pháp điều trị ngược lại với phương pháp điều trị mà họ được khuyến khích thực hiện. Cụ thể là,  $(1) = 0$  và  $(0) = 1$ .

Đồ thị nhân quả khác nhau Điều quan trọng là các tầng này có đồ thị nhân quả khác nhau. Trong khi cách đối xử mà những người tuân thủ và chống đối áp dụng phụ thuộc vào sự khuyến khích (công cụ), thì cách đối xử mà những người luôn chấp nhận và không bao giờ chấp nhận lại không phụ thuộc vào sự đối xử đó. Do đó, người tuân thủ và người thách thức có đồ thị nhân quả bình thường (Hình 9.6), trong khi người luôn nhận và người không bao giờ nhận có cùng đồ thị nhân quả nhưng loại bỏ cạnh (Hình 9.7). Điều này có nghĩa là tác động nhân quả của lên là bằng 0 đối với người luôn nhận và người không bao giờ nhận. Khi đó, do hạn chế loại trừ, điều này có nghĩa là tác động nhân quả của lên là bằng 0 đối với người luôn nhận và không bao giờ nhận. Điều này sẽ rất quan trọng cho việc dẫn xuất sắp tới.

Không thể xác định được tầng Với một số giá trị quan sát được của và , chúng ta thực sự không thể xác định được mình đang ở tầng nào. Có bốn kết hợp của các biến nhị phân và ; đối với mỗi kết hợp này, chúng tôi sẽ lưu ý rằng có nhiều tầng tương thích với các kết hợp giá trị được quan sát.

1. ???? \_ \_ \_ \_ \_ hoặc không bao giờ nhận 4.  
 $= 1$ ,  $= 1$ . Tầng lớp tương thích: người tuân thủ hoặc luôn nhận

Điều này có nghĩa là chúng tôi không thể xác định liệu một đơn vị nhất định là người tuân thủ, người thách thức, người luôn nhận hay không bao giờ nhận.

### 9.5.3 ATE cục bộ

Mặc dù chúng tôi sẽ không thể sử dụng các biến công cụ để xác định ATE một cách phi tham số khi có sự nhiễu loạn không quan sát được (Phần 9.2), nhưng chúng tôi sẽ có thể xác định một cách phi tham số cái được gọi là ATE cục bộ . Hiệu quả can thiệp trung bình cục bộ (LATE) còn được gọi là tác động nhân quả trung bình của người tuân thủ (CACE), vì đây là ATE của những người tuân thủ.

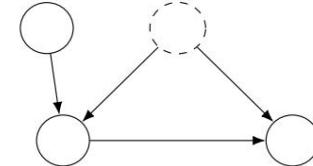
**Định nghĩa 9.2 (Hiệu quả can thiệp trung bình cục bộ (LATE) / Hiệu quả nhân quả trung bình của người tuân thủ (CACE))**

$$[ \quad (\quad = 1) \quad \quad (\quad = 0) \mid \quad (\quad = 1) = 1, \quad (\quad = 0) = 0 ] \quad (9.18)$$

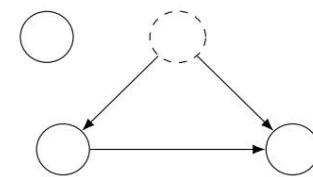
Để xác định LATE, mặc dù chúng ta sẽ không cần giả định tuyến tính nữa (Giả định 9.4), chúng ta sẽ cần đưa ra một giả định mới được gọi là tính đơn điệu.

**Giả định 9.5 (Tính đơn điệu)**

$$, \quad (\quad = 1) \geq \quad (\quad = 0) \quad (9.19)$$



Hình 9.6: Biểu đồ nhân quả của người tuân thủ và người thách thức.



Hình 9.7: Biểu đồ nhân quả của người luôn nhận và không bao giờ nhận.

**Bài tập đọc tích cực:** Đảm bảo rằng bạn hiểu được lý do tại sao đây là các tầng tương thích cho từng tổ hợp giá trị được quan sát này.

Tính đơn điệu có nghĩa là nếu chúng ta được khuyến khích thực hiện điều trị ( $\alpha = 1$ ), chúng ta có nhiều khả năng hoặc có khả năng thực hiện điều trị nhiều hơn nếu chúng ta được khuyến khích không thực hiện điều trị ( $\alpha = 0$ ).

Điều quan trọng là điều này có nghĩa là chúng ta đang giả định rằng không có người thách thức. Điều này là do những người tuân thủ thỏa mãn  $\alpha(1) > \alpha(0)$ , những người luôn chấp nhận và không bao giờ chấp nhận thỏa mãn  $\alpha(1) = \alpha(0)$ , nhưng những người thách thức không thỏa mãn cả hai điều này; trong số những người thách thức,  $\alpha(1) < \alpha(0)$ , điều này vi phạm giả định về tính đơn điệu.

Bây giờ chúng tôi đã giới thiệu các khái niệm chính về tầng lớp chính và giả định về tính đơn điệu. Điều quan trọng là chúng ta đã thấy rằng tác động nhân quả của đối với  $\alpha$  là bằng 0 đối với những người luôn chấp nhận và không bao giờ chấp nhận (Phần 9.5.2), và chúng ta vừa thấy rằng giả định tính đơn điệu ngụ ý rằng không có người thách thức. Với ý nghĩ này, giờ đây chúng ta đã sẵn sàng rút ra kết quả nhận dạng phi tham số cho ước tính LATE.

**Định lý 9.3 (Nhận dạng không tham số LATE)** Cho rằng  $\alpha$  là một công cụ, và là các biến nhị phân và tính đơn điệu đó đúng, điều sau đây đúng:

$$\alpha = [\alpha(1) - \alpha(0) | \alpha(1) = 1, \alpha(0) = 0] = \frac{[\alpha(1) - \alpha(0) | \alpha(1) = 1]}{[\alpha(1) - \alpha(0) | \alpha(1) = 0]} = 0 \quad (9.20)$$

Bằng chứng. Bởi vì chúng ta quan tâm đến tác động nhân quả của  $\alpha$  lên  $\alpha$  và vì biết rằng chúng ta sẽ sử dụng công cụ  $\alpha$ , nên chúng ta sẽ bắt đầu với tác động nhân quả của  $\alpha$  lên  $\alpha$  và phân tách nó thành các tác động nhân quả cụ thể theo tầng có trọng số bằng cách sử dụng định luật của tổng xác suất:

$$\begin{aligned} & [\alpha(1) - \alpha(0)] \\ &= [\alpha(1) - \alpha(0) | \alpha(1) = 1, \alpha(0) = 0] + [\alpha(1) - \alpha(0) | \alpha(1) = 0, \alpha(0) = 1] \\ &+ [\alpha(1) - \alpha(0) | \alpha(1) = 1, \alpha(0) = 1] + [\alpha(1) - \alpha(0) | \alpha(1) = 0, \alpha(0) = 0] \quad (9.21) \end{aligned}$$

Số hạng đầu tiên tương ứng với những người tuân thủ, số hạng thứ hai tương ứng với những người thách thức, số hạng thứ ba tương ứng với những người luôn chấp nhận và số hạng cuối cùng tương ứng với những người không bao giờ chấp nhận. Như chúng ta đã thảo luận trong Phần 9.5.2, tác động nhân quả của  $\alpha$  lên  $\alpha$  giữa những người luôn nhận và không bao giờ nhận là bằng 0, vì vậy chúng ta có thể loại bỏ những số hạng đó.

$$\begin{aligned} &= [\alpha(1) - \alpha(0) | \alpha(1) = 1, \alpha(0) = 0] + [\alpha(1) - \alpha(0) | \alpha(1) = 0, \alpha(0) = 1] \quad (9.22) \end{aligned}$$

Bởi vì chúng ta đã đưa ra giả định về tính đơn điệu, nên chúng ta biết rằng không có phần tử thách thức nào ( $\alpha(1) = 0, \alpha(0) = 1$ ), do đó số hạng phần tử thách thức cũng bằng 0.

$$= [\alpha(1) - \alpha(0) | \alpha(1) = 1, \alpha(0) = 0] + [\alpha(1) - \alpha(0) | \alpha(1) = 0, \alpha(0) = 0] \quad (9.23)$$

Bây giờ, nếu chúng ta giải quyết tác động này của  $\alpha$  lên  $\alpha$  trong số những người tuân thủ, chúng ta nhận được

Trình tựa thủ:  $(1) = 1, (0) = 0$

Người luôn nhận:  $(1) = 1, (0) = 1$

Người không bao giờ nhận:  $(1) = 0, (0) = 0$

Các yếu tố thách thức:  $(1) = 0, (0) = 1$

sau đây:

$$0) [ ( = 1) ( = 0) | (1) = 1, (0) = \frac{[ ( = 1) ( = 0)]}{0} = (1) = 1, (0) = 0) \quad (9.24)$$

Và bởi vì đây là những người tuân thủ, những người sẽ thực hiện bất kỳ phương pháp điều trị nào mà họ được khuyến khích thực hiện,  $( = 1)$  và  $( = 0)$  thực sự bằng  $( = 1)$  và  $( = 0)$ , tương ứng, vì vậy chúng ta có thể thay đổi vé trái của phương trình 9.24 thành LATE, ước tính nhân quả mà chúng ta đang cố gắng xác định:

$$[ ( = 1) ( = 0) | (1) = 1, (0) = 0] \quad (9.25)$$

$$= \frac{[ ( = 1) ( = 0)]}{0} = (1) = 1, (0) = 0 \quad (9.26)$$

Bây giờ, chúng ta áp dụng giả định về tính vô căn cứ công cụ (Giả định 9.3) để xác định tử số.

$$= \frac{[ | = 1]}{0} \quad (1) = 1, (0) = 0 \quad (9.27)$$

Tất cả những gì còn lại là xác định mẫu số, xác suất trở thành người tuân thủ. Tuy nhiên, chúng tôi đã đề cập rằng chúng tôi không thể xác định những người tuân thủ trong Phần 9.5.2, vậy chúng tôi có thể thực hiện việc này bằng cách nào? Đây là lúc chúng ta cần phải khéo léo một chút. Chúng ta sẽ có được xác suất này bằng cách lấy tất cả mọi người (xác suất 1) và trừ đi những người luôn chấp nhận và những người tuân thủ, vì không có người thách thức, do tính đơn điệu (Giả định 9.5).

$$= \frac{[ | = 1]}{0} \quad (1) = 0 | = 1) \quad ( = 1 | = 0) \quad (9.28)$$

Để hiểu làm thế nào chúng ta có được đẳng thức trên, hãy xem xét rằng mọi người đều có  $= 1$  hoặc  $= 0$ . Chúng ta có thể trừ đi tất cả những người không bao giờ nhận bằng cách loại bỏ những cá thể có  $= 0$  trong số  $= 1$  quần thể ( $( = 0 | = 1) = 1$ ). Tương tự, chúng ta có thể loại bỏ tất cả những người luôn nhận bằng cách loại bỏ những người có  $= 1$  trong số  $= 0$  ( $( = 1 | = 0) = 0$ ). Chúng ta biết rằng điều này sẽ loại bỏ tất cả những người không bao giờ nhận và luôn nhận vì không có người thách thức và vì chúng ta đã xem xét cả tiêu quần thể  $= 1$  và tiêu quần thể  $= 0$ . Bây giờ chúng ta chỉ cần thực hiện một chút thao tác:

$$= \frac{[ | = 1]}{0} \quad (1) = 0 | = 1) \quad ( = 1 | = 0) \quad (9.29)$$

$$= \frac{[ | = 1]}{0} \quad ( = 1 | = 1) \quad ( = 1 | = 0) \quad (9.30)$$

Cuối cùng, vì là biến nhị phân nên chúng ta có thể đổi xác suất  $= 1$  thành kỳ vọng:

$$= \frac{[ | = 1]}{[ | = 1]} \quad [ | = 0] \quad (9.31)$$

□

Đây chính xác là ước tính Wald mà chúng ta đã thấy trong thiết lập tuyến tính (Phần 9.3) trong Công thức 9.7. Tuy nhiên, lần này, đó là ước tính thống kê tương ứng của ATE cục bộ  $[ \cdot (\cdot = 1) \quad (\cdot = 0) | \cdot (1) = 1, \quad (\cdot = 0) = 0 ]$ , còn được gọi là tác động nhân quả trung bình của trình biên dịch (CACE). Ước tính nhân quả LATE/CACE này trái ngược với ước tính nhân quả ATE mà chúng ta đã thấy trong Phần 9.3:  $[ \cdot (\cdot = 1) \cdot (\cdot = 0) ]$ . Sự khác biệt là tác động nhân quả trung bình của người tuân thủ là ATE, đặc biệt trong nhóm nhỏ những người tuân thủ, chứ không phải trên tổng dân số. Nó mang tính chất cục bộ (MUỘN) đối với tiêu quần thể đó, thay vì mang tính toàn cầu đối với toàn bộ quần thể như ATE. Vì vậy, chúng ta đã thấy hai giả định khác nhau đưa chúng ta đến ước tính Wald với các biến công cụ:

1. Tính tuyến tính (hay nói chung hơn là tính đồng nhất)
2. Sự đơn điệu

Các vấn đề với LATE/CACE Tuy nhiên, có một số vấn đề với ước tính Wald cho LATE. Đầu tiên là sự đơn điệu có thể không được thỏa mãn trong bối cảnh mà bạn quan tâm. Thứ hai là, ngay cả khi tính đơn điệu được thỏa mãn, bạn có thể không quan tâm đến tác động nhân quả cụ thể giữa những người tuân thủ, đặc biệt là vì bạn thậm chí không thể xác định được những người tuân thủ là ai (xem Phần 9.5.2). Đúng hơn, ATE thông thường thường là một đại lượng hữu ích hơn để biết.

## 9.6 Cài đặt chung khác để nhận dạng ATE

Một cách thiết lập biến công cụ phổ biến hơn là xem xét rằng kết quả được tạo ra bởi một hàm phức tạp của việc điều trị và các biến đồng thời được quan sát cộng với một số yếu tố gây nhiễu bổ sung không được quan sát:

$$:= \cdot (\cdot, \cdot) + \quad (9,32)$$

Xem, ví dụ, Hartford và cộng sự. [75] và Xu và cộng sự. [76] về việc sử dụng deep learning để lập mô hình . Xem tài liệu tham khảo trong các bài viết đó để biết cách sử dụng các mô hình khác, chẳng hạn như phương pháp kernel cho mô hình Trong các mô hình đó và cho rằng nhập vào phương trình cấu trúc của một cách bổ sung, bạn có thể nhận dạng được bằng cách biến công cụ.

Ngoài ra, chúng ta có thể từ bỏ việc xác định điểm của các tác động nhân quả, thay vào đó giải quyết việc xác định tập hợp (nhận dạng một phần) và sử dụng các biến công cụ để đạt được giới hạn về các tác động nhân quả. Để biết thêm về điều đó, xem Pearl [18, Phần 8.2]. Ngoài ra, việc giải quyết việc xác định một tập hợp, thay vì một điểm, cho phép chúng ta giảm bớt nhiều cộng tính giả định ở trên trong phương trình 9.32. Ví dụ, Kilbertus và cộng sự. [77] xem xét cài đặt trong đó không nhập phương trình cấu trúc cho một cách bổ sung:

$$:= \cdot (\cdot, \cdot) \quad (9,33)$$

[75]: Hartford và cộng sự. (2017), 'Deep IV: Một cách tiếp cận linh hoạt để dự đoán phản ứng'.

[76]: Xu et al. (2020), 'Tìm hiểu các tính năng sâu trong hồi quy biến công cụ'

[18]: Ngoc Trai (2009), Nhân Quả

[77]: Kilbertus và cộng sự. (2020), 'Lớp thuật toán cho các mô hình biến công cụ chung'

# Sự khác biệt trong sự khác biệt 10

Lưu ý: chương sau khó khăn hơn nhiều so với bình thường và hiện tại không chứa nhiều số liệu và trực giác như tương ứng bài học.

## 10.1 Vòng sơ loại

Lần đầu tiên chúng tôi đưa ra giả định về tính vô căn cứ (Giả định 2.1) trong Chương 2:

$$( \quad (1), \quad (0) ) \quad (10.1)$$

Hãy nhớ lại rằng điều này tương đương với việc giả định rằng không có đường dẫn cửa sau từ đến trong biểu đồ nhân quả. Khi gặp trường hợp này, chúng ta có mối liên hệ đó là nhân quả. Nói cách khác, nó mang lại cho chúng ta nhận dạng sau đây (hy vọng là quen thuộc) của ATE:

$$[ \quad (1) \quad (0) ] = [ \quad (1) ] \quad [ \quad (0) ] \quad (10.2)$$

$$= [ \quad (1) \mid \quad = 1 ] \quad [ \quad (0) \mid \quad = 0 ] = \quad (10.3)$$

$$[ \quad | \quad = 1 ] \quad [ \quad | \quad = 0 ] \quad (10.4)$$

nơi chúng tôi đã sử dụng tính vô căn cứ này trong phương trình 10.3.

Tuy nhiên, ATE không phải là tác động nhân quả trung bình duy nhất mà chúng ta có thể quan tâm. Thông thường, các học viên quan tâm đến ATE đặc biệt ở tiêu chuẩn thể được điều trị. Đây được gọi là mức trung bình tác dụng điều trị lên đối tượng được điều trị (ATT):  $[ \quad (1) \quad (0) \mid \quad = 1 ]$ . Chúng ta có thể làm một giả định yếu hơn nếu chúng ta chỉ quan tâm đến ATT, hơn là

ẤN:

$$( \quad ) \quad (10.5)$$

Chúng ta chỉ cần giả sử rằng  $(\theta)$  ở đây không bị nhầm lẫn, thay vì giả định như vậy cả  $(0)$  và  $(1)$  đều vô căn cứ. Chúng tôi thể hiện điều này ở phần sau bằng chứng:

$$[ \quad (1) \quad (0) \mid \quad = 1 ] = [ \quad (1) \mid \quad = 1 ] \quad [ \quad (0) \mid \quad = 1 ] \quad (10.6)$$

$$= [ \quad | \quad = 1 ] \quad [ \quad (0) \mid \quad = 1 ] \quad (10.7)$$

$$= [ \quad | \quad = 1 ] \quad [ \quad (0) \mid \quad = 0 ] \quad (10.8)$$

$$= [ \quad | \quad = 1 ] \quad [ \quad | \quad = 0 ] \quad (10.9)$$

nơi chúng tôi sử dụng tính vô căn cứ yếu hơn này trong phương trình 10.8.

Nhìn chung, chúng tôi quan tâm đến ước tính ATT với những khác biệt trong khác biệt, nhưng chúng tôi sẽ sử dụng một giả định nhận dạng khác.

10.1 Sơ bộ . . . . .	95
10.2 Thời gian giới thiệu . . . . .	96
10.3 Nhận dạng . . . . .	96
Giả định . . . . .	96
Kết quả chính và bằng chứng . . . . .	97
10.4 Các vấn đề chính . . . . .	98

## 10.2 Thời gian giới thiệu

Bây giờ chúng ta sẽ giới thiệu chiềut thời gian. Việc sử dụng thông tin theo chiềut thời gian sẽ là chìa khóa để chúng ta có được nhận dạng mà không cần phải già định sự vô căn cứ thông thường. Chúng ta sẽ sử dụng cho biến thời gian.

Cài đặt Như thường lệ, chúng ta có nhóm can thiệp ( $\tau = 1$ ) và nhóm đổi chứng ( $\tau = 0$ ). Tuy nhiên, bây giờ cũng có thời gian và nhóm điều trị chỉ được điều trị sau một thời gian nhất định. Vì vậy, chúng ta có một khoảng thời gian  $\tau = 1$  biểu thị thời gian sau khi việc điều trị được thực hiện cho nhóm can thiệp và một khoảng thời gian  $\tau = 0$  biểu thị một khoảng thời gian trước khi việc điều trị được thực hiện cho nhóm điều trị. Bởi vì nhóm đổi chứng không bao giờ được điều trị nên nhóm đổi chứng chưa được điều trị tại một trong hai thời điểm  $\tau = 0$  hoặc tại thời điểm  $\tau = 1$ . Chúng ta sẽ biểu thị

biến ngẫu nhiên về kết quả tiềm năng khi điều trị tại thời điểm  $\tau$  là  $(\cdot)$ . Sau đó, ước tính nhận quả mà chúng ta quan tâm là sự khác biệt trung bình về kết quả có thể xảy ra sau khi thực hiện điều trị (trong khoảng thời gian  $\tau = 1$ ) trong nhóm điều trị:

$$[1(1) \quad 1(0) | \tau = 1] \quad (10.10)$$

Nói cách khác, chúng ta quan tâm đến ATT sau khi thực hiện điều trị.

## 10.3 Nhận dạng

### 10.3.1 Giả định

Bạn chỉ có thể coi  $\tau = 1$  và  $\tau = 0$  là hai biến ngẫu nhiên khác nhau. Vì vậy, mặc dù hiện tại chúng ta có chỉ số thời gian, chúng ta vẫn có nhận dạng tầm thường thông qua tính nhất quán (hãy nhớ lại Giả định 2.5) khi giá trị bên trong dấu ngoặc đơn cho kết quả tiềm năng khớp với giá trị điều kiện cho  $\tau$ :

**Giả định 10.1 (Tính nhất quán)** Nếu can thiệp là  $\cdot$  thì kết quả quan sát được tại thời điểm  $\tau$  là kết quả tiềm năng của can thiệp  $\cdot$ . chính thức,

$$\cdot, \tau = 1 = \cdot, \tau = 0 = (\cdot) \quad (10.11)$$

Chúng ta có thể viết điều này tương đương như sau:

$$\cdot, \tau = 1 = (\cdot) \quad (10.12)$$

Tính nhất quán là điều cho chúng ta biết rằng ước tính nhận quả  $[ \cdot (1) | \tau = 1 ]$  bằng ước tính thống kê  $[\cdot | \tau = 1]$ , và tương tự,  $[\cdot (0) | \tau = 0] = [\cdot | \tau = 0]$ . Ngược lại,  $[\cdot (1) | \tau = 0]$  và  $[\cdot (0) | \tau = 1]$  là các ước tính nhận quả phản thực, do đó tính nhất quán không trực tiếp xác định các đại lượng này cho chúng ta. Lưu ý: Trong các dân số của chúng tôi trong chương này, chúng tôi cũng ngầm giả định giả định không có nhiều (Giả định 2.4) được mở rộng cho cài đặt này khi chúng tôi có chỉ số thời gian.

Bây giờ chúng ta đã đi đến giả định xác định về sự khác biệt trong sự khác biệt: các xu hướng song song. Giả định này cho thấy xu hướng (theo thời gian) ở nhóm điều trị sẽ phù hợp với xu hướng ở nhóm đối chứng (theo thời gian) nếu nhóm điều trị không được điều trị.

Giả định 10.2 (Xu hướng song song)

$$[1(0) \quad 0(0) | = 1] = [1(0) \quad 0(0) | = 0] \quad (10.13)$$

Điều này giống như một giả định về tính vô căn cứ giữa sự khác biệt:

Sự vô căn cứ thường xuyên:

$$(1(0) \quad 0(0)) \quad (10.14)$$

(0)

(10.5 được sửa lại)

Vì vậy, bạn có thể thấy điều này giống như sự vô căn cứ thông thường mà chúng ta đã thấy trong Công thức 10.5, nhưng việc điều trị không phụ thuộc vào sự khác biệt của các kết quả tiềm năng, thay vì độc lập với bản thân kết quả tiềm năng.

Sau đó, chúng ta cần một giả định cuối cùng. Đây là giả định rằng việc điều trị không có tác dụng đối với nhóm điều trị trước khi nó được thực hiện.

Giả định 10.3 (Không có tác động tiền xử lý)

$$[0(1) \quad 0(0) | = 1] = 0 \quad (10.15)$$

Giả định này có vẻ như rõ ràng là đúng, nhưng thực tế không nhất thiết phải như vậy. Ví dụ: nếu người tham gia dự kiến được việc điều trị thì họ có thể

### 10.3.2 Kết quả chính và chứng minh

Sử dụng các giả định ở phần trước, chúng tôi có thể chỉ ra rằng ATT bằng với chênh lệch giữa những khác biệt theo thời gian ở mỗi nhóm can thiệp. Chúng tôi phát biểu điều này một cách toán học trong mệnh đề sau.

Mệnh đề 10.1 (Xác định sự khác biệt trong sự khác biệt) Với sự nhất quán, các xu hướng song song và không có tác động tiền xử lý, chúng ta có kết quả sau:

$$\begin{aligned} & [1(1) \quad 1(0) | = 1] \\ & = ([1 | = 1] - [0 | = 1]) - ([1 | = 0] - [0 | = 0]) \quad (10.16) \end{aligned}$$

**Bài tập đọc tích cực:** Bạn ước tính ước lượng thống kê ở vé phái của Công thức 10.16 như thế nào?

Bằng chứng. Như thường lệ, chúng ta bắt đầu với tính tuyến tính của kỳ vọng:

$$[1(1) \quad 1(0) | = 1] = [1(1) | = 1] + [1(0) | = 1] \quad (10.17)$$

Chúng ta có thể xác định ngay kết quả tiềm năng được điều trị trong nhóm được can thiệp bằng cách sử dụng tính nhất quán

$$= [1 | = 1] - [1(0) | = 1] \quad (10.18)$$

Vì vậy, chúng tôi đã xác định được số hạng đầu tiên, nhưng số hạng thứ hai vẫn chưa được xác định. Để làm được điều đó, chúng ta sẽ giải thuật ngữ này theo giả định về các xu hướng song song:1

$$[1(0) | = 1] = [0(0) | = 1] + [1(0) | = 0] [0(0) | = 0] \quad (10.19)$$

Chúng ta có thể sử dụng tính nhất quán để xác định hai thuật ngữ cuối:

$$= [0(0) | = 1] + [1 | = 0] [0 | = 0] \quad (10.20)$$

Nhưng thuật ngữ đầu tiên là phản thực tế. Đây là lúc chúng ta cần giả định không có tác dụng trước điều trị:2

$$= [0(1) | = 1] + [1 | = 0] [0 | = 0] \quad (10.21)$$

<sup>1</sup> Giả định về xu hướng song song (Giả định 10.2):

$$\begin{aligned} & [1(0) | = 1] [0(0) | = 1] \\ & = [1(0) | = 0] [0(0) | = 0] \end{aligned} \quad (\text{xem lại 10.13})$$

<sup>2</sup> Không có giả định về hiệu quả tiền xử lý (Giả định 10.3):

$$[0(1) | = 1] [0(0) | = 1] = 0 \quad (\text{xem lại 10.15})$$

Bây giờ, chúng ta có thể sử dụng tính nhất quán để hoàn thành việc xác định:

$$= [0 | = 1] + [1 | = 0] [0 | = 0] \quad (10.22)$$

Bây giờ chúng tôi đã xác định được  $[1(0) | = 1]$ , ta có thể thế phương trình 10.22 trở lại phương trình 10.18 để hoàn thành chứng minh:

$$\begin{aligned} & [1(1) | = 1] [1(0) | = 1] \\ & = [1 | = 1] ([0 | = 1] + [1 | = 0] [0 | = 0]) \\ (10.23) & = ([1 | = 1] [0 | = 1]) ([1 | = 0] [0 | = 0]) \quad (10.24) \end{aligned}$$

□

## 10.4 Các vấn đề chính

Vấn đề lớn đầu tiên với các phương pháp sai biệt trong khác biệt để ước tính tác động nhân quả là giả định các xu hướng song song thường không được thỏa mãn. Chúng ta có thể cố gắng khắc phục điều này bằng cách kiểm soát các yếu tố gây nhiễu có liên quan và cố gắng thỏa mãn giả định về các xu hướng song song được kiểm soát:

Giả định 10.4 (Xu hướng song song được kiểm soát)

$$[1(0) 0(0) | = 1, ] = [1(0) 0(0) | = 0, ] \quad (10.25)$$

Điều này thường được thực hiện trong thực tế, nhưng vẫn có thể không thỏa mãn được phiên bản yếu hơn này của giả định xu hướng song song. Ví dụ: nếu có một số hạng tương tác giữa điều trị và thời gian trong phương trình cấu trúc của , thì chúng ta sẽ không bao giờ có các xu hướng song song.

Ngoài ra, giả định về các xu hướng song song có tính chất cụ thể theo quy mô. Ví dụ: nếu chúng ta thỏa mãn các xu hướng song song, điều này không có nghĩa là chúng ta thỏa mãn các xu hướng song song dưới một phép biến đổi nào đó của . Logarit là một dạng phổ biến

sự biến đổi như vậy. Điều này là do giả định các xu hướng song song là một giả định về sự khác biệt, khiến nó không hoàn toàn phi tham số. Theo nghĩa này, giả định các xu hướng song song là bán tham số. Và, tương tự, phương pháp sai phân trong khác biệt là phương pháp bán tham số phương pháp.

# Khám phá nhân quả từ Dữ liệu quan sát

# 11

Trong suốt cuốn sách này, chúng tôi đã thực hiện suy luận nhân quả, giả sử chúng tôi biết đồ thị nhân quả. Nếu chúng ta không biết biểu đồ thì sao? Chúng ta có thể học nó không? BẰNG bạn có thể mong đợi, dựa trên việc đây là chủ đề đang diễn ra trong cuốn sách này, nó sẽ phụ thuộc vào những giả định mà chúng ta sẵn sàng đưa ra. Chúng tôi sẽ giới thiệu đối với vấn đề này như là nhận dạng cấu trúc, khác biệt với nguyên nhân nhận dạng ước tính mà chúng ta đã thấy trong cuốn sách cho đến nay.

## 11.1 Khám phá nhân quả dựa trên độc lập

### 11.1.1 Giả định và định lý

Giả định chính mà chúng ta đã thấy liên quan đến biểu đồ với phân bố là giả định Markov. Giả định Markov cho chúng ta biết liệu các biến có được phân tách bằng d trong đồ thị thì chúng độc lập trong phân bố (Định lý 3.1):

$$| = | \quad (xem lại 3.20)$$

Có lẽ chúng ta có thể phát hiện sự độc lập trong dữ liệu và sau đó sử dụng nó để suy ra đồ thị nhân quả. Tuy nhiên, điều từ sự độc lập trong phân bố cho các khoảng cách d trong đồ thị không phải là điều mà Giả định Markov cho chúng ta (xem Công thức 3.20 ở trên). Đúng hơn là chúng ta cần nghịch đảo của giả định Markov. Đây được gọi là sự trung thành giả thiết.

Giả định 11.1 (Sự trung thành)

$$| = | \quad (11.1)$$

Giả định này cho phép chúng ta suy ra các phân tách d trong biểu đồ từ tính độc lập trong phân phối. Sự trung thành, cùng với Markov giả định, thực sự ngụ ý sự tối thiểu (Giả định 3.2), vì vậy nó là một giả định mạnh mẽ hơn. Sự chung thủy là một giả định kém hấp dẫn hơn nhiều hơn giả định Markov vì dễ dàng nghĩ ra các phản ví dụ (trong đó hai biến độc lập trong , nhưng có các biến không bị chặn đường đi giữa chúng trong ).

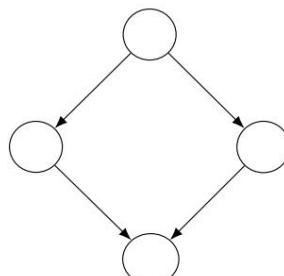
Phản ví dụ về độ trung thực Xét và trong biểu đồ nhân quả với các hệ số ở hình 11.1. Chúng ta đã vi phạm lòng chung thủy khi đường dẫn triệt tiêu đường dẫn . Để thấy cụ thể điều này có thể xảy ra như thế nào, hãy xem xét SCM mà biểu đồ này thể hiện:

$$:= \quad (11.2)$$

$$:= \quad (11.3)$$

$$:= + \quad (11.4)$$

11.1 Nhân quả dựa trên tính độc lập	100
Khám phá . . . . .	100
Giả sử và Định lý	100
Thuật toán PC . . . . .	102
Chúng ta có thể nhận dạng tốt hơn không? . . . . .	104
11.2 Phân tích nhân quả bán tham số	
che phủ . . . . .	104
Không có khả năng nhận dạng nếu không có Giả định tham số .	105
Nhiều phi Gauss tuyến tính	105
Mô hình phi tuyến . . . . .	108
11.3 Tài nguyên bổ sung . . . . .	109



Hình 11.1: Phản ví dụ về sự trung thực đồ thị.

Chúng ta có thể giải sự phụ thuộc giữa và bằng cách thay và vào phương trình 11.4 để có được kết quả sau:

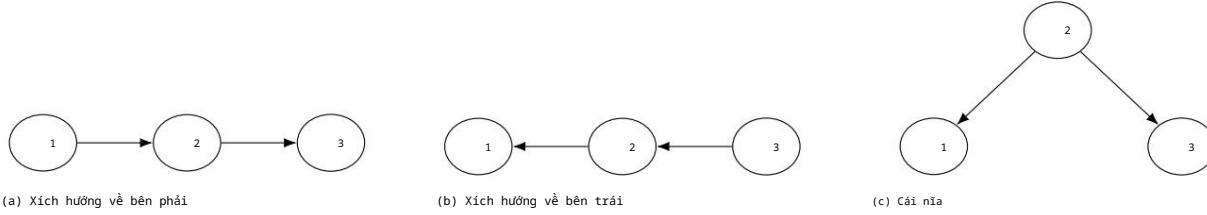
$$= ( \quad + \quad ) \quad (11.5)$$

Điều này có nghĩa là sự liên kết đi từ đến là + trong ví dụ này. Hai đường đi sẽ hủy nhau nếu = , điều này sẽ tạo nên . Sự vi phạm tính trung thực này sẽ khiến chúng ta tin rằng không có đường dẫn nào giữa và trong biểu đồ.

Ngoài tính trung thực, nhiều phương pháp còn cho rằng không có yếu tố gây nhiễu nào không được quan sát, được gọi là tính đủ nhân quả.

**Giả định 11.2 (Tính đủ nhân quả)** Không có yếu tố gây nhiễu nào không được quan sát đối với bất kỳ biến nào trong biểu đồ.

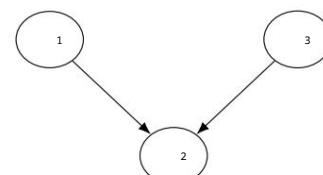
Sau đó, theo giả định Markov, độ trung thực, tính đầy đủ nhân quả và tính chu kỳ, chúng ta có thể xác định một phần biểu đồ nhân quả. Chúng ta không thể xác định hoàn toàn đồ thị nhân quả vì các đồ thị khác nhau tương ứng với cùng một tập hợp các yếu tố độc lập. Ví dụ, hãy xem xét các đồ thị trong Hình 11.2.



Hình 11.2: Ba đồ thị tương đương Markov

Mặc dù đây đều là những biểu đồ riêng biệt nhưng chúng tương ứng với cùng một tập hợp các giả định về tính độc lập/phụ thuộc. Nhớ lại Mục 3.5 rằng  $1 \perp\!\!\!\perp 3 | 2$  trong phân bố Markov đối với bất kỳ đồ thị nào trong ba đồ thị trên Hình 11.2. Chúng ta cũng thấy rằng cực tiểu cho chúng ta biết rằng 1 và 2 là phụ thuộc và 2 và 3 là phụ thuộc. Và giả định về độ trung thực mạnh hơn còn cho chúng ta biết rằng trong bất kỳ phân bố nào đúng với bất kỳ đồ thị nào trong số này, 1 và 3 đều phụ thuộc nếu chúng ta không đặt điều kiện vào 2. Vì vậy, việc sử dụng sự hiện diện/không có tính độc lập (có điều kiện) trong dữ liệu là không đủ để phân biệt ba biểu đồ này với nhau; những đồ thị này tương đương với Markov;

Chúng ta nói rằng hai đồ thị là tương đương Markov nếu chúng tương ứng với cùng một tập hợp các tính độc lập có điều kiện. Cho một biểu đồ, chúng ta gọi lớp tương đương Markov của nó là tập hợp các biểu đồ mã hóa các tính độc lập có điều kiện giống nhau. Với tính trung thực, chúng ta có thể xác định một biểu đồ khỏi sự độc lập có điều kiện trong dữ liệu nếu nó là biểu đồ duy nhất trong lớp tương đương Markov của nó. Bất kỳ ví dụ nào về đồ thị là đồ thị duy nhất trong lớp tương đương Markov của nó đều có tính phi đạo đức cơ bản mà chúng tôi trình bày trong Hình 11.3. Hãy nhớ lại Phần 3.6 rằng sự vô đạo đức khác biệt với hai khái niệm xây dựng đồ họa cơ bản khác (chuỗi và nhánh) ở chỗ trong Hình 11.3, 1 độc lập (vô điều kiện) với 3, và 1 và 3 trở nên phụ thuộc nếu chúng ta đặt điều kiện vào 2. Điều này có nghĩa là trong khi các chuỗi và nhánh cơ bản trong Hình 11.2 thuộc cùng một lớp tương đương Markov, thì bản thân sự vô đạo đức cơ bản lại nằm trong lớp tương đương Markov của chính nó.



Hình 11.3: Những hành vi vô đạo đức nằm trong lớp tương đương Markov của riêng chúng.

Chúng ta đã thấy rằng chúng ta có thể xác định được biểu đồ nhân quả nếu đó là một hành vi trái đạo đức cơ bản, nhưng chúng ta có thể xác định được điều gì khác? Chúng ta đã thấy rằng các chuỗi và nhánh đều thuộc cùng một lớp tương đương Markov, nhưng điều đó không có nghĩa là chúng ta không thể lấy bất kỳ thông tin nào từ các phân phối Markov và trung thực đối với các biểu đồ đó. Tất cả các chuỗi và nhánh trong Hình 11.2 có điểm gì chung? Họ có chung một bộ xương. Khung của đồ thị là cấu trúc mà chúng ta có được nếu thay thế tất cả các cạnh có hướng của nó bằng các cạnh vô hướng. Chúng ta mô tả bộ xương của một chuỗi cơ bản và một nhánh cơ bản trong Hình 11.4.

Khung của đồ thị cũng cung cấp cho chúng ta thông tin độc lập có điều kiện quan trọng mà chúng ta có thể sử dụng để phân biệt nó với các đồ thị có các khung khác nhau. Ví dụ: nếu chúng ta thêm cạnh 1 → 3 vào chuỗi trong Hình 11.2a, chúng ta sẽ có được đồ thị hoàn chỉnh Hình 11.5. Trong biểu đồ này, không giống như biểu đồ chuỗi hoặc biểu đồ ngã ba, 1 và 3 không độc lập khi chúng ta điều kiện hóa trên 2. Vì vậy, đồ thị này không cùng loại với lớp tương đương Markov như các chuỗi và nhánh trong Hình 11.2. Và chúng ta có thể thấy điều đó bằng đồ họa bởi thực tế là biểu đồ này có khung khác với các biểu đồ đó (biểu đồ này có một cạnh bổ sung giữa 1 và 3).

Tóm lại, chúng tôi đã chỉ ra hai đặc điểm cấu trúc mà chúng tôi có thể sử dụng để phân biệt các biểu đồ với nhau:

#### 1. Vô đạo đức 2.

#### Bộ xương

Và hóa ra là chúng ta có thể xác định xem đồ thị thuộc lớp tương đương Markov giống hay khác nhau bằng cách sử dụng hai đặc tính cấu trúc này, nhờ kết quả của Verma và Pearl [78] và Frydenberg [79]:

Mệnh đề 11.1 (Tương đương Markov thông qua các bộ xương vô đạo đức) Hai đồ thị tương đương với Markov khi và chỉ khi chúng có cùng một khung và những điểm vô đạo đức giống nhau.

Điều này có nghĩa là, bằng cách sử dụng tính độc lập có điều kiện trong dữ liệu, chúng ta không thể phân biệt các biểu đồ có cùng khung và các điểm vô đạo đức giống nhau.

Ví dụ: chúng ta không thể phân biệt đồ thị hai nút với

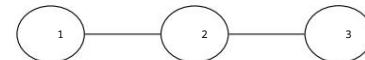
chỉ bằng cách sử dụng thông tin độc lập có điều kiện.<sup>2</sup> Nhưng chúng ta có thể hy vọng tìm hiểu được bộ khung và những điểm vô đạo đức của đồ thị; đây được gọi là biểu đồ thiết yếu hoặc CPDAG (Biểu đồ tuần hoàn được định hướng một phần đã hoàn thành). Một thuật toán phổ biến để học đồ thị cơ bản là thuật toán PC.

### 11.1.2 Thuật toán PC

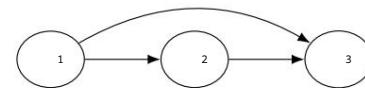
PC [80] bắt đầu với một đồ thị vô hướng hoàn chỉnh, sau đó cắt bớt và định hướng các cạnh thông qua ba bước:

1. Xác định bộ xương.
2. Nhận diện những điều vô đạo đức và định hướng chúng.
3. Định hướng các cạnh định hướng tới máy và chạm.

Chúng ta sẽ sử dụng đồ thị thực trong Hình 11.6 làm ví dụ cụ thể khi giải thích từng bước này.



Hình 11.4: Khung xích/ngã ba.



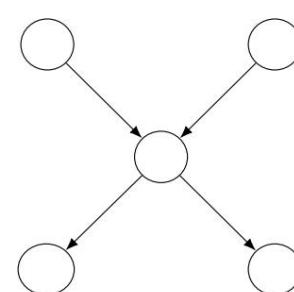
Hình 11.5: Biểu đồ hoàn chỉnh.

<sup>1</sup> Hãy nhớ lại rằng một đồ thị hoàn chỉnh là một đồ thị có cạnh nối mọi cặp đồ thị điểm giao.

[78]: Verma và Pearl (1990), 'Sự tương đương và tổng hợp các mô hình nhân quả'

[79]: Frydenberg (1990), 'Biểu đồ chuỗi Tài sản Markov'

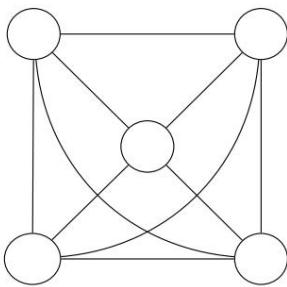
<sup>2</sup> *Bài tập đọc tích cực: Kiểm tra xem các biểu đồ này có mã hóa các tính độc lập có điều kiện giống nhau không .*



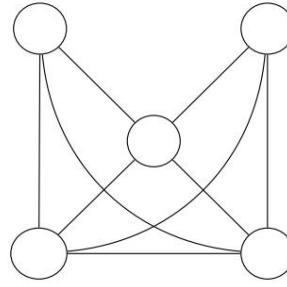
Hình 11.6: Ví dụ đồ thị đúng cho PC.

[80]: Spirtes và cộng sự. (2001), Nguyên nhân, Dự đoán và Tìm kiếm

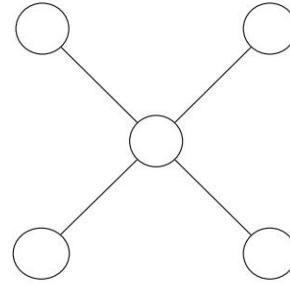
Xác định bộ xưởng Chúng ta khám phá bộ xưởng bằng cách bắt đầu với một biểu đồ hoàn chỉnh (Hình 11.7a) và sau đó loại bỏ các cạnh | cho một số bộ điều hòa (có thể trống). Vì vậy, trong ví dụ của chúng ta, chúng ta sẽ bắt đầu với tập điều hòa trống và phát hiện ra rằng (vì đường đi duy nhất từ đến trong Hình 11.6 bị chặn bởi máy va chạm ); điều này có nghĩa là chúng ta có thể loại bỏ cạnh , mang lại cho chúng ta đồ thị thị trong Hình 11.7b. Sau đó, chúng ta sẽ chuyển sang các tập điều kiện có kích thước một và thấy rằng điều hòa trên cho chúng ta biết rằng mọi cặp biến khác đều độc lập có điều kiện cho trước , điều này cho phép chúng ta loại bỏ tất cả các cạnh không liên quan đến , dẫn đến đồ thị trong Hình 11.7c. Và quả thực đây chính là khung của đồ thị thực trong Hình 11.6. PC tổng quát hơn sẽ tiếp tục với các bộ điều hòa lớn hơn, để xem liệu chúng ta có thể loại bỏ nhiều cạnh hơn hay không, nhưng các bộ điều hòa có kích thước một là đủ để khám phá khung trong ví dụ này.



(a) Hoàn thành đồ thị vô hướng mà chúng ta bắt đầu với



(b) Đồ thị vô hướng còn lại sau khi loại bỏ các cạnh trong đó

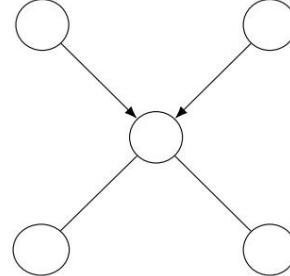


(c) Đồ thị vô hướng còn lại sau khi loại bỏ các cạnh trong đó |

Hình 11.7: Minh họa quy trình bước 1 của PC, trong đó chúng ta bắt đầu với biểu đồ hoàn chỉnh (trái) và loại bỏ các cạnh cho đến khi chúng ta xác định được khung của biểu đồ (phải), với điều kiện là biểu đồ thực là biểu đồ trong Hình 11.6.

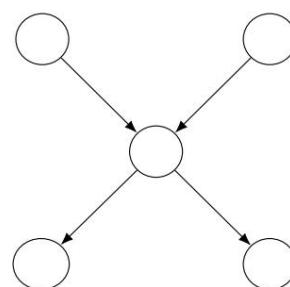
Xác định sự trái đạo đức Bây giờ cho bất kỳ đường dẫn nào trong biểu đồ làm việc của chúng tôi, nơi chúng tôi phát hiện ra rằng không có cạnh nào giữa và trong bước trước, nếu không nằm trong tập điều kiện làm cho và độc lập có điều kiện, thì chúng ta biết tạo thành một sự vô đạo đức. Nói cách khác, điều này có nghĩa là , là một tính chất vô đạo đức giúp phân biệt nó với dây xích và nĩa (Phần 3.6), vì vậy chúng ta có thể định hướng các cạnh này để có được . Trong ví dụ của chúng tôi, điều này đưa chúng ta từ Hình 11.7c đến Hình 11.8.

Sự cố định hướng các cạnh đủ điều kiện trên máy va chạm Ở bước cuối cùng, chúng tôi lợi dụng thực tế là chúng tôi có thể định hướng nhiều cạnh hơn vì chúng tôi biết rằng chúng tôi đã phát hiện ra tất cả những điểm vô đạo đức ở bước trước. Bất kỳ cạnh nào của đường đi có hướng một phần có dạng , trong đó không có cạnh nào nối và , đều có thể định hướng là . Điều này là do nếu đồ thị đúng có cạnh , thì chúng ta đã tìm được đồ thị này ở bước trước vì điều đó sẽ tạo thành đồ thị vô đạo đức . Vì chúng ta không tìm thấy sự vô đạo đức đó ở bước trước nên chúng ta biết rằng hướng thực sự là . Trong ví dụ của chúng ta, điều này có nghĩa là chúng ta có thể định hướng hai cạnh cuối cùng còn lại, đưa chúng ta từ Hình 11.8 đến Hình 11.9. Hóa ra trong ví dụ này, chúng ta thật may mắn khi có thể định hướng tất cả các cạnh còn lại ở bước cuối cùng này, nhưng nói chung thì không phải vậy. Ví dụ: chúng ta đã thảo luận rằng chúng ta sẽ không thể phân biệt các biểu đồ chuỗi đơn giản và các biểu đồ phân nhánh đơn giản với nhau .



Hình 11.8: Đồ thị từ PC sau khi chúng ta đã hướng tới những điều trái đạo đức.

<sup>3</sup> Điều này được gọi là lan truyền định hướng.



Hình 11.9: Đồ thị từ PC sau khi chúng ta đã định hướng các cạnh sẽ tạo thành các điểm vô đạo đức nếu chúng được định hướng theo hướng khác (không chính xác).

Loại bỏ các giả định Có những thuật toán cho phép chúng ta loại bỏ các giả định khác nhau. Thuật toán FCI (Suy luận nhân quả nhanh) [80] hoạt động mà không cần giả định tính đầy đủ của nhân quả (Giả định 11.2). Thuật toán CCD [81] hoạt động mà không giả định tính chu kỳ. Và có nhiều công trình khác nhau về khám phá nhân quả dựa trên SAT cho phép chúng ta loại bỏ cả hai giả định trên [82, 83].

Độ khó của việc kiểm tra tính độc lập có điều kiện Tất cả các phương pháp dựa vào các bài kiểm tra tính độc lập có điều kiện như PC, FCI, thuật toán dựa trên SAT, v.v. đều có một vấn đề thực tế quan trọng liên quan đến chúng. Kiểm tra tính độc lập có điều kiện rất khó và đôi khi có thể cần nhiều dữ liệu để có kết quả kiểm tra chính xác [84]. Nếu chúng tôi có dữ liệu vô hạn thì đây không phải là vấn đề, nhưng trên thực tế, chúng tôi không có dữ liệu vô hạn.

#### 11.1.3 Chúng ta có thể nhận được thông tin nhận dạng nào tốt hơn không?

Chúng ta đã thấy rằng việc giả định giả định Markov và tính trung thực chỉ có thể giúp chúng ta tiến xa; với những giả định đó, chúng ta chỉ có thể xác định một biểu đồ có lớp tương đương Markov của nó. Nếu chúng ta đưa ra nhiều giả định hơn, liệu chúng ta có thể xác định biểu đồ chính xác hơn lớp tương đương Markov của nó không?

Chà, nếu chúng ta ở trong trường hợp phân phối là đa thức, chúng ta không thể [85]. Hoặc nếu chúng ta ở trong trường hợp đồ chơi thông thường trong đó SCM tuyển tính với nhiễu Gaussian thì chúng ta không thể [86]. Vì vậy, chúng tôi có kết quả đầy đủ sau đây do Geiger và Pearl [86] và Meek [85]:

**Định lý 11.2 (Tính đầy đủ Markov)** Nếu chúng ta có phân phối đa thức hoặc phương trình cấu trúc Gauss tuyển tính, chúng ta chỉ có thể xác định một đồ thị lên đến lớp tương đương Markov của nó.

Điều gì sẽ xảy ra nếu chúng ta không có phân phối đa thức và không có phân phối tuyển tính Tuy nhiên, SCM Gaussian?

## 11.2 Khám phá nhân quả bán tham số

Trong Định lý 11.2, chúng ta đã thấy rằng, nếu chúng ta đang ở trong bối cảnh Gaussian tuyển tính, điều tốt nhất chúng ta có thể làm là xác định lớp tương đương Markov; chúng ta không thể hy vọng xác định được các đồ thị nằm trong các lớp tương đương Markov không đơn lẻ. Nhưng điều gì sẽ xảy ra nếu chúng ta không ở trong cài đặt Gaussian tuyển tính? Chúng ta có thể xác định đồ thị nếu chúng ta không ở trong cài đặt Gaussian tuyển tính không? Chúng tôi xem xét cài đặt nhiều phi tuyển tính Gaussian trong Phần 11.2.2 và cài đặt nhiều công phi tuyển trong Phần 11.2.3. Hóa ra trong cả hai cách cài đặt này, chúng ta đều có thể xác định được biểu đồ nhân quả. Và chúng ta không cần phải thừa nhận tính trung thực (Giả định 11.1) trong những bối cảnh này.

Bằng cách xem xét các cài đặt này, chúng tôi đang đưa ra các giả định bán tham số (về dạng hàm). Nếu chúng ta không đưa ra bất kỳ giả định nào về dạng hàm, chúng ta thậm chí không thể xác định được hướng của cạnh trong biểu đồ hai nút. Chúng tôi nhấn mạnh điều này trong phần tiếp theo trước khi chuyển sang các giả định bán tham số cho phép chúng tôi xác định biểu đồ.

[80]: Spirtes và cộng sự. (2001), Nguyên nhân, Dự đoán và Tìm kiếm

[81]: Richardson (1996), 'Mô hình phản hồi: Giải thích và khám phá'

[82]: Hyttinen và cộng sự. (2013), 'Khám phá các mô hình nhân quả tuần hoàn với các biến tiềm ẩn: Quy trình chung dựa trên SAT'

[83]: Hyttinen et al. (2014), 'Khám phá nguyên nhân dựa trên ràng buộc: Giải quyết xung đột bằng lập trình tập hợp câu trả lời'

[84]: Shah và Peters (2020), 'Độ cứng của kiểm tra tính độc lập có điều kiện và thước đo hiệp phương sai tổng quát'

[85]: Meek (1995), 'Tính hoàn chỉnh và trung thực mạnh mẽ trong Mạng Bayesian'

[86]: Geiger và Pearl (1988), 'Về logic của các mô hình nhân quả'

### 11.2.1 Không thể nhận dạng nếu không có giả định tham số

Phối cảnh Markov Xem xét cài đặt hai biến, trong đó hai tùy chọn của đồ thị nhân quả là và . Lưu ý rằng hai đồ thị này tương đương với Markov. Cả hai đều không mã hóa bất kỳ giả định độc lập có điều kiện nào, vì vậy cả hai đều có thể mô tả phân phối tùy ý ( , ). Điều này có nghĩa là sự độc lập có điều kiện trong dữ liệu không thể giúp chúng ta phân biệt giữa và . Sử dụng tính độc lập có điều kiện, điều tốt nhất chúng ta có thể làm là khám phá đồ thị cơ bản tương ứng .

Quan điểm của SCM Sẽ thế nào nếu chúng ta xem xét vấn đề này từ góc độ của SCM; bằng cách nào đó chúng ta có thể phân biệt với bằng cách sử dụng SCM không? Đối với SCM, chúng tôi muốn viết một biến dưới dạng hàm của biến kia và một số biến thuật ngữ nhiều. Như bạn có thể mong đợi, nếu chúng ta không đưa ra bất kỳ giả định nào, sẽ tồn tại các SCM với biểu đồ nhân quả ngẫu ý và SCM với biểu đồ nhân quả ngẫu ý cả hai đều tạo ra dữ liệu theo ( , ).

Mệnh đề 11.3 (Không thể nhận dạng được đồ thị hai nút) Đối với mọi phân phối chung ( , ) trên hai biến ngẫu nhiên có giá trị thực, sẽ có một SCM theo một trong hai hướng tạo ra dữ liệu nhất quán với ( , ).

Về mặt toán học, tồn tại hàm sao cho

$$= \dots ( , \dots ), \quad (11.6)$$

và tồn tại hàm số sao cho

$$= \dots ( , \dots ), \quad (11.7)$$

trong đó và là các biến ngẫu nhiên có giá trị thực.

Xem, ví dụ, Peters et al. [14, tr. 44] để có một bằng chứng ngắn gọn. Tương tự, kết quả không thể nhận dạng này có thể được mở rộng cho các đồ thị tổng quát hơn có nhiều hơn hai biến [xem, ví dụ, 14, tr. 135].

[14]: Peters và cộng sự. (2017), Các yếu tố của suy luận nhân quả: Nền tảng và thuật toán học tập

Tuy nhiên, nếu chúng ta giả định về dạng tham số của SCM, chúng ta có thể phân biệt với và xác định đồ thị một cách tổng quát hơn. Đó là những gì chúng ta sẽ thấy trong phần còn lại của chương này.

### 11.2.2 Nhiều tuyển tính phi Gaussian

Chúng ta đã thấy trong Định lý 11.2 rằng chúng ta không thể phân biệt các đồ thị trong cùng một lớp tương đương Markov nếu các phương trình cấu trúc là tuyển tính với nhiều Gaussian . Ví dụ, điều này có nghĩa là chúng ta không thể phân biệt với . Tuy nhiên, nếu hạng nhiều không phải Gaussian thì chúng ta có thể xác định được đồ thị nhân quả. Như thường lệ, chúng tôi đưa ra giả định quan trọng về phi Gaussianity trong hộp riêng của nó:

Giả định 11.3 (Tuyển tính phi Gaussian) Tất cả các phương trình cấu trúc (nhân quả

cơ chế tạo ra dữ liệu) có dạng sau:

$$:= (\quad) + \quad (11.8)$$

trong đó  $\mu$  là hàm tuyến tính,  $\sigma^2$ , và  $\epsilon$  được phân phối dưới dạng không phải Gaussian biến ngẫu nhiên.

Sau đó, trong cài đặt phi Gaussian tuyến tính này, chúng ta có thể xác định cái nào trong số đồ thị và là đồ thị nhân quả thực sự. Đầu tiên chúng tôi sẽ trình bày định lý và chứng minh rồi đi đến trực giác.

**Định lý 11.4** (Nhận dạng được trong bối cảnh tuyến tính phi Gaussian) Trong thiết lập tuyến tính phi Gaussian, nếu SCM thực sự là

$$:= (\quad) + , , \quad (11.9)$$

thì không tồn tại SCM theo hướng ngược lại

$$:= (\quad) + , , \quad (11.10)$$

có thể tạo ra dữ liệu phù hợp với  $(\mu, \sigma^2)$ .

Bằng chứng. Đầu tiên chúng tôi sẽ giới thiệu một kết quả quan trọng từ Darmois [87] và Skitovich [88] và Skitovich [88] mà chúng ta sẽ sử dụng để chứng minh định lý này:

**Định lý 11.5 (Darmois-Skitovich)** Giả sử  $1, \dots, n$  là các biến ngẫu nhiên độc lập, không suy biến. Nếu tồn tại các hệ số  $a_1, \dots, a_n$  và  $b_1, \dots, b_n$  đều khác 0 sao cho hai tổ hợp tuyến tính

$$\begin{aligned} &= 1 \cdot 1 + \dots + n \cdot n \quad \text{và} \\ &= 1 \cdot 1 + \dots + n \cdot n \end{aligned}$$

độc lập thì mỗi tổ hợp có phân phối chuẩn.

[87]: Darmois (1953), 'Analyse générale des liaisons entre variables aléatoires: étude particulière de l'analyse factorielle linéaire'

[88]: Skitovich (1954), 'Đặng tuyến tính của các biến ngẫu nhiên độc lập và chuẩn luật phân phối mal'

[88]: Skitovich (1954), 'Đặng tuyến tính của các biến ngẫu nhiên độc lập và quy luật phân phối chuẩn'

Chúng ta sẽ sử dụng phản chứng của trường hợp đặc biệt của định lý này cho

= 2 để thực hiện gần như toàn bộ công việc cho chứng minh này:

Hệ quả 11.6 Nếu một trong hai biến ngẫu nhiên độc lập  $1$  hoặc  $2$  là không phải Gaussian thì không có tổ hợp tuyến tính

$$\begin{aligned} &= 1 \cdot 1 + 2 \cdot 2 \quad \text{và} \\ &= 1 \cdot 1 + 2 \cdot 2 \end{aligned}$$

sao cho  $1$  và  $2$  độc lập (vì vậy  $1$  và  $2$  phải phụ thuộc).

Sơ lược chứng minh Với hệ quả tất yếu ở trên, chiến lược chứng minh của chúng ta là viết  $\mu$  và  $\sigma^2$  dưới dạng tổ hợp tuyến tính của  $1$  và  $2$ . Bằng cách làm này, chúng tôi đang ánh xạ một cách hiệu quả các biến của chúng tôi trong Công thức 11.9 và 11.10 lên các biến trong hệ quả như sau:  $l_1$  lên  $1$ ,  $l_2$  lên  $2$ ,  $l_3$  lên  $1$ , và  $l_4$  lên  $2$ . Khi đó, chúng ta có thể áp dụng hệ quả tất yếu trên của Darmois-Skitovich Định lý để  $l_1$  và  $l_2$  phải phụ thuộc, vì phạm định lý hướng ngược lại SCM trong phương trình 11.10. Vậy giờ chúng ta tiến hành chứng minh.

Chúng ta đã có rằng chúng ta có thể viết dưới dạng tổ hợp tuyến tính của và , vì chúng ta đã giả sử phương trình cầu trúc đúng trong phương trình 11.9 là tuyến tính:

$$= + \quad (11.11)$$

Sau đó, để có được là tổ hợp tuyến tính của và , chúng ta lấy giả thuyết-kích thước SCM đảo ngược

$$= - + - \quad (11.12)$$

từ phương trình 11.10, giải cho , và thế vào phương trình 11.11 cho :

$$= \quad (13.11)$$

$$- = - \quad (11.14)$$

$$+ ) = (1 - ) + - \quad (11.15)$$

Do đó, chúng ta đã viết cả và dưới dạng tổ hợp tuyến tính của biến ngẫu nhiên độc lập và . Điều này cho phép chúng ta áp dụng Hệ quả 11.6 của Định lý Darmois-Skitovish để có được rằng và phải là người phụ thuộc: . Điều này vi phạm hướng ngược lại SCM:

$$:= ( ) + , \quad ( xem lại ngày 11.10)$$

□

Ở đây chúng ta đã chứng minh chỉ hai biến, nhưng nó có thể được mở rộng sang cài đặt tổng quát hơn với nhiều biến số (xem [89] và [14, Mục 7.1.4]).

#### Trực giác đồ họa

Khi chúng ta khớp dữ liệu theo hướng nhân quả, chúng ta sẽ nhận được phần dư độc lập với biến đầu vào, nhưng khi chúng ta khớp dữ liệu vào hướng phản nhân quả, chúng ta nhận được phần dư phụ thuộc vào đầu vào

**Biến đổi.** Chúng ta mô tả đường hồi quy mà chúng ta có được nếu chúng ta hồi quy tuyến tính trên (hướng nhân quả) trong Hình 11.10a, và chúng tôi mô tả sự hồi quy dòng chúng ta nhận được nếu chúng ta hồi quy tuyến tính trên (hướng ngược nhân quả) trong Hình 11.10b. Chỉ từ những sự phù hợp này, bạn có thể thấy rằng mô hình chuyển tiếp (vừa vặn theo hướng nhân quả) trông đẹp mắt hơn mô hình lắc hậu (phù hợp theo hướng phản nhân quả).

Để làm cho trực giác đồ họa này rõ ràng hơn, chúng tôi vẽ biểu đồ phần dư của mô hình thuận (hướng nhân quả) và mô hình lùi (hướng phản nhân quả) trong Hình 11.11. Phần dư theo hướng thuận tương ứng với các giá trị sau:  $\hat{y} = \hat{y}(\cdot)$ . Và phần dư trong hướng lùi tương ứng như sau:  $\hat{y} = \hat{y}(\cdot)$ . Như bạn có thể xem trong Hình 11.11a, phần dư của mô hình kỳ hạn trông độc lập của biến đầu vào (trên trực x). Tuy nhiên trong Hình 11.10b, phần dư của mô hình lùi trông không độc lập với đầu vào biến (trên trực x). Rõ ràng, phạm vi của phần dư (trên dọc) thay đổi khi chúng ta di chuyển qua các giá trị của (từ trái sang phải).

[89]: Shimizu và cộng sự. (2006), 'Mô hình tuần hoàn tuyến tính phi Gaussian để khám phá nguyên nhân'

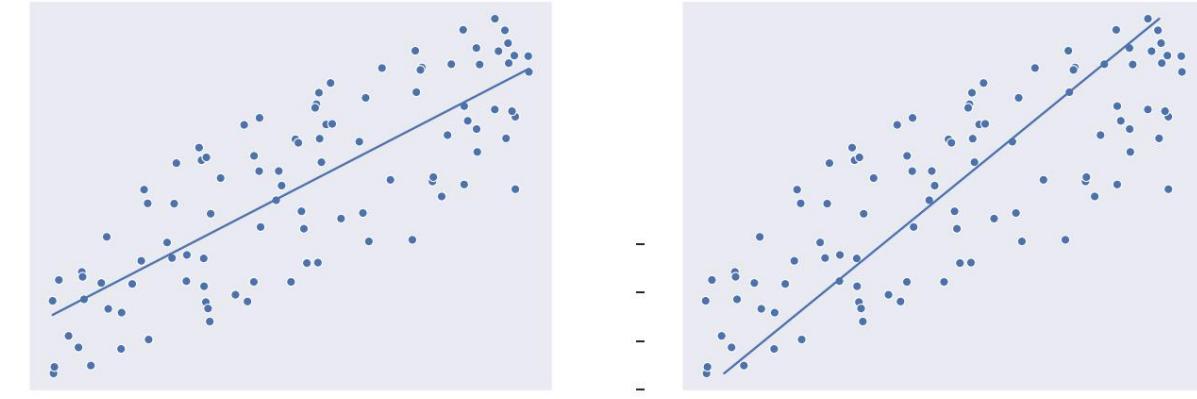
[14]: Peters và cộng sự. (2017), Yếu tố nhân quả Suy luận: Cơ sở và thuật toán học tập

Chuyển tiếp mô hình SCM:

$$:= ( ) + , \quad ( xem lại ngày 11.9)$$

Mô hình lắc hậu SCM:

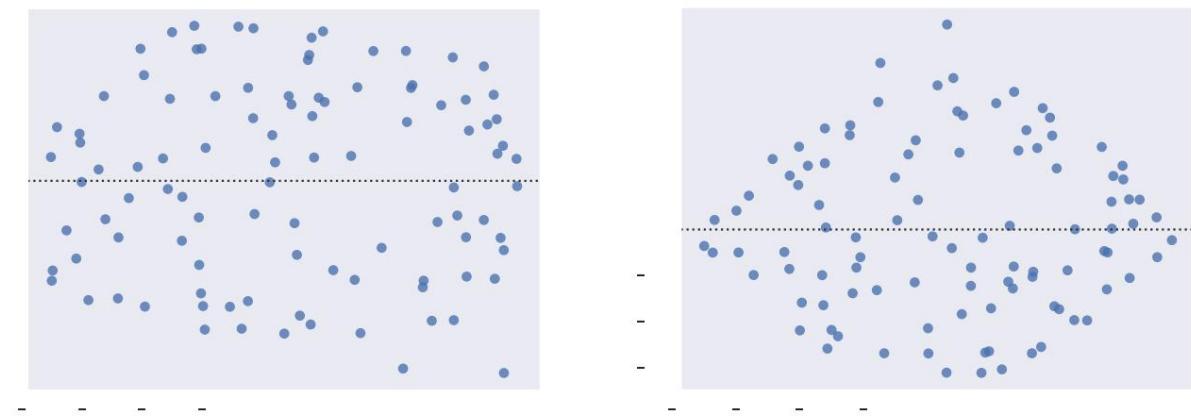
$$:= ( ) + , \quad ( xem lại ngày 11.10)$$



(a) Sự phù hợp theo hướng nhân quả: sự phù hợp tuyến tính là kết quả của sự hồi quy trên

(b) Phù hợp hướng phản nhân quả: phù hợp tuyến tính là kết quả của sự hồi quy trên

Hình 11.10: Khớp tuyến tính (theo cả hai hướng) của dữ liệu tuyến tính phi Gaussian.



(a) Phần dư theo hướng nhân quả: phần dư là kết quả của sự hồi quy tuyến tính trên

(b) Phần dư hướng phản nhân quả: phần dư là kết quả của sự hồi quy sớm linh trên

Hình 11.11: Phần dư của mô hình tuyến tính (theo cả hai hướng) phù hợp với dữ liệu tuyến tính phi Gaussian.

### 11.2.3 Mô hình phi tuyến

Cài đặt nhiều cộng phi tuyến Chúng ta cũng có thể nhận dạng được biểu đồ nhân quả trong cài đặt nhiều cộng phi tuyến [90, 91]. Điều này đòi hỏi giả định nhiều cộng tính phi tuyến (bên dưới) và các giả định kỹ thuật khác mà chúng tôi giới thiệu cho bạn Hoyer et al. [90] và Peters và cộng sự. [91] cho.

[90]: Hoyer và cộng sự. (2009), 'Khám phá nhân quả phi tuyến với các mô hình nhiều phụ gia' [91]: Peters et al. (2014), 'Khám phá nhân quả với các mô hình tiếng ồn cộng gộp liên tục'

**Giả định 11.4 (Nhiều cộng phi tuyến)** Tất cả các cơ chế nhân quả đều phi tuyến trong đó nhiều đi vào theo hướng cộng. Về mặt toán học,

$$\text{trong đó } \text{ là phi tuyến và } \text{biểu thị cha mẹ của} \quad (11.16)$$

Cài đặt hậu phi tuyến Điều gì sẽ xảy ra nếu bạn không tin rằng tiếng ồn thực sự đi vào một cách bổ sung. Điều này thúc đẩy các mô hình hậu phi tuyến, trong đó có một phép biến đổi phi tuyến khác sau khi thêm nhiều như trong

Giả sử 11.5 dưới đây. Cài đặt này cũng có thể mang lại khả năng nhận dạng (trong điều kiện kỹ thuật khác). Xem Zhang và Hyvärinen [92] để biết thêm chi tiết.

[92]: Zhang và Hyvärinen (2009), 'Về khả năng nhận dạng của hậu phi tuyến Mô hình nhân quả'

#### Giả định 11.5 (Hậu phi tuyến)

$$\text{trong đó } \text{ là phi tuyến và } \text{ biểu thị cha mẹ của } . \quad (11.17)$$

trong đó là phi tuyến và pa biểu thị cha mẹ của .

### 11.3 Tài nguyên bổ sung

Chúng tôi kết thúc chương này bằng cách chỉ cho bạn một số tài nguyên có liên quan về nơi bắt đầu tìm hiểu thêm (ngoài các tài liệu tham khảo trong chương này). Những tài liệu tham khảo này cũng được sử dụng làm nguồn cảm hứng khi hình thành chương này. Xem Eberhardt [93] và Glymour et al. [94] vì hai bài báo đánh giá tuyệt vời của những người ở tuyến đầu của nghiên cứu khám phá nhân quả. Và sau đó nếu bạn muốn có cả một cuốn sách về chủ đề này, Peters et al. [14] đã viết một bài phỏ biến!

[93]: Eberhardt (2017), 'Giới thiệu về nền tảng của khám phá nhân quả' [94]: Glymour et al. (2019), 'Đánh giá các phương pháp khám phá nhân quả dựa trên mô hình đồ họa'

[14]: Peters và cộng sự. (2017), Các yếu tố của suy luận nhân quả: Nền tảng và thuật toán học tập

# Khám phá nhân quả từ Dữ liệu can thiệp

# 12

12.1 Can thiệp mang tính cấu trúc

12.1.1 Can thiệp vào một nút

Sắp ra mắt

12.1.2 Can thiệp vào nhiều nút

Sắp ra mắt

12.2 Can thiệp tham số

12.2.1 Sắp ra mắt

12.3 Tương đương Markov can thiệp

12.3.1 Sắp ra mắt

12.4 Các cài đặt khác

12.4.1 Sắp ra mắt

12.1 Can thiệp mang tính cấu trúc . . . 110

Can thiệp đơn nút 110

Can thiệp đa nút . 110

12.2 Can thiệp tham số . . . 110

Sắp có . . . . . 110

12.3 Can thiệp Markov

Sự tương đương . . . . . 110

Sắp có . . . . . 110

12.4 Bộ khác-

tings . . . . . 110

Sắp có . . . . . 110

## Chuyển giao học tập và Khả năng vận chuyển

13

13.1 Hiểu biết sâu sắc về nguyên nhân của việc học chuyển tiếp

13.1.1 Sắp ra mắt

13.2 Khả năng vận chuyển của các tác động nhân quả qua  
Quần thể

13.2.1 Sắp ra mắt

13.1 Hiểu biết sâu sắc về nguyên nhân chuyển giao	
Học tập .	111 · · · · ·
Sắp có .	111 · · · · ·
13.2 Khả năng vận chuyển của nhân quả	
Hiệu ứng trên toàn bộ dân số	111
Sắp có .	111 · · · · ·

# 14

## Phản thực tế và hòa giải

14.1 Khái niệm cơ bản về phản thực tế . . . . .	112
Sắp có . . . . .	112
14.1.1 Sắp ra mắt	
14.2 Úng dụng quan trọng: Hòa giải . . . . .	112
Sắp có . . . . .	112
14.2.1 Sắp ra mắt	

ruột thừa

# MỘT

## Bằng chứng

### A.1 Chứng minh phương trình 6.1 từ Mục 6.1

Khẳng định Cho đồ thị nhân quả là Hình A.1,  $(\quad | \text{do}(\quad)) = (\quad | \quad)$ .

Bằng chứng. Trước tiên chúng tôi áp dụng hệ số hóa mạng Bayes (Định nghĩa 3.1):

$$(\quad, \quad, \quad, \quad) = (\quad) (\quad | \quad) (\quad | \quad) (\quad | \quad, \quad) \quad (\text{A.1})$$

Tiếp theo, chúng ta áp dụng hệ số rút gọn (Mệnh đề 4.1):

$$(\quad, \quad, \quad | \text{do}(\quad)) = (\quad) (\quad | \quad) (\quad | \quad, \quad) \quad (\text{A.2})$$

Cuối cùng, chúng tôi loại trừ và :

$$(\quad, \quad, \quad | \text{làm}(\quad)) = (\quad) (\quad | \quad) (\quad | \quad, \quad) \quad (\text{A.3})$$

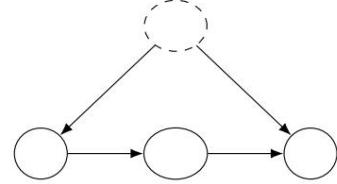
$$(\quad | \text{làm}(\quad)) = (\quad) (\quad | \quad) (\quad | \quad, \quad) \quad (\text{A.4})$$

$$= (\quad | \quad) \quad (\text{A.5})$$

A.1 Chứng minh phương trình 6.1  
từ Mục 6.1 . . . . .

A.2 Bằng chứng về Định lý Điểm Xu hướng (7.1) . . . . . 114

A.3 Bằng chứng ước tính IPW (7.18) 115



□

### A.2 Định lý Bằng chứng về Điểm Xu hướng (7.1)

Yêu cầu  $(\quad(1), \quad(0)) | \quad = (\quad(1), \quad(0)) | \quad(\quad)$ .

Bằng chứng. Giả sử  $(\quad(1), \quad(0)) | \quad$ , chúng ta sẽ chứng minh  $(\quad(1), \quad(0)) | \quad(\quad)$  bằng cách chứng minh rằng  $(\quad = 1, | \quad(\quad), \quad(\quad))$  không phụ thuộc vào  $\quad(\quad)$ , trong đó  $\quad(\quad)$  là kết quả tiềm năng.

Đầu tiên, vì là hệ nhị phân nên có thể biến xácsuất này thành kỳ vọng:

$$(\quad = 1, | \quad(\quad), \quad(\quad)) = [\quad | \quad(\quad), \quad(\quad)] \quad (\text{A.6})$$

Sau đó, bằng cách sử dụng luật kỳ vọng lặp, chúng ta có thể giới thiệu :

$$= [\quad [\quad | \quad(\quad), \quad(\quad), \quad] | \quad(\quad), \quad(\quad)] \quad (\text{A.7})$$

Bởi vì bây giờ chúng ta đã quy định tất cả và ( ) là hàm của , nên nó dư thừa, nên chúng ta có thể loại bỏ ( ) khỏi kỳ vọng bên trong:

$$= [ \quad | \quad ( ), \quad | \quad ( ), \quad ( ) ] \quad (A.8)$$

Bây giờ, chúng ta áp dụng giả định vô căn cứ mà chúng ta đã bắt đầu để loại bỏ ( ) khỏi kỳ vọng bên trong:

$$= [ \quad | \quad ] | \quad ( ), \quad ( ) ] \quad (A.9)$$

Một lần nữa, bằng cách sử dụng thực tế là là hệ nhị phân, chúng ta có thể giảm kỳ vọng bên trong xuống ( = 1 | ) ( ), một điều kiện đã được điều chỉnh dựa trên:

$$= [ \quad ( \quad = 1 | \quad ) | \quad ( ), \quad ( ) ] \quad (A.10)$$

$$= [ \quad ( \quad ) | \quad ( ), \quad ( ) ] \quad (A.11)$$

$$( \quad ) = ( \quad ) \quad (A.12)$$

Vì điều này không phụ thuộc vào ( ) nên chúng tôi đã chứng minh rằng độc lập với ( ) cho trước ( ).  $\square$

### A.3 Bằng chứng ước tính IPW (7.18)

Tuyên bố Dưới sự vô căn cứ và tích cực,  $[ \quad ( \quad ) ] = \frac{( \quad )}{( \quad | \quad )}$ .

Bằng chứng. Chúng ta sẽ bắt đầu với ước tính thống kê mà chúng ta có được từ công thức điều chỉnh (Định lý 2.1). Với tính không nhiễu và dương, công thức điều chỉnh cho chúng ta biết

$$[ \quad ( \quad ) ] = [ \quad | \quad , \quad ] \quad (A.13)$$

Chúng ta sẽ giả sử biến là rời rạc để chia các kỳ vọng này thành tổng (thay thế bằng tích phân nếu liên tục):

$$= ( \quad | \quad , \quad ) - ( \quad ) \quad (A.14)$$

Để có ( | ) trong đó, chúng ta nhân với  $\frac{( \quad )}{( \quad | \quad )}$ :

$$= ( \quad | \quad , \quad ) - ( \quad ) \frac{( \quad | \quad )}{( \quad | \quad )} \quad (A.15)$$

Sau đó, nhận thấy rằng ( | , ) ( | ) ( ) là phân phối chung:

$$= ( \quad | \quad , \quad ) \frac{1}{( \quad | \quad )} \quad (A.16)$$

( , , ) gần như ( ) = [ ], nhưng vì = và = có xác suất nên các số hạng của tổng này chỉ khác 0 nếu = và = . Do đó, chúng ta lấy biến ngẫu nhiên chỉ báo cho sự kiện này với kỳ vọng nằm trên cả ba biến ngẫu nhiên ( , và ) :

$$= \frac{1}{[( = , = )]} \quad (A.17)$$

1 Bây giờ, phần ~~tồn~~ lại là kỳ vọng có trọng số trên . ( | ) Việc tích hợp điều này có nghĩa là vì hiện tại chúng ta đang vượt quá nên trở thành một biến ngẫu nhiên ( ) và = bên trong chỉ báo trở nên dư thừa. Điều này mang lại cho chúng tôi những điều sau đây:

$$= \frac{( = )}{( | )} \quad (A.18)$$

□

Lưu ý: Đối với một số người, việc chuyển thẳng từ Công thức A.16 sang Công thức A.18 có thể sẽ tự nhiên hơn .

# Thư mục

Dưới đây là các tài liệu tham khảo theo thứ tự trích dẫn.

- [1] Tyler Vigen. Tương quan giả. <https://www.tylervigen.com/spurious-correlations>. 2015 (trích dẫn ở trang 3).
- [2] Jerzy Spława-Neyman. 'Về việc áp dụng lý thuyết xác suất vào các thí nghiệm nông nghiệp. Tiêu luận về Nguyên tắc. Phần 9.' Dịch. của DM Dabrowska và TP Speed. Trong: Khoa học thống kê 5.4 (1923 [1990]), trang 465-472 (trích dẫn trang 6).
- [3] Donald B. Rubin. 'Ước tính tác động nhân quả của các phương pháp điều trị trong các nghiên cứu ngẫu nhiên và không ngẫu nhiên.' Trong: Tạp chí Tâm lý giáo dục 66.5 (1974), tr. 688 (trích dẫn trang 6, 7).
- [4] Jasjeet S. Sekhon. 'Mô hình suy luận và ước tính nhân quả của Neyman-Rubin thông qua các phương pháp so khớp'. Trong: Cẩm nang phương pháp chính trị Oxford (2008), trang 271- (trích dẫn trang 6).
- [5] Paul W. Hà Lan. 'Thống kê và suy luận nhân quả'. Trong: Tạp chí của Hiệp hội Thống kê Hoa Kỳ 81.396 (1986), trang 945-960. doi: [10.1080/01621459.1986.10478354](https://doi.org/10.1080/01621459.1986.10478354) (trích dẫn trang 8, 42).
- [6] Alexander D'Amour, Peng Ding, Avi Feller, Lihua Lei và Jasjeet Sekhon. Sự chồng chéo trong các nghiên cứu quan sát với các hiệp phương sai chiều cao. 2017 (trích dẫn trang 13).
- [7] Miguel A Hernán và James M Robins. Suy luận nhân quả: Nếu như. Boca Raton: Chapman & Hall/CRC, 2020 (trích dẫn trang 14, 27, 90).
- [8] Miguel Angel Luque-Fernandez, Michael Schomaker, Daniel Redondo-Sanchez, Maria Jose Sanchez Perez, Anand Vaidya và Mireille E Schnitzer. 'Lưu ý mang tính giáo dục: Hiệu ứng máy và chậm nghịch lý trong phân tích dữ liệu dịch tễ học về bệnh không lây nhiễm: một minh họa có thể tái tạo và ứng dụng web'. Trong: Tạp chí Dịch tễ học Quốc tế 48.2 (Tháng 12 năm 2018), trang 640-653. doi: [10.1093/ije/dyy275](https://doi.org/10.1093/ije/dyy275) (trích dẫn trang 16, 45).
- [9] Salim S. Virani và cộng sự. 'Thống kê về bệnh tim và đột quy-Cập nhật năm 2020: Báo cáo từ Hiệp hội Tim mạch Hoa Kỳ'. Trong: Lưu hành (tháng 3 năm 2020), trang 640-653. doi: [10.1161/cir.0000000000000757](https://doi.org/10.1161/cir.0000000000000757) (trích dẫn trang 16).
- [10] Trevor Hastie, Robert Tibshirani và Jerome Friedman. Các yếu tố của việc học thống kê. Dòng Springer trong Thống kê. New York, NY, USA: Springer New York Inc., 2001 (trích dẫn trang 17, 63).
- [11] Dominik Janzing, David Balduzzi, Moritz Grosse-Wentrup, và Bernhard Schölkopf. 'Định lượng ảnh hưởng nhân quả'. Trong: Ann. Nhà thống kê. 41.5 (tháng 10 năm 2013), trang 2324-2358. doi: [10.1214/13-AOS1145](https://doi.org/10.1214/13-AOS1145) (trích dẫn trang 18).
- [12] Stephen L. Morgan và Christopher Winship. Phản thực tế và suy luận nhân quả: Phương pháp và nguyên tắc nghiên cứu xã hội. tái bản lần thứ 2. Phương pháp phân tích cho nghiên cứu xã hội. Nhà xuất bản Đại học Cambridge, 2014 (trích dẫn trang 18, 78).
- [13] Daphne Koller và Nir Friedman. Mô hình đồ họa xác suất: Nguyên tắc và kỹ thuật. Tính toán thích ứng và học máy. Nhà xuất bản MIT, 2009 (trích dẫn trang 21, 29, 47).
- [14] J. Peters, D. Janzing, và B. Schölkopf. Các yếu tố của suy luận nhân quả: Nền tảng và thuật toán học tập. Cambridge, MA, USA: MIT Press, 2017 (trích dẫn trang 21, 105, 107, 109).
- [15] Judea Pearl, Madelyn Glymour và Nicholas P Jewell. Suy luận nhân quả trong thống kê: Sơ lược. John Wiley & Sons, 2016 (trích dẫn trang 25).
- [16] Ngọc trai Judea. Lý luận xác suất trong các hệ thống thông minh: Mạng lưới suy luận hợp lý. San Francisco, CA, Mỹ: Morgan Kaufmann Publishing Inc., 1988 (trích dẫn trang 29).
- [17] Ngọc trai Judea. 'Suy luận nhân quả trong thống kê: Tổng quan'. Trong: Thống kê. Số 3 (2009), trang 96-146. doi: [10.1214/09-SS057](https://doi.org/10.1214/09-SS057) (trích dẫn trang 42).

- [18] Ngọc trai Judea. Nhân quả. Nhà xuất bản Đại học Cambridge, 2009 (trích dẫn trang 44, 47, 48, 59, 87, 94).
- [19] Felix Elwert và Christopher Winship. 'Thành kiến lựa chọn nội sinh: Vấn đề điều hòa đối với một biến va chạm.' Trong: Tạp chí xã hội học hàng năm 40 (2014), trang 31-53 (trích dẫn trang 44).
- [20] David Galles và Judea Pearl. 'Một đặc điểm tiên đề của các phản thực tế nhân quả'. Trong: Foundation-khoa học 3.1 (1998), trang 151-182. doi: [10.1023/A:1009602825894](https://doi.org/10.1023/A:1009602825894) (trích dẫn trang 47).
- [21] Joseph Y. Halpern. 'Tiên đề hóa lý luận nhân quả'. Trong: Kỷ yếu của Hội nghị lần thứ mười bốn về sự không chắc chắn trong trí tuệ nhân tạo. UAI'98. Madison, Wisconsin: Morgan Kaufmann Publishing Inc., 1998, trang 202-210 (trích dẫn trang 47).
- [22] Elizabeth L. Ogburn và Tyler J. VanderWeele. 'Sơ đồ nguyên nhân gây nhiễu'. Trong: Thống kê. Khoa học. 29,4 (Tháng 11 năm 2014), trang 559-578. doi: [10.1214/14-STSS501](https://doi.org/10.1214/14-STSS501) (trích dẫn trang 48).
- [23] J. Ngọc trai. 'Về quy tắc nhất quán trong suy luận nhân quả: tiên đề, định nghĩa, giả định hay định lý?' TRONG: Dịch tễ học 21.6 (tháng 11 năm 2010), trang 872-875 (trích dẫn trang 48).
- [24] Ngọc trai Judea. 'Sơ đồ nhân quả cho nghiên cứu thực nghiệm'. Trong: Biometrika 82.4 (tháng 12 năm 1995), trang 669-688. doi: [10.1093/biomet/82.4.669](https://doi.org/10.1093/biomet/82.4.669) (trích dẫn trang 55, 56).
- [25] Ilya Shpitser và Ngọc trai Judea. 'Xác định các phân phối can thiệp chung trong các mô hình nguyên nhân bán Markovian đệ quy'. Trong: Kỷ yếu Hội nghị Trí tuệ nhân tạo toàn quốc lần thứ 21 - Tập 2. AAAI'06. Boston, Massachusetts: AAAI Press, 2006, trang 1219-1226 (trích dẫn trang 57, 60).
- [26] Yimin Huang và Marco Valtorta. 'Tính toán can thiệp của Pearl đã hoàn tất'. Trong: Kỷ yếu của Hội nghị lần thứ 22 về tính không chắc chắn trong trí tuệ nhân tạo. UAI'06. Cambridge, MA, USA: AUAI Press, 2006, trang 217-224 (trích dẫn trang 57).
- [27] Jin Tian và Judea Pearl. 'Điều kiện nhận dạng chung về tác động nhân quả'. Trong: Hội nghị quốc gia về trí tuệ nhân tạo lần thứ 18. Edmonton, Alberta, Canada: Hiệp hội Trí tuệ Nhân tạo Hoa Kỳ, 2002, trang 567-573 (trích dẫn trang 59).
- [28] Ilya Shpitser và Judea Pearl. 'Xác định các phân phối can thiệp có điều kiện'. Trong: Kỷ yếu của Hội nghị lần thứ 22 về tính không chắc chắn trong trí tuệ nhân tạo. UAI'06. Cambridge, MA, USA: AUAI Press, 2006, trang 437-444 (trích dẫn trang 60).
- [29] F. Pedregosa và cộng sự. 'Scikit-learn: Học máy bằng Python'. Trong: Tạp chí Nghiên cứu Máy học 12 (2011), trang 2825-2830 (trích dẫn trang 62, 64).
- [30] Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel và Bin Yu. 'Người học siêu âm để ước tính hiệu quả điều trị không đồng nhất bằng cách sử dụng máy học'. Trong: Kỷ yếu của Viện Hàn lâm Khoa học Quốc gia 116.10 (2019), trang 4156-4165. doi: [10.1073/pnas.1804597116](https://doi.org/10.1073/pnas.1804597116) (trích ở trang 64-67).
- [31] Uri Shalit, Fredrik D. Johansson và David Sontag. 'Ước tính hiệu quả điều trị riêng lẻ: giới hạn tổng quát hóa và các thuật toán'. Trong: biên tập. của Doina Precup và Yee Whye Teh. Tập. 70. Kỷ yếu nghiên cứu học máy. Trung tâm Hội nghị Quốc tế, Sydney, Úc: PMLR, tháng 6 năm 2017, trang 3076-3085 (trích dẫn trang 65, 66).
- [32] Paul R. Rosenbaum và Donald B. Rubin. 'Vai trò trung tâm của điều xu hướng trong các nghiên cứu quan sát về tác động nhân quả'. Trong: Biometrika 70.1 (tháng 4 năm 1983), trang 41-55. doi: [10.1093/biomet/70.1.41](https://doi.org/10.1093/biomet/70.1.41) (trích dẫn trang 67).
- [33] DG Horvitz và DJ Thompson. 'Tổng quát hóa việc lấy mẫu không thay thế từ một vũ trụ hữu hạn'. Trong: Tạp chí của Hiệp hội Thống kê Hoa Kỳ 47.260 (1952), trang 663-685. doi: [10.1080/01621459.1952.10483446](https://doi.org/10.1080/01621459.1952.10483446) (trích dẫn trang 69).
- [34] Jason Abrevaya, Yu-Chin Hsu, và Robert P. Lieli. 'Ước tính hiệu quả điều trị trung bình có điều kiện'. Trong: Tạp chí Thống kê Kinh tế & Kinh doanh 33.4 (2015), trang 485-505. doi: [10.1080/07350015.2014.975555](https://doi.org/10.1080/07350015.2014.975555) (trích dẫn trang 70).
- [35] Joseph DY Kang và Joseph L. Schafer. 'Làm sáng tỏ tính chắc chắn kép: So sánh các chiến lược thay thế để ước tính trung bình dân số từ dữ liệu không đầy đủ'. Trong: Thống kê. Khoa học. 22.4 (tháng 11 năm 2007), trang 523-539. doi: [10.1214/07-STS227](https://doi.org/10.1214/07-STS227) (trích dẫn trang 70).
- [36] Paul N Zivich và Alexander Breskin. Học máy để suy luận nhân quả: về việc sử dụng các công cụ ước tính phù hợp chéo. 2020 (trích dẫn trang 70).

- [37] Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott và Dan Cervone. 'Phương pháp tự động so với phương pháp tự làm để suy luận nhân quả: Bài học rút ra từ cuộc thi phân tích dữ liệu'. Trong: *Thống kê. Khoa học.* 34.1 (tháng 2 năm 2019), trang 43-68. doi: [10.1214/18-STS667](https://doi.org/10.1214/18-STS667) (trích dẫn trang 70).
- [38] Shaun R. Seaman và Stijn Vansteelandt. 'Giới thiệu về các phương pháp mạnh mẽ gấp đôi cho dữ liệu không đầy đủ'. Trong: *Thống kê. Khoa học.* 33.2 (tháng 5 năm 2018), trang 184-197. doi: [10.1214/18-STS647](https://doi.org/10.1214/18-STS647) (trích dẫn trang 70).
- [39] Anastasios Tsiatis. *Lý thuyết bán tham số và dữ liệu còn thiếu.* Khoa học & Truyền thông Kinh doanh Springer, 2007 (trích dẫn trang 70).
- [40] James M. Robins, Andrea Rotnitzky, và Lue Ping Zhao. 'Ước tính các hệ số hồi quy khi một số biến hồi quy không phải lúc nào cũng được quan sát'. Trong: *Tạp chí của Hiệp hội Thống kê Hoa Kỳ* 89.427 (1994), trang 846-866. doi: [10.1080/01621459.1994.10476818](https://doi.org/10.1080/01621459.1994.10476818) (trích dẫn trang 70).
- [41] Heejung Bang và James M. Robins. 'Ước tính mạnh mẽ gấp đôi trong các mô hình suy luận nhân quả và dữ liệu bị thiếu'. Trong: *Sinh trắc học* 61.4 (2005), trang 962-973. doi: [10.1111/j.1541-0420.2005.00377.x](https://doi.org/10.1111/j.1541-0420.2005.00377.x) (trích dẫn trang 70).
- [42] Mark J Van Der Laan và Daniel Rubin. 'Có khả năng học tập tối đa được nhắm mục tiêu'. Trong: *Quốc tế tạp chí thống kê sinh học* 2.1 (2006) (trích dẫn trang 70).
- [43] Megan S. Schuler và Sherri Rose. 'Ước tính khả năng xảy ra tối đa có mục tiêu cho suy luận nhân quả trong nghiên cứu quan sát'. Trong: *Tạp chí Dịch tễ học Hoa Kỳ* 185.1 (tháng 1 năm 2017), trang 65-73. doi: [10.1093/aje/kww165](https://doi.org/10.1093/aje/kww165) (trích dẫn trang 70).
- [44] Mark J Van der Laan và Sherri Rose. *Học tập có mục tiêu: suy luận nhân quả cho dữ liệu quan sát và thử nghiệm.* Springer Science & Business Media, 2011 (trích dẫn trang 70).
- [45] Elizabeth A. Stuart. 'Các phương pháp so sánh để suy luận nhân quả: Đánh giá và hướng tới tương lai'. Trong: *Thống kê. Khoa học.* 25.1 (tháng 2 năm 2010), trang 1-21. doi: [10.1214/09-STS313](https://doi.org/10.1214/09-STS313) (trích dẫn trang 71).
- [46] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey và James Robins. 'Học máy kép/debiased cho các thông số cấu trúc và xử lý'. Trong: *Tạp chí Kinh tế lượng* 21.1 (2018), trang C1-C68. doi: [10.1111/ectj.12097](https://doi.org/10.1111/ectj.12097) (trích dẫn trang 71).
- [47] Chris Felton. Chernozhukov và cộng sự. về Học máy đôi / Debiased. [https://scholar.princeton.edu/sites/default/files/bstewart/files/felton.chern\\_.slides.20190318.pdf](https://scholar.princeton.edu/sites/default/files/bstewart/files/felton.chern_.slides.20190318.pdf). 2018 (trích dẫn trang 71).
- [48] Vasilis Syrgkanis. Học máy trực giao/kép. <https://econml.azurewebsites.net/spec/Estimate/dml.html>. Truy cập: ngày 16 tháng 9 năm 2020. 2019 (trích dẫn trang 71).
- [49] Dylan J. Foster và Vasilis Syrgkanis. Học thống kê trực giao. 2019 (trích dẫn trang 71).
- [50] Susan Athey và Guido Imbens. 'Phân vùng để quy cho các hiệu ứng nhân quả không đồng nhất'. Trong: *Kỷ yếu của Viện Hàn lâm Khoa học Quốc gia* 113.27 (2016), trang 7353-7360. doi: [10.1073/pnas.1510489113](https://doi.org/10.1073/pnas.1510489113) (trích dẫn trang 71).
- [51] Stefan Wager và Susan Athey. 'Ước tính và suy luận về hiệu quả xử lý không đồng nhất bằng cách sử dụng rurgeon ngẫu nhiên'. Trong: *Tạp chí của Hiệp hội Thống kê Hoa Kỳ* 113.523 (2018), trang 1228-1242. doi: [10.1080/01621459.2017.1319839](https://doi.org/10.1080/01621459.2017.1319839) (trích dẫn trang 71).
- [52] Susan Athey, Julie Tibshirani và Stefan Wager. 'Rút ngẫu nhiên tổng quát'. Trong: *Ann. Nhà thống kê.* 47,2 (Tháng 4 năm 2019), trang 1148-1178. doi: [10.1214/18-AOS1709](https://doi.org/10.1214/18-AOS1709) (trích dẫn trang 71).
- [53] Charles F. Manski. Xác định một phần phân phối xác suất: Chuỗi Springer trong thống kê. Tiếng Anh. Springer, 2003 (trích dẫn trang 73, 82).
- [54] Charles Manski. 'Giải phẫu vấn đề lựa chọn'. Trong: *Tạp chí Nhân sự* 24.3 (1989), trang 343-360 (trích dẫn trang 73, 82).
- [55] Charles F. Manski. 'Giới hạn phi tham số về hiệu quả điều trị'. Trong: *Tạp chí Kinh tế Mỹ* 80.2 (1990), tr. 319-323 (trích dẫn trang 73, 74, 79-82).
- [56] Charles F. Manski. 'Các vấn đề về nhận dạng trong khoa học xã hội'. Trong: *Phương pháp xã hội học* 23 (1993), trang 1-56 (trích dẫn trang 73, 82).

- [57] Charles F. Manski. 'Vấn đề lựa chọn'. Trong: *Những tiến bộ trong kinh tế lượng: Đại hội thế giới lần thứ sáu*. Ed. của Christopher A. Editor Sims. Tập. 1. Chuyên khảo của Hiệp hội Kinh tế lượng. Nhà xuất bản Đại học Cambridge, 1994, trang 143-170. doi: [10.1017/CCOL0521444594.004](https://doi.org/10.1017/CCOL0521444594.004) (trích dẫn trang 73, 82).
- [58] Charles F. Manski. 'Phản ứng điều trị đơn điệu'. Trong: *Econometrica* 65.6 (1997), tr. 1311-1334 (trích dẫn trang 73, 76, 82).
- [59] Charles F. Manski và John V. Pepper. 'Các biến công cụ đơn điệu: Với một ứng dụng cho Trở lại trường học'. Trong: *Econometrica* 68.4 (2000), trang 997-1010 (trích dẫn trang 73, 78, 82).
- [60] Charles F. Manski. Chính sách công trong một thế giới không chắc chắn. Nhà xuất bản Đại học Harvard, 2013 (trích dẫn trang 73, 82).
- [61] Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl và Elias Bareinboim. 'Phân tích độ nhạy của các mô hình nhân quả cấu trúc tuyến tính'. Trong: *biên tập.* của Kamalika Chaudhuri và Ruslan Salakhutdinov. Tập. 97. Kỷ yếu nghiên cứu học máy. Long Beach, California, Hoa Kỳ: PMLR, tháng 9 năm 2019, trang 1252-1261 (trích dẫn trang 84).
- [62] PR Rosenbaum và DB Rubin. 'Đánh giá độ nhạy đối với một hiệp phương sai nhị phân không được quan sát trong một nghiên cứu quan sát với kết quả nhị phân'. Trong: *Tạp chí của Hiệp hội Thống kê Hoàng gia. Series B (Phương pháp luận)* 45.2 (1983), trang 212-218 (trích dẫn trang 85).
- [63] Guido W. Imbens. 'Độ nhạy cảm với các giả định ngoại sinh trong đánh giá chương trình'. Trong: *American Economic Review* 93.2 (tháng 5 năm 2003), trang 126-132. doi: [10.1257/000282803321946921](https://doi.org/10.1257/000282803321946921) (trích dẫn trang 85).
- [64] Carlos Cinelli và Chad Hazlett. 'Tạo ý nghĩa về độ nhạy: mở rộng độ lệch biến bị bỏ qua'. Trong: *Tạp chí của Hiệp hội Thống kê Hoàng gia: Series B (Phương pháp thống kê)* 82.1 (2020), trang 39-67. doi: [10.1111/rssb.12348](https://doi.org/10.1111/rssb.12348) (trích dẫn trang 85).
- [65] Victor Veitch và Anisha Zaveri. Phân tích cảm giác và độ nhạy: Phân tích hậu kỳ đơn giản về sai lệch do nhiều không quan sát được. 2020 (trích dẫn trang 85).
- [66] W. Liu, SJ Kuramoto, và EA Stuart. 'Giới thiệu về phân tích độ nhạy đối với các yếu tố gây nhiễu không quan sát được trong nghiên cứu phòng ngừa phi thực nghiệm'. Trong: *Prev Sci* 14.6 (Tháng 12 năm 2013), trang 570-580 (trích dẫn trang 85).
- [67] Paul Rosenbaum. Nghiên cứu quan sát. Tháng 1 năm 2002 (trích dẫn trang 85).
- [68] Paul R Rosenbaum. Thiết kế nghiên cứu quan sát. Tập. 10. Springer, 2010 (trích dẫn trang 85).
- [69] Paul R Rosenbaum. Quan sát và thí nghiệm. Nhà xuất bản Đại học Harvard, 2017 (trích dẫn trang 85).
- [70] Alexander M. Franks, Alexander D'Amour và Avi Feller. 'Phân tích độ nhạy linh hoạt cho các nghiên cứu quan sát không có ý nghĩa quan sát được'. Trong: *Tạp chí của Hiệp hội Thống kê Hoa Kỳ* 0.0 (2019), trang 1-33. doi: [10.1080/01621459.2019.1604369](https://doi.org/10.1080/01621459.2019.1604369) (trích dẫn trang 85).
- [71] Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi và Lu Tian. Giới hạn về hiệu quả can thiệp có điều kiện và trung bình với các yếu tố gây nhiễu không quan sát được. 2020 (trích dẫn trang 85).
- [72] TJ Vanderweele và OA Arah. 'Các công thức sai lệch để phân tích độ nhạy của yếu tố gây nhiễu không đo lường được đối với kết quả chung, phương pháp điều trị và yếu tố gây nhiễu'. Trong: *Dịch tễ học* 22.1 (tháng 1 năm 2011), trang 42-52 (trích dẫn trang 85).
- [73] P. Ding và TJ VanderWeele. 'Phân tích độ nhạy không có giả định'. Trong: *Dịch tễ học* 27.3 (tháng 5 2016), trang 368-377 (trích dẫn trang 85).
- [74] Abraham Wald. 'Việc ghép các đường thẳng nếu cả hai biến đều có lỗi'. Trong: *Ann. Toán học. Nhà thống kê*. 11.3 (tháng 9 năm 1940), trang 284-300. doi: [10.1214/aoms/1177731868](https://doi.org/10.1214/aoms/1177731868) (trích dẫn trang 88).
- [75] Jason Hartford, Greg Lewis, Kevin Leyton-Brown và Matt Taddy. 'Deep IV: Một cách tiếp cận linh hoạt để dự đoán phản ứng'. Trong: *biên tập.* của Doina Precup và Yee Whye Teh. Tập. 70. Kỷ yếu nghiên cứu học máy. Trung tâm Hội nghị Quốc tế, Sydney, Úc: PMLR, tháng 6 năm 2017, trang 1414-1423 (trích dẫn trang 94).
- [76] Liyuan Xu, Yutian Chen, Siddarth Srinivasan, Nando de Freitas, Arnaud Doucet, và Arthur Gretton. Học các tính năng sâu trong hồi quy biến công cụ. 2020 (trích dẫn trang 94).

- [77] Niki Kilbertus, Matt J. Kusner và Ricardo Silva. Một lớp thuật toán cho biến công cụ chung Người mẫu. 2020 (trích dẫn trang 94).
- [78] Thomas Verma và Judea Pearl. 'Sự tương đương và tổng hợp các mô hình nhân quả'. Trong: Kỷ yếu của Hội nghị thường niên lần thứ sáu về sự không chắc chắn trong trí tuệ nhân tạo. UAI '90. Hoa Kỳ: Elsevier Science Inc., 1990, trang 255-270 (trích dẫn trang 102).
- [79] Morten Frydenberg. 'Thuộc tính Markov của biểu đồ chuỗi'. Trong: Tạp chí Thống kê Scandinavia 17.4 (1990), trang 333-353 (trích dẫn trang 102).
- [80] Peter Spirtes, Clark Glymour, và Richard Scheines. Nguyên nhân, dự đoán và tìm kiếm. Nhà xuất bản MIT, tháng 1. 2001 (trích dẫn trang 102, 104).
- [81] Thomas Richardson. 'Mô hình phản hồi: Giải thích và khám phá'. luận án tiến sĩ. 1996 (trích dẫn trên trang 104).
- [82] Antti Hyttinen, Patrik O. Hoyer, Frederick Eberhardt, và Matti Järvisalo. 'Khám phá các mô hình nhân quả tuần hoàn với các biến tiềm ẩn: Quy trình chung dựa trên SAT'. Trong: Kỷ yếu của Hội nghị lần thứ 29 về tính không chắc chắn trong trí tuệ nhân tạo. UAI'13. Bellevue, WA: Nhà xuất bản AUAI, 2013, trang 301-310 (trích dẫn trang 104).
- [83] Antti Hyttinen, Frederick Eberhardt, và Matti Järvisalo. 'Khám phá nhân quả dựa trên ràng buộc: Giải quyết xung đột bằng lập trình tập hợp câu trả lời'. Trong: Kỷ yếu của Hội nghị lần thứ 30 về tính không chắc chắn trong trí tuệ nhân tạo. UAI'14. Thành phố Quebec, Quebec, Canada: Nhà xuất bản AUAI, 2014, trang 340-349 (trích dẫn trang 104).
- [84] Rajen D. Shah và Jonas Peters. 'Độ cứng của kiểm tra tính độc lập có điều kiện và thước đo hiệp phương sai tổng quát'. Trong: Ann. Nhà thống kê. 48.3 (tháng 6 năm 2020), trang 1514-1538. doi: [10.1214/19-AOS1857](https://doi.org/10.1214/19-AOS1857) (trích dẫn trang 104).
- [85] Christopher Meek. 'Tính hoàn chỉnh và trung thực mạnh mẽ trong Mạng Bayesian'. Trong: Kỷ yếu Hội nghị lần thứ 11 về tính không chắc chắn trong trí tuệ nhân tạo. UAI'95. Montréal, Québec, Canada: Morgan Kaufmann Publishing Inc., 1995, trang 411-418 (trích dẫn trang 104).
- [86] Dan Geiger và Judea Pearl. 'Về logic của các mô hình nhân quả'. Trong: Kỷ yếu của Hội nghị thường niên lần thứ tư về tính không chắc chắn trong trí tuệ nhân tạo. UAI '88. NLD: North-Holland Publishing Co., 1988, trang 3-14 (trích dẫn trang 104).
- [87] G. Darmois. 'Analyse générale des liaisons stochastiques: étude particulière de l'analyse Factorielle linéaire'. Trong: Revue de l'Institut International de Statistique / Review of the International Statistical Institute 21.1/2 (1953), tr. 2-8 (trích dẫn trang 106).
- [88] VP Skitovich. 'Dạng tuyển tính của các biến ngẫu nhiên độc lập và luật phân phối chuẩn'. TRONG: Izvestia Akademii Nauk SSSR, Serija Matematicheskie 18 (1954), trang 185-200 (trích dẫn trang 106).
- [89] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, và Antti Kerminen. 'Mô hình tuần hoàn phi Gaussian tuyển tính để khám phá nguyên nhân'. Trong: Tạp chí Nghiên cứu Học máy 7.72 (2006), trang 2003-2030 (trích dẫn trang 107).
- [90] Patrik Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, và Bernhard Schölkopf. 'Khám phá nhân quả phi tuyển với các mô hình nhiều bồ sung'. Trong: Những tiến bộ trong hệ thống xử lý thông tin thần kinh. Ed. của D. Koller, D. Schuurmans, Y. Bengio và L. Bottou. Tập. 21. Curran Associates, Inc., 2009, trang 689-696 (trích dẫn trang 108).
- [91] Jonas Peters, Joris M. Mooij, Dominik Janzing, và Bernhard Schölkopf. 'Khám phá nhân quả với các mô hình tiếng ồn cộng gộp liên tục'. Trong: Tạp chí Nghiên cứu Học máy 15.58 (2014), trang 2009-2053 (trích dẫn trang 108).
- [92] Kun Zhang và Aapo Hyvärinen. 'Về khả năng nhận dạng của mô hình nhân quả phi tuyển'. Trong: Kỷ yếu của Hội nghị lần thứ 25 về tính không chắc chắn trong trí tuệ nhân tạo. UAI '09. Montréal, Québec, Canada: Nhà xuất bản AUAI, 2009, trang 647-655 (trích dẫn trang 109).
- [93] Frederick Eberhardt. 'Giới thiệu về nền tảng của khám phá nhân quả'. Trong: Tạp chí Quốc tế về Khoa học Dữ liệu và Phân tích 3.2 (2017), trang 81-91. doi: [10.1007/s41060-016-0038-6](https://doi.org/10.1007/s41060-016-0038-6) (trích dẫn trang 109).

[94] Clark Glymour, Kun Zhang, và Peter Spirtes. 'Đánh giá các phương pháp khám phá nhân quả dựa trên mô hình đồ họa'. Trong: Biên giới trong Di truyền học 10 (2019), tr. 524. doi: [10.3389/fgene.2019.00524](https://doi.org/10.3389/fgene.2019.00524) (trích dẫn trang 109).

## Danh mục theo thứ tự chữ cái

- 2SLS, 89
- công thức điều chỉnh, 11  
tổ tiên, 19
- liên kết, 4 liên  
kết nhân quả, 4, 30 liên  
kết nhiễu, 4, 30 sai phân liên  
kết,  
8
- ATT, 95
- hiệu quả điều trị trung bình (ATE),  
8, 62
- điều chỉnh cửa sau, 38 tiêu  
chí cửa sau, 37 đường dẫn  
cửa sau, 37
- Mạng Bayes, quy tắc chuỗi  
21, phân tích  
nhân tử 21, 21
- Nghịch lý Berkson, 27 con  
đường bị chặn, 25, 26, 28
- mạng Bayes nhân quả,  
35
- nhân quả  
trung bình, 8,  
62 trung bình tuân thủ,  
91 trung bình có điều kiện,  
62 cá nhân, 7, 62  
trung bình cá nhân,  
62  
cấp đơn vị, 7  
ước lượng nhân quả, 15, 33
- rừng nhân quả, 71
- đô thị nhân quả, 22  
không chặt chẽ,  
23 chặt  
chẽ, 22 cơ chế nhân quả, 34,  
41 nhân quả đủ, 101 cây  
nhân quả, 71  
nguyên nhân,  
22, 41
- con, 19 máy  
va chạm, 26 độ lệch máy va chạm, 43
- Công cụ ước tính COM,  
63 nguyên nhân chung,  
4 hỗ trợ chung, 13 so  
sánh, 49
- tác động nhân quả trung bình của trình biên dịch  
(CACE), 91
- hiệu quả can thiệp trung bình có  
điều kiện (CATE),  
62 công cụ có điều kiện, 87 mô  
hình kết quả có điều kiện,  
63  
yếu tố gây  
nhiễu, 4 nhất  
quán, 70 tương  
quan, 4 phản thực,  
8 cân bằng đồng biến, 49  
CPDAG, 102 lời  
nguyên của chiều, 13 chu kỳ, 19
- kết nối d, 29 phân  
tách d, 29 phân  
tách d, 29 quả  
trình tạo dữ liệu, 27 hậu duệ,  
19 nguyên nhân  
trực tiếp, 22, 41 đồ  
thị chu kỳ có hướng (DAG),  
19  
đô thị có hướng, 19  
đường dẫn có hướng,  
19 phép tính do,  
55 toán tử do, 32  
độ bền gấp đôi, 70 độ bền  
gấp đôi, 70
- cạnh, 19  
nội sinh, 41  
Equationtown, 53 biểu  
đò cơ bản, 102 ước  
tính, 15  
nhân quả, 15,  
33 thống kê, 15,  
33 ước lượng,  
15 ước lượng, 15,  
16 ước lượng,  
15 khả năng trao đổi, 9,  
50 ngoại sinh,  
41 ngoại suy, 13
- thực tế, 8  
độ trung thực, 100
- điều chỉnh cửa trước, 53 tiêu  
chí cửa trước, 53
- rừng ngẫu nhiên tổng quát,  
71
- giả định Markov toàn cầu,  
29
- đô thị, 19
- tiêu chí phòng ngừa,  
60 tính đồng nhất, 90
- khả năng nhận dạng, 10,  
33 nhận dạng, 10, 16, 33
- không tham số, 57  
tham số, 57, 63 câu  
trúc, 100
- vô minh, 9 điều  
vô đạo đức, 20, 26
- điều trị cá nhân  
(ITE), 7, 62
- hiệu quả điều trị trung  
bình cá nhân hóa  
(IATE), 62
- công cụ, 86 điều  
kiện, 87 biến
- công cụ, 86 can thiệp, 13  
phân phối can  
thiệp,  
32
- SCM can thiệp, trọng số xác  
suất nghịch đảo 42,  
68
- IPW, 68
- hiệu quả điều trị trung bình cục bộ  
(LATE), 91
- giả định Markov địa phương,  
20
- biến ẩn, 4
- Độ lệch M, độ  
phóng đại 44, 47,  
độ phóng đại  
43, đô thị thao tác 43, 34, 55  
Khả năng tương thích Markov, 21  
Tương đương Markov, 101  
Markovian, 41  
trung gian hòa  
giải, 43 mức tối  
chiều, 21 đặc tả sai,  
18 ước tính được hỗ trợ bởi mô hình,  
16

công cụ ước tính có sự hỗ trợ của mô hình, 16,	dường	nhận dạng cấu trúc, 100 bộ
17	dẫn, 19 bị chặn,	điều chỉnh đủ, 38
đáp ứng điều trị đơn điệu	25, 26 bị chặn	SUTVA, 14
(MTR), 76	được bỏ chặn, 25, 27	
đơn điệu, 91	được bỏ chặn, 28	ước tính khả năng tối đa được
không có giả định nào bị ràng	tích cực, 12	nhắm mục tiêu, 70
buộc, 75	sau can thiệp, 33 đồng	TARNet, súng
nút, 19 không phải	biến sau điều trị,	máy thuật ngữ 65 ,
Markovian, 41 không	44	19
tham số, 40 nhận dạng không tham số,	kết quả tiềm năng, 6	TMLE, 70
57	trước can thiệp, 33	cơ chế phân công điều
dữ liệu quan sát, 33	điểm xu hướng, 67 quần	trị, 49 nhân tử
phân phối quan sát,	thể giả, 68	rút gọn, 35 bình phương nhỏ
33	thử nghiệm đối chứng ngẫu nhiên	nhất hai giai đoạn, 89
phân rã quan sát-phản thực, 75	(RCT), 49	con đường không bị chặn, 25,
lựa chọn điều trị	thí nghiệm ngẫu nhiên,	27, 28 đứa trẻ vô tư
tối ưu	49	tiêu chí, 59
(OTS), 79	bán Markovian, 41	độ không nhiều, 11 đồ
lai truyền định hướng, 103	Nghịch lý Simpson, 1	thị vô hướng, 19 hiệu
chồng chéo, 13	khung xương,	ứng can thiệp cấp đơn vị, 7
xu hướng song song,	102 tương quan giả, 3 ước	định, 19
96 nhận dạng tham số, 57,	tính thống kê, 15, 33 mô hình	Ước tính Wald, 88
63	nhân quả cấu trúc (SCM),	Công cụ ước tính Wald, 88
cha mẹ, 19	40	Người học X, 66