



UNIVERSITAT DE
BARCELONA

Facultat de Matemàtiques
i Informàtica

DOBLE GRAU DE MATEMÀTIQUES I ENGINYERIA INFORMÀTICA

Treball final de grau

MODELS CAUSALS ESTRUCTURALS I RESULTATS POTENCIALS

Autor: Mario VILAR

Directors: Dr. Jordi VITRIÀ

Dr. Josep VIVES

**Realitzat a: Departament de Matemàtiques
i Informàtica**

Sitges, 10 de juny de 2026

TREBALL DE FINAL DE GRAU

Títol MODELS CAUSALS ESTRUCTURALS I RESULTATS POTENCIALS
(EL CIMENT DEL MÓN)

Autor Mario VILAR

Tutors Dr. Jordi VITRIÀ
Dr. Josep VIVES

Grau en MATEMÀTIQUES
ENGINYERIA INFORMÀTICA

Publicat a [Dipòsit digital](#)
Realitzat a UNIVERSITAT DE BARCELONA
Facultat de Matemàtiques i Informàtica

Maquetació Mario VILAR
amb L^AT_EX i pdfT_EX (bestclass, v. 3.0 [2026/06/02])

Curs 2025/2026
Semestre Primavera

[AMS \(2020\)](#) Primària: **62D20**, 62H22, 62A01, 62-01, 05C20
Secundària: 68T37, 68T30, 05C20, 60A05, 62P25

Keywords causal inference, structural causal models, potential outcomes, causal effect identification, d-separation, experimental and observational studies

Repositori <https://github.com/mariovilar/TFG/>

Aquesta obra està subjecta a una llicència de [Creative Commons](#)
«Reconeixement-NoComercial-SenseObraDerivada 3.0 No adaptada».



Solcant baixades, solcant costes,
tirant pel dret, sempre endavant,
potser com a la vida ve que ens toca
anar més lluny, pujar al més alt.

Enmig de fantasmagories
i dels espais i dels instants,
enmig de traves i d'ajuts benignes,
la vida corre a on vol anar.

I on vol anar, d'hoste o a casa,
és a perviure i viure en fi
com reuiu un revolt en la distància
seguint la traça del camí.

El camí (1957), Borís PASTERNAK

Aquest treball presenta una introducció teòrica a la inferència causal. El fil del treball és la distinció entre associació i causalitat: les dades observacionals poden mostrar regularitats, però no determinen per si soles què passaria sota una intervenció ni què hauria passat en un escenari contrafactual. A partir d'aquesta motivació, i a través de les diferents capes de la jerarquia causal de Pearl, es desenvolupen els models causals estructurals de Pearl, les xarxes bayesianes causals, la d -separació i el problema de la identificació d'efectes causals. També s'estudia el model de resultats potencials de Neyman–Rubin, el problema fonamental de la inferència causal, els estimands més utilitzats i el paper del mecanisme d'assignació. El treball posa especial èmfasi en la connexió entre aquests dos llenguatges i en les condicions que permeten transformar una pregunta causal en una expressió estimable a partir de dades observacionals. Es distingeixen estudis experimentals i observacionals. Abans d'estudiar el cas de les admissions a Berkeley, s'estudien aquests algorismes d'identificació i la paradoxa de Simpson. Mostrem com les inferències permeses depenen de la pregunta causal i del model assumit.

Este trabajo presenta una introducción teórica a la inferencia causal. El hilo conductor del trabajo es la distinción entre asociación y causalidad: los datos observacionales pueden mostrar regularidades, pero no determinan por sí solos qué pasaría bajo una intervención ni qué habría pasado en un escenario contrafactual. A partir de esta motivación, y a través de las diferentes capas de la jerarquía causal de Pearl, se desarrollan los modelos causales estructurales de Pearl, las redes bayesianas causales, la d -separación y el problema de la identificación de efectos causales. También se estudia el modelo de resultados potenciales de Neyman-Rubin, el problema fundamental de la inferencia causal, los estimandos más utilizados y el papel del mecanismo de asignación. El trabajo pone especial énfasis en la conexión entre estos dos lenguajes y en las condiciones que permiten transformar una pregunta causal en una expresión estimable a partir de datos observacionales. Se distinguen estudios experimentales y observacionales. Antes de estudiar el caso de las admisiones en Berkeley, se estudian estos algoritmos de identificación y la paradoja de Simpson. Mostramos cómo las inferencias permitidas dependen de la pregunta causal y del modelo asumido.

This work presents a theoretical introduction to causal inference. The common thread of the dissertation is the distinction between association and causation: observational data may show regularities, but they do not by themselves determine what would happen under an intervention or what would have happened in a counterfactual scenario. Based on this motivation, and through the different layers of Pearl's causal hierarchy, we develop Pearl's structural causal models, causal Bayesian networks, d -separation, and the problem of identifying causal effects. The Neyman-Rubin potential outcomes framework is also studied, along with the fundamental problem of causal inference, the most commonly used estimands, and the role of the assignment mechanism. The work places special emphasis on the connection between these two frameworks and on the conditions that allow transforming a causal question into an estimable expression from observational data. Experimental and observational studies are distinguished. Before studying the case of Berkeley admissions, these identification algorithms and Simpson's paradox are studied. We show how the permitted inferences depend on the causal question and the assumed model.

Agraïments

Primer de tot, al Dr. Jordi Vitrià, per despertar-me l'interès pel fascinant univers de la causalitat. Al Dr. Josep Vives, per la seva orientació, tan precisa com oportuna. Sense tots dos, aquest treball no hauria estat possible. To Dr. Bareinboim, for kindly providing some essential resources that were key to this work.

A aquells que ja no hi són: els meus avis i, en particular, el meu germanet de quatre potes. Per totes les llargues hores d'estudi que vas passar al meu costat, mirant per la finestra. Allà on sigueu, va per vosaltres.

A la meva germana, àvies i tietes, però sobretot als meus pares, pel vostre suport incondicional. Us estimo. Als meus companys de trinxera, sis anys haurien estat encara més eterns sense vosaltres; al David, camarada de mil batalles. Als amics que puc comptar amb els dits d'una mà.

A tots aquells qui m'he deixat. Com diria Fermat, aquest marge és massa estret.

I, finalment, a l'Alex. Per ser-hi sempre, per treure'm un somriure quan volia tirar la tovallola, per créixer junts. Que aquesta Facultat, que ens va unir fa ja quatre anys, no marqui sinó un revolt en la distància, seguint la traça del camí.

Resum	IV
Agraïments	V
Introducció	XIX
1 Models causals estructurals	1
1.1 Xarxes bayesianes	2
1.2 d -separació, d -connexió i $\mathcal{I}$ -mapes	4
1.3 Models causals estructurals (SCM)	8
1.4 Configuracions gràfiques elementals	12
2 Identificació d'efectes causals	14
2.1 Diagrames causals induïts per SCMs	15
2.2 Xarxes bayesianes causals (CBNs)	16
2.3 Identificabilitat i <i>do-calculus</i>	21
3 Model de resultats potencials	27
3.1 El problema fonamental de la inferència causal	27
3.2 Efecte causal i estimands	30
3.3 Mecanisme d'assignació de tractaments	33
3.4 Identificabilitat per ajustament	38
3.5 Pearl i Neyman–Rubin	39
4 Estudis experimentals i observacionals	46
4.1 Estudis experimentals	47
4.2 Estudis observacionals	49
5 Cas pràctic	53
5.1 C-boscós i hedges	53
5.2 Algorismes ID i IDC	55
5.3 Paradoxa de Simpson	61
5.4 Admissions a Berkeley (1973)	65
Conclusions	71
A Demostracions seleccionades	75
A.1 Models causals i gràfics dirigits	75
A.2 Identificabilitat d'efectes causals	80
A.3 Model de resultats potencials	82

A.4	Estudis experimentals i observacionals	87
A.5	Cas pràctic	92
B	Probabilitats i estadística	99
B.1	Mesura i funcions mesurables	99
B.2	Variables i vectors aleatoris	104
B.3	Independència i independència condicional	107
B.4	Esperances i moments	110
B.5	Models estadístics i estimació puntual	112
B.6	Biaix, variància i mitjanes mostrals	115
B.7	Confiança i testos	117
B.8	Taules de contingència i mesures d'associació	120
B.9	Distribucions	122
C	Teoria de grafs	123
C.1	Grafs i subgrafs	123
C.2	Grafs dirigits	126
C.3	Operacions estructurals sobre grafs	127
	Bibliografia	129

Índex de figures

1	Vinyl sales and Costco Wholesale's stock price	XXII
1.1	Una xarxa bayesiana de cinc variables. Font: [Pea09, Sec. 1.2.2, p. 15, figura 1.2].	4
2.1	Diagrames causals possibles per a l'exemple de la glucosa.	16
2.2	Xarxes bayesianes on es mostren possibles relacions entre el gen G , el tabaquisme T i el càncer de pulmó C . En (a), T té un efecte causal directe sobre C , mentre que en (b) no.	22
2.3	Diagrama causal semi-Markovià $\mathcal{G}_M$ sobre $\{T, Q, C\}$ (esquerra) i graf augmentat $\mathcal{G}_M^+$ restringit a G, T, Q, C (dreta) per al mateix SCM.	24
2.4	Subgrafs mutilats de $\mathcal{G} := \mathcal{G}_M^+$ que justifiquen les aplicacions successives de les regles (R2) i (R3) en la derivació de l'ajustament <i>front-door</i> en l'exemple 2.39.	26
5.1	Una CBN semi-Markoviana amb tres C-components maximals. La parella $(\mathcal{G}[\mathbf{S}_1], \mathcal{H})$, on $\mathcal{H}$ és el subgraf reduït al node M , determina un <i>hedge</i> per a $p(y \mid \text{do}(a))$	55
5.2	Exemple de Simpson quan l'estrat Z actua com a causa comuna de tractament i resultat.	64
5.3	Taxa d'admissió per gènere i departament en UCBAmissions.	66
5.4	Tres diagrames per a dues hipòtesis causals sobre el cas Berkeley. La subfigura (a) representa un SCM Markovià $\mathcal{M}_0$. Les subfigures (b) i (c) corresponen a un altre SCM $\mathcal{M}_1$ amb confusió latent entre D i Y : (b) en mostra el graf augmentat $\mathcal{G}_{\mathcal{M}_1}^+$ i (c) en mostra el diagrama causal $\mathcal{G}_{\mathcal{M}_1}$ obtingut per projecció latent sobre les variables observades.	68
B.1	Distribucions discretes i absolutament contínues.	122
C.1	Exemple de graf i les seves versions mutilades.	128

Índex de resultats

Capítol 1	MODELS CAUSALS ESTRUCTURALS	
Notació 1.1		1
Notació 1.2		1
Definició 1.3 — Pares Markovians		2
Propietat 1.4 — Regla de la cadena per a xarxes bayesianes		2
Observació 1.5		3
Teorema 1.6 — Ordered Markov conditioning [Pea88]; [Pea09, Sec. 1.2.3, Th. 1.2.6]		3
Definició 1.7 — Propietat local de Markov [Lau96, Sec. 3.2.2, p. 50]		3
Propietat 1.8		3
Teorema 1.9		3
Definició 1.10 — Xarxa bayesiana		3
Exemple 1.11		3
Definició 1.12 — Camí i camí dirigit		4
Nota 1.13		4
Definició 1.14 — Camí blocat [PJS17, Sec. 6.2, Def. 6.1]		4
Definició 1.15 — d -connexió		5
Definició 1.16 — d -separació [Pea88]; [Pea09, Sec. 1.2.3, Def. 1.2.3]		5
Propietat 1.17 — Global de Markov		5
Teorema 1.18 — [PJS17, Sec. 6.5, Th. 6.22] i [Lau96, Sec. 3.2.2, Th. 3.27]		5
Definició 1.19 — Mapa d'independències condicionals de P [KF09, Def. 3.2]		5
Definició 1.20 — I -mapa [KF09, Def. 3.3]		5
Teorema 1.21 — [KF09, Ths. 3.1, 3.2, 3.3]		5
Nota 1.22		6
Observació 1.23		6
Teorema 1.24 — [KF09, Sec. 3.3.2, Th. 3.5]		6
Definició 1.25 — Cobertura de Markov o <i>Markov blanket</i> , [Pea88]; [PJS17, Sec. 6.5, Def. 6.26]		6
Corol·lari 1.26		6
Proposició 1.27 — <i>Markov blanket</i> , [PJS17, Sec. 6.5, Prop. 6.27]		6
Definició 1.28 — I -equivalència [KF09, Sec. 3.3.4, Def. 3.9]		6
Definició 1.29 — Equivalència de Markov [PJS17, Sec. 6.5]		6
Nota 1.30		6
Nota 1.31 — Classes d'equivalència de Markov [PJS17, Sec. 6.5]		6
Definició 1.32 — Esquelet [KF09, Sec. 3.3.4, Def. 3.10]		7
Teorema 1.33 — [KF09, Sec. 3.3.4, Th. 3.7]		7
Observació 1.34		7
Definició 1.35 — Immoralitat [KF09, Sec. 3.3.4, Def. 3.11]		7

Nota 1.36	7
Teorema 1.37 — [KF09, Th. 3.8]; [PJS17, Lem. 6.25]; [VP90]	7
Definició 1.38 — Fidelitat [PJS17, Sec. 6.5, Def. 6.33]	7
Observació 1.39	7
Definició 1.40 — Model causal estructural (SCM) [Bar+22, Def. 27.1]; [Pea09, Sec. 1.4.1]	8
Observació 1.41	9
Definició 1.42 — SCM Markovià [Bar+22, Sec. 27.2]	9
Definició 1.43 — SCM semi-Markovià [Bar+22, Sec. 27.2]	9
Nota 1.44	9
Proposició 1.45 — Distribució induïda [PJS17, Prop. 6.3]	9
Nota 1.46 — C-components al nivell de l'SCM [Bar+20, Def. 18]	10
Definició 1.47 — DAG observacional induït per un SCM [Bar+20, Def. 26]	10
Observació 1.48	10
Teorema 1.49 — SCM-BN observacional [Bar+20, App. C, Th. 8]	10
Proposició 1.50 — Representabilitat [PJS17, Prop. 7.1]	10
Nota 1.51	10
Definició 1.52 — Minimalitat causal o de Markov	11
Proposició 1.53 — [PJS17, Sec. 6.5, Prop. 6.35]	11
Proposició 1.54 — Minimalitat causal [PJS17, Sec. 6.5, Prop. 6.36]	11
Corol·lari 1.55	11
Definició 1.56 — Submodel causal estructural [Pea09, Def. 7.1.2]	11
Definició 1.57 — Modularitat, segons [Nea20]	11
Definició 1.58 — Cadena (<i>chain</i>)	12
Observació 1.59	12
Definició 1.60 — Bifurcació (<i>fork</i>)	12
Definició 1.61 — Col·lisionador (<i>collider</i>)	12
Propietat 1.62	12
Definició 1.63 — Blocar per defecte	13
Exemple 1.64 — <i>Chain, fork i collider</i>	13
Nota 1.65	13
Exemple 1.66 — <i>Explaining away</i>	13

Notació 2.1	14
Definició 2.2 — Llenguatges de la jerarquia causal, [Bar+22, Sec. 27.3, Def. 27.7]	14
Teorema 2.3 — Causal Hierarchy Theorem [Bar+22, Sec. 27.3, Th. 27.1]	15
Definició 2.4 — Diagrama causal, cas Markovià [Bar+22, Sec. 27.4, Def. 27.10]	15
Definició 2.5 — Diagrama causal, cas semi-Markovià [Bar+22, Sec. 27.4, Def. 27.12]	15
Proposició 2.6 — Condició causal de Markov [PJS17, Prop. 6.31]	16
Nota 2.7	16
Exemple 2.8	16

Definició 2.9 — Operador $do()$ [Pea09, Sec. 3.2.1, Def. 3.2.1]	17
Definició 2.10 — Distribució intervinguda	17
Nota 2.11 — <i>Efectivitat i graf mutilat</i>	17
Exemple 2.12 — 2.8, continuació	17
Nota 2.13	17
Definició 2.14 — Xarxa bayesiana causal Markoviana [Bar+22, Def. 27.11, Th. 27.2]	17
Nota 2.15 — [Bar+22, Sec. 27.4]	18
Teorema 2.16 — SCM-CBN Markoviana [Bar+22, Sec. 27.4, Th. 27.2]	18
Definició 2.17 — Efecte d'una acció [Pea09, Sec. 3.2.1, Def. 3.2.1]	18
Teorema 2.18 — Factorització truncada per a CBNs Markovianes [Bar+22, Sec. 27.4, Th. 27.3]	18
Nota 2.19 — Suficiència causal i confusió	19
Definició 2.20 — C-component, [TP02; SP06]	19
Definició 2.21 — C-component maximal [SP06; SP08]	19
Lema 2.22	19
Definició 2.23 — Pares estesos [Bar+22, Sec. 27.4]	19
Definició 2.24 — Distribució semi-Markoviana [Bar+22, Sec. 27.4, Def. 27.14]	20
Observació 2.25 — [Bar+22, Sec. 27.4, Defs. 27.14–27.15]	20
Definició 2.26 — CBN semi-Markoviana [Bar+22, Sec. 27.4, Def. 27.15 i Th. 27.4]	20
Teorema 2.27 — Connexió SCM-CBN semi-Markoviana [Bar+22, Th. 27.4]	20
Nota 2.28 — Interpretació de la CBN Semi-Markoviana	21
Definició 2.29 — Criteri de la porta enrere [Pea93]; [Par23, Sec. 1.1.4]	21
Nota 2.30	21
Teorema 2.31 — <i>Back-door adjustment</i> [Pea93]; [Par23, Sec. 1.1.4, Th. 1]	21
Observació 2.32	22
Nota 2.33	22
Exemple 2.34 — Tabac i càncer de pulmó	22
Definició 2.35 — Criteri de la porta endavant (<i>front-door</i>) [Par23, Sec. 1.1.4]	23
Teorema 2.36 — <i>Front-door adjustment</i> , [Pea93]; [Par23, Sec. 1.1.4, Th. 2]	23
Teorema 2.37 — <i>do-calculus</i> , [Pea09, Sec. 3.4.2, Th. 3.4.1]; [Par23, Sec. 1.1.4, Th. 3]	23
Nota 2.38	23
Exemple 2.39 — Tabac i càncer de pulmó, continuació	23
Nota 2.40	25
Definició 2.41 — Identificabilitat d'efectes causals [SP06, Def. 2]; [Bar+22, Sec. 27.4, Def. 27.16]	25
Nota 2.42	26

Notació 3.1 — Convenció notacional del capítol	27
Notació 3.2 — Tractaments i resultats potencials en població finita	27
Definició 3.3 — No-interferència entre unitats	28
Exemple 3.4 — No-interferència	28
Nota 3.5 — <i>Spillover</i>	28

Definició 3.6 — Absència de múltiples versions	28
Exemple 3.7	28
Definició 3.8 — SUTVA	28
Definició 3.9 — Consistència	28
Nota 3.10	28
Definició 3.11 — Resultats potencials observats i no observats [IR15, Secs. 1.7 i 2.6]	28
Nota 3.12 — Convenció per als resultats observats	29
Observació 3.13 — [IR15, Secs. 1.7 i 2.6]	29
Exemple 3.14 — de l'aspirina	29
Nota 3.15 — Problema fonamental de la inferència causal, [Hol86]	29
Definició 3.16 — Covariant [IR15, Sec. 1.8]	30
Exemple 3.17 — de l'aspirina, continuació	30
Notació 3.18 — Matriu de covariants, [IR15, Secs. 1.8 i 1.10]	30
Nota 3.19 — Efecte causal, [IR15, Secs. 1.3 i 1.10]	30
Definició 3.20 — Efecte causal individual, [IR15, Sec. 1.10]	31
Nota 3.21	31
Definició 3.22 — SATE, [Ima13, Sec. 1.3, p. 3]	31
Definició 3.23 — SATT, [Ima13, Sec. 1.3, p. 3]	31
Notació 3.24 — Superpoblació	31
Definició 3.25 — PATE i PATT	32
Nota 3.26 — Mostra finita i població	32
Definició 3.27 — CATE	32
Definició 3.28 — QTE, [Ima13, Sec. 1.3, p. 3]	33
Observació 3.29	33
Nota 3.30 — Mostra de superpoblació i assignació finita	33
Definició 3.31 — Mecanisme d'assignació [IR15, Sec. 3.3, Def. 3.1]	33
Observació 3.32	34
Definició 3.33 — Probabilitat d'assignació individual, [IR15, Sec. 3.3, Def. 3.2]	34
Nota 3.34	34
Definició 3.35 — Mecanisme individualístic, [IR15, Sec. 3.4, Def. 3.4]	34
Observació 3.36 — Conjunt admissible	35
Nota 3.37	35
Definició 3.38 — Mecanisme probabilístic	35
Definició 3.39 — Assignació no confosa en població finita, [IR15, Sec. 3.4]	35
Definició 3.40 — Valor de propensió per a poblacions finites [IR15, Sec. 3.3]	35
Nota 3.41	36
Proposició 3.42	36
Definició 3.43 — Valor de propensió de superpoblació, [IR15, Sec. 3.5, Def. 3.7]	36
Definició 3.44 — Assignació probabilística en superpoblació, [IR15, Sec. 3.5, Def. 3.8]	37
Definició 3.45 — Assignació no confosa en superpoblació, [IR15, Sec. 3.5, Def. 3.9]	37
Definició 3.46 — Positivitat o superposició	37
Definició 3.47 — Assignació fortament ignorable, [RR83, Sec. 1.3]	37

Nota 3.48	37
Observació 3.49	37
Nota 3.50	38
Definició 3.51 — Regularitat en població finita	38
Definició 3.52 — Identificabilitat, [HR20, Fine Point 3.1, p. 29]	38
Proposició 3.53 — Identificació per ajustament o estandardització, [HR20, Sec. 2.3 i Technical Point 4.1]	38
Notació 3.54 — Notació global	39
Notació 3.55 — Submodel i resultat potencial, [Pea09, Def. 7.1.2]; [GP98, Defs. 2.2–2.4]	39
Definició 3.56 — Enunciat contrafactual i conseqüència semàntica	40
Nota 3.57 — Lectura estructural dels contrafactuals	40
Observació 3.58 — Connexió amb els resultats potencials	40
Propietat 3.59 — Composició, [GP98, Sec. 3]	41
Nota 3.60 — Interpretació de la composició	41
Corol·lari 3.61 — Consistència	41
Propietat 3.62 — Efectivitat, [GP98, Sec. 3]	41
Observació 3.63	41
Propietat 3.64 — Reversibilitat, [GP98, Sec. 3]	41
Teorema 3.65	41
Teorema 3.66	41
Teorema 3.67	42
Definició 3.68 — Ordenació causal	42
Nota 3.69 — [GP98, Sec. 3, Def. 3.2]	42
Propietat 3.70 — Definitud	42
Propietat 3.71 — Unicitat	42
Teorema 3.72	42
Nota 3.73	42
Nota 3.74 — Eliminació de contrafactuals niats [Bar+20, Th. 7]	42
Exemple 3.75 — Tabac i càncer de pulmó, continuació	43
Nota 3.76	45

Notació 4.1	46
Definició 4.2 — Experiment completament aleatoritzat [IR15, Def. 4.2 i Sec. 4.4]	47
Proposició 4.3 — Aleatorització i no-confusió [IR15, Sec. 4.7]	47
Nota 4.4	48
Definició 4.5 — Estimador de diferència de mitjanes [IR15, Sec. 6.3]	48
Nota 4.6	48
Teorema 4.7 — Absència de biaix de l'estimador de Neyman [IR15, Sec. 6.3, Th. 6.1]	48
Notació 4.8 — Variàncies de mostra finita	48
Teorema 4.9 — Variància de Neyman [IR15, Sec. 6.4, Th. 6.2]	49

Observació 4.10	49
Corol·lari 4.11 — Estimador conservador de la variància	49
Observació 4.12	49
Definició 4.13 — Valor d'equilibratge [RR83]; [HR20, Technical Point 15.1]	51
Proposició 4.14 — Equilibratge, ignorabilitat i positivitat [IR15, Sec. 12.3, Lema 12.2]	51
Proposició 4.15 — El valor de propensió és d'equilibratge [RR83, Th. 1]; [IR15, Lema 12.1]	51
Proposició 4.16 — [RR83, Th. 2]; [IR15, Sec. 12.3, Lema 12.3]	51
Definició 4.17 — Estimador per subclassificació	52
Proposició 4.18 — Ponderació inversa i PATE [IR15, Sec. 17.8]	52
Nota 4.19	52

Capítol v	CAS PRÀCTIC
Notació 5.1 — Convenció notacional del capítol	53
Definició 5.2 — Bosc i arbre causal, [SP06; SP08]	54
Nota 5.3 — [SP08]	54
Definició 5.4 — C-arbre, <i>C-tree</i> [SP06, Def. 4]; [SP08, Def. 11]	54
Definició 5.5 — C-bosc, <i>C-forest</i> [SP06, Def. 5]; [SP08, Def. 15]	54
Definició 5.6 — <i>Hedge</i> [SP06, Def. 6]; [SP08, Def. 16]	54
Observació 5.7 — [SP06; SP08; Pea09]	54
Exemple 5.8	55
Teorema 5.9 — Criteri d'identificació de <i>hedges</i> [SP06, Cor. 3]; [SP08, Th. 19]	55
Lema 5.10 — Factorització per C-components, [TP02, Lem. 1]; [Bar+20, Ap. D, Th. 11]	56
Teorema 5.11 — [SP06, Lem. 3, Th. 5, Cor. 2]	56
Lema 5.12 — [SP06, Lem. 5]	57
Lema 5.13 — [SP06, Lem. 6]	57
Lema 5.14 — [SP06, Lem. 4]	58
Teorema 5.15 — [SP06, Th. 6]; [SP08, Th. 39]	58
Lema 5.16 — [SP06, Lem. 8]	58
Lema 5.17 — Reducció del cas condicionat [SP12, Cor. 1 i Th. 6]	59
Teorema 5.18 — Correctesa i completesa d'IDC [SP12, Th. 7 i Th. 8]	60
Definició 5.19 — Reversió d'associació [Sam93, p. 81]	61
Exemple 5.20	61
Nota 5.21 — [GM87]	62
Definició 5.22 — Mesura d'associació [GM87]	62
Definició 5.23 — Paradoxa d'amalgamació	62
Corol·lari 5.24 — Condició d'enllaç doble [Sam93, pp. 81–83]	63
Exemple 5.25 — 5.20, continuació	63
Nota 5.26 — Disseny de mostreig	63
Proposició 5.27 — <i>Sure thing principle</i> [Pea09, Sec. 6.1.4]	64
Exemple 5.28 — 5.20, continuació	64
Observació 5.29	65

Observació 5.30 — El <i>bow</i> i l'error de causaleffect	69
---	----

Capítol B	PROBABILITATS I ESTADÍSTICA
Definició B.1 — Mesura i interval a $\mathbb{R}^n$ [Vil23, Sec. 1, Def. 1.1]	99
Definició B.2 — Volum d'un interval	99
Definició B.3 — Mesura exterior de Lebesgue	99
Propietat B.4	99
Propietat B.5 — Invariant per translacions	100
Propietat B.6 — σ -subadditivitat	100
Nota B.7	100
Definició B.8 — σ -àlgebra	100
Definició B.9 — Mesura positiva	100
Definició B.10 — Continuitat absoluta i mesura dominant [Dur19, Ap. A.4]; [Nag23]	100
Teorema B.11 — Radon–Nikodym [Dur19, Ap. A.4, Thm. A.4.8]; [Nag23]	100
Observació B.12	101
Definició B.13 — Conjunt mesurable Lebesgue	101
Nota B.14	101
Definició B.15 — Mesura de Lebesgue [Vil23, Sec. 1, Def. 1.11]	101
Teorema B.16 — Teorema de Carathéodory	101
Notació B.17 — Almost everywhere, a.e.	101
Propietat B.18	101
Proposició B.19	102
Teorema B.20	102
Proposició B.21	102
Definició B.22 — Funció mesurable [Vil23, Sec. 2, Prop. 2.1 i Def. 2.2]	102
Proposició B.23	103
Proposició B.24	103
Propietat B.25	103
Definició B.26 — Integral de Lebesgue [Vil23, Sec. 3, Def. 3.8]	103
Definició B.27 — Funció integrable [Vil23, Sec. 3, Def. 3.9]	103
Proposició B.28	103
Proposició B.29	104
Definició B.30 — Àlgebra de Borel	104
Definició B.31 — Espai de probabilitat [San99, Sec. 1.2]; [Vil24b, Sec. 1, Def. 1.1.1]	104
Definició B.32 — Variable aleatòria [Vil24b, Sec. 2, Def. 2.1.2]	105
Definició B.33 — Llei d'una variable aleatòria [San99, Sec. 2.2, Def. 2.2]	105
Definició B.34 — Variable aleatòria real	105
Proposició B.35	105
Definició B.36 — Variable aleatòria discreta	105
Definició B.37 — Funcions de distribució	105
Propietat B.38 — Propietats de les CDF	106

Definició B.39 — PMF i PDF [Vil24b, Sec. 2, Defs. 2.2.3, 2.4.1 i 2.4.2]	106
Definició B.40 — Variable aleatòria absolutament contínua	106
Definició B.41 — Vector aleatori real	106
Definició B.42 — Distribució conjunta i marginal	106
Definició B.43 — Independència estadística	107
Teorema B.44	107
Definició B.45 — Probabilitat condicionada	107
Teorema B.46 — Distribució condicional regular	107
Teorema B.47 — Probabilitats compostes	107
Teorema B.48 — Probabilitats totals	108
Teorema B.49 — Regla de Bayes	108
Definició B.50 — Independència condicional, [Lau96, Sec. 3.1, pp. 28–29]	108
Observació B.51	108
Teorema B.52	109
Teorema B.53	109
Corol·lari B.54	109
Definició B.55 — Model d'independència, [Lau96, Sec. 3.1, pp. 30–31]	109
Nota B.56	109
Definició B.57 — Models d'independència probabilística, [Lau96, Sec. 3.1, pp. 28–31]	110
Notació B.58	110
Proposició B.59	110
Definició B.60 — Esperança matemàtica [Vil24b, Sec. 3, Def. 3.1.6]	110
Definició B.61 — Esperança condicionada [Dur19, Sec. 4.1]	110
Proposició B.62 — Esperances iterades [Dur19, Sec. 4.1, Thm. 4.1.13]	110
Proposició B.63	111
Proposició B.64 — Esperança d'una variable absolutament contínua	111
Procés B.65 — Càlcul de moments [Vil24b, Sec. 3, Defs. 3.2.1 i 3.2.4]	111
Proposició B.66 — Desigualtat de Jensen	112
Proposició B.67	112
Proposició B.68	112
Definició B.69 — Covariància	112
Definició B.70 — Model estadístic [Vil24a, Sec. 1.1, Def. 1.1.1]	112
Definició B.71 — Model paramètric [Vil24a, Sec. 1.1, Def. 1.1.2]	112
Definició B.72 — Funció de versemblança [Vil24a, Sec. 1.2, Def. 1.2.1]	113
Observació B.73	113
Definició B.74 — Model exponencial [Vil24a, Sec. 1.2, Def. 1.2.4]	113
Definició B.75 — Estadístic [Vil24a, Sec. 2.1]	113
Definició B.76 — Estadístic suficient [Vil24a, Sec. 1.3, Def. 1.3.3]	114
Teorema B.77 — Factorització de Neyman–Fisher [Vil24a, Sec. 1.3, Thm. 1.3.4]	114
Definició B.78 — Model regular [Vil24a, Sec. 1.3, Def. 1.3.6]	114
Definició B.79 — Informació de Fisher [Vil24a, Sec. 1.4, Def. 1.4.1]	114
Definició B.80 — Estimador [Vil24a, Sec. 2.1, Def. 2.1.1]	114

Definició B.81 — Funció de pèrdua [Vil24a, Sec. 2.1]	115
Definició B.82 — Funció de risc [Vil24a, Sec. 2.1, Def. 2.1.3]	115
Definició B.83 — Estadístic integrable [Vil24a, Sec. 2.1, Def. 2.1.6]	115
Definició B.84 — Biaix i absència de biaix [Vil24a, Defs. 2.1.7–2.1.8]	115
Definició B.85 — Variància i risc quadràtic [Vil24a, Defs. 2.1.3 i 2.1.8]	115
Corol·lari B.86 — Mitjana mostral [Vil24a, Sec. 2.1, Cor. 2.1.13]	115
Corol·lari B.87 — Variància mostral corregida [Vil24a, Prop. 2.1.14 i Cor. 2.1.15]	116
Definició B.88 — UMV [Vil24a, Sec. 2.2, Def. 2.2.1]	116
Definició B.89 — Estadístic complet [Vil24a, Sec. 2.2, Def. 2.2.2]	116
Definició B.90 — Regió de confiança [Vil24a, Sec. 3.1, Def. 3.1.1]	117
Definició B.91 — Funció pivotant [Vil24a, Sec. 3.1, Def. 3.1.2]	117
Definició B.92 — Hipòtesis [Vil24a, Sec. 4.1, Defs. 4.1.2–4.1.4]	118
Definició B.93 — Test, potència i errors [Vil24a, Sec. 4.1, Defs. 4.1.9–4.1.11]	118
Observació B.94 — Intervals de confiança i tests [Vil24a, Sec. 4.1, Obs. 4.1.12]	118
Definició B.95 — UMP [Vil24a, Sec. 4.1, Def. 4.1.13]	119
Observació B.96 — Existència i asimetria dels tests UMP [Vil24a, Sec. 4.1, Obs. 4.1.14]	119
Lema B.97 — Neyman–Pearson [Vil24a, Sec. 4.1, Lema 4.1.15]	119
Definició B.98 — Contrast de Pearson d’independència	120
Definició B.99 — Odds ratio	120
Definició B.100 — Estimador de Mantel–Haenszel [MH59]	121
Definició B.101 — Contrast de Cochran–Mantel–Haenszel [Coc54; MH59]	121
Observació B.102	121

Capítol C

TEORIA DE GRAFS

Definició C.1 — Graf, [BM08]	123
Definició C.2 — Conceptes bàsics	123
Definició C.3 — Graf simple i complet	124
Nota C.4 — Notació	124
Definició C.5 — Subgraf	124
Definició C.6 — Graf simple subjacent	124
Definició C.7 — Subgraf induït	124
Nota C.8 — Notació d’eliminació	124
Definició C.9 — Passeig, recorregut i camí [BM08]	125
Observació C.10 — Notació de camins	125
Definició C.11 — Connexió i components	125
Definició C.12 — Cicle	125
Definició C.13 — Graf dirigit	126
Observació C.14 — Tipus d’arestes i grafs mixtos	126
Definició C.15 — Conceptes de digrafs	126
Definició C.16 — Relacions familiars en digrafs	126
Notació C.17	126

Nota C.18 — Convenció causal	126
Nota C.19	127
Definició C.20 — DAG i ordre topològic	127
Teorema C.21	127
Definició C.22 — Conjunt terminal, [BM08]	128
Definició C.23 — Connexió en grafs dirigits, [BM08]	128
Notació C.24 — Operacions de mutilació, [SP08, Sec. 2]; [Pea09, Sec. 3.2.1]	128

Introducció

La naturalesa de l'experiència consisteix en això: recordem haver tingut exemples freqüents de l'existència d'una mena d'objectes; recordem també que els individus pertanyents a una altra mena d'objectes han acompanyat sempre els primers, i que han existit segons un ordre regular de contigüïtat i successió amb ells. D'aquesta manera, recordem haver vist aquesta mena d'objectes que anomenem flama, i haver sentit aquesta mena de sensació que anomenem calor. I de la mateixa manera recordem mentalment la seva conjunció constant en tots els casos passats. Sense més preàmbuls, anomenem als uns causa i als altres efecte i inferim l'existència dels uns dels altres.

David HUME, *Tractat sobre la naturalesa humana* (1739)

En la seva obra cabdal, *Tractat sobre la naturalesa humana*, David Hume va postular que allò que observem directament no és una connexió necessària entre dos fets, sinó una successió regular d'esdeveniments. Veiem que es frega un llumí i que, gairebé a l'instant, el llumí s'encén¹; però la connexió que converteix el primer fet en causa del segon no apareix davant nostre de la mateixa manera que apareixen la flama o la calor. Aquesta tensió entre el que observem i el que afirmem quan diem «això ha causat allò» és el punt de partida d'aquest treball. A [Hum00], Hume va descriure la causalitat d'una manera molt il·lustrativa: «el ciment de l'univers».

La metàfora no és exagerada. Sense causalitat, cap de les nostres accions respondria als nostres pensaments i, per tant, no existiria el concepte de responsabilitat moral. No hi hauria cabuda a poders judicials ni a altres sistemes legals, ja que el *culpable* és únicament aquell o aquella que ha *causat* un dany o perjudici. Tampoc no existirien ni la ciència ni la tecnologia. La causalitat ens interpel·la, fins i tot, en quelcom tan fonamental com el lliure albir, ja que està estretament lligada a la nostra idea d'agència i, per tant, a la llibertat humana. A la seva *Ètica* [Spi18, I, Ax. 3], Spinoza diu que «si hi ha una causa determinada, se segueix necessàriament un efecte, i viceversa: si no es dona aquesta causa, és impossible que se segueixi cap efecte». Però si les causes necessitessin forçosament els seus efectes, no estaria tot determinat, esdevenint nosaltres esclaus de la necessitat? I si les nostres accions fossin pura contingència, com escaparíem de l'atzar en el món?

Malgrat que el raonament causal ens acompanya de molt abans que la revolució cognitiva [GL18], no hem sabut formalitzar el seu llenguatge fins fa ben poc. Perquè per establir causes i efectes hem d'anar amb compte amb les afirmacions que fem. En aquest treball volem explorar com la inferència causal ens ofereixen un llenguatge rigorós per donar forma a inquietuds tan humanes com «per què ha passat això?», «què passaria si féssim això altre?» o «què hauria passat si haguéssim actuat diferent?» en expressions que puguin ser analitzades amb precisió. Probablement, la pregunta més important que tractarem és com podem passar de dades que mostren associacions a enunciats que parlen d'intervencions, explicacions o mons que no hem observat.

¹ El que resulta inobservable, segons el plantejament humià, és el vincle mateix pel qual la fricció causa que el llumí s'encengui. La connexió causal que hi suposem no és un element perceptible de la realitat, sinó una inferència de la nostra ment.

En aquesta introducció partirem de la intuïció humana de causa i efecte i mostrarem per què la correlació no és suficient. En aquest sentit, el fil conductor serà entendre per què les dades observacionals no parlen per si soles, quines hipòtesis cal fer explícites per donar-los una interpretació causal i en quines condicions una pregunta causal es pot convertir en una quantitat estimable. Les dades poden contenir associacions, correlacions i independències; la causalitat apareix quan hi afegim hipòtesis, coneixement extern i una estructura causal que digui com s'han generat les dades que observem. En el llenguatge de Pearl, la inferència causal combina supòsits, preguntes i dades per obtenir, quan és possible, un estimand i finalment una estimació [PM18]. A la memòria, introduïrem models que representin mecanismes causals i arribarem a la connexió entre els dos grans llenguatges que estructuren bona part de la causalitat moderna: els models causals estructurals de Pearl i el model de resultats potencials de Neyman-Rubin.

PER QUÈ ÉS RELLEVANT LA CAUSALITAT?

Considerem un escenari causal, a tall de reflexió. Imaginem que es posa a prova un sistema automàtic de frenada en una sèrie de vehicles, i que el 50% d'aquests acaben patint accidents greus durant les setmanes posteriors. Aquest resultat podria semblar alarmant, i seria fàcil concloure que el sistema és defectuós o fins i tot perillós. Però podem afirmar directament que provoqui els accidents?

La primera observació rellevant és que aquest cas mostra fins a quin punt les afirmacions causals són habituals en el nostre dia a dia. Si diem que un sistema de seguretat provoca accidents, estem formulant una afirmació de caràcter causal. De fet, el que estem dient és que hi ha una relació de *causa i efecte*. De la mateixa manera, quan afirmem que una avaria paralitza Rodalies o que unes certes gravacions desencadenen una crisi de govern, estem fent servir verbs causals: «paralitzar», «desencadenar». Aquests verbs expressen accions que, d'alguna manera, provoquen una conseqüència.

Tornant al sistema de frenada, pel simple fet que un accident s'esdevingui després de l'activació del sistema de frenada no implica necessàriament que aquest en sigui la causa. I si els vehicles inclosos a la prova ja tenien, per exemple, un defecte estructural al xassís amb capacitat de provocar un alt índex d'accidents? Fins i tot si aquest no fos el cas, hi podria haver una explicació alternativa molt més plausible: potser, durant la prova, hi va haver una tempesta sobtada que va afectar la conducció a la zona on s'estaven fent els tests. En aquest escenari, la causa dels accidents no seria el sistema de frenada, sinó un «factor ambiental» completament aliè a l'experiment.

Una distinció important que cal fer és entre els conceptes de *causalitat* i *correlació*. A l'estadística, la correlació es defineix com una relació recíproca o mútua entre dues o més coses, o el grau d'interdependència entre diverses variables. Hume va capgirar la relació entre el problema de la causalitat i el de la correlació: segons ell, és la correlació la que explica la causalitat, i no a l'inrevés. La teoria de Hume sostenia que només a través de l'observació de regularitats —la *repetició*—, obtenim qualsevol idea de causalitat. En altres paraules, si un tipus de succés és seguit habitualment per un altre —ho anomenen *conjunció constant*— i es troba en relacions similars de precedència i contigüitat en l'espai i el temps, podem creure que el primer és la causa del segon [Hum75, Seccions II-VI-XIV, part 3].

El problema d'aquesta aproximació ve quan comencem a preguntar-nos com dos esdeveniments que estan connectats causalment poden estar-ho si la connexió causal no forma part de la nostra experiència. En efecte, Hume admet que «semblen conjuntats, però mai connectats» [Hum75, Secció VII, part 2].

Així, el concepte de causalitat segons la seva manera de pensar és molt fràgil, perquè no distingeix entre una connexió genuïnament causal i una simple coincidència. En efecte, si algú que aposta sempre guanya quan porta una corbata de la sort, l'humà hauria de dir que la corbata de la sort causa cada cop la victòria en l'aposta [MA13]. La causalitat, segons aquesta visió, és una relació completament *contingent*. Aquest error es produeix perquè Hume ignora la naturalesa contrafactual de la causalitat: per definir un efecte real, hem de comparar el que ha passat amb el que hauria passat si la causa hagués estat una altra [Hol86]. En parlarem més endavant.

Malgrat haver-ne fet un petit tast i ser terriblement interessants, queden fora l'abast d'aquest treball un ampli ventall de teories de la causalitat: des de l'esmentat regularisme de Hume, passant pel mecanicisme i el pluralisme, fins les teories dels poders causals d'Aristòtil. Convé haver-les introduït, però, perquè el debat al voltant d'aquest concepte travessa un bon grapat de disciplines científiques i socials: moltes de les seves afirmacions centrals són, precisament, causals.

LA CAUSALITAT A LA CIÈNCIA

La història moderna de la causalitat en estadística és una de les moltes innovacions que es produeixen en el si d'una discussió gairebé tabú. D'una banda, els primers estadístics, influenciats per l'empirisme de Hume i pel pensament positivista, van rebutjar efusivament la idea que la ciència —i particularment l'estadística— pogués o hagués de fer afirmacions causals. Això es basava en la idea que només es poden registrar correlacions observables, i que la causalitat era una noció metafísica, especulativa o inobservable. D'aquí neix la famosa frase «correlació no implica causalitat». Com Francis Galton, Karl Pearson considerava la causalitat un cas especial de correlació i va proposar elaborar *taules de contingència*, on cada cop que un tipus de fenomen es correlacionava amb un altre es marcava a la taula.

Aquest enfocament s'utilitza encara a dia d'avui (cf. Secció B.8) i pot ajudar en alguns casos, però no en tots, ja que no distingeix les correlacions genuïnament causals de les que són purament accidentals. Per exemple, a [Vig15] (cf. figura 1) podem veure com durant diversos anys hi ha hagut una forta correlació entre l'augment en la venda de LPs i l'increment en el preu de les accions de Costco.

Tot i que no podem descartar que l'audiència de vinils tingui poders inexplicables sobre Wall Street, tot apunta que es tracta d'una *correlació espúria* (*spurious correlation*). En definitiva, les dades per si soles, fins i tot quan disposem de conjunts molt amplis i detallats, posen de manifest que considerar només les correlacions és un error, ja que la lectura pot ser diferent quan s'inclouen altres factors.

D'ençà del segle xx, el concepte de causalitat s'ha anat formalitzant progressivament, especialment amb les aportacions de Neyman i Rubin i els seus «resultats potencials». Jerzy Neyman va ser el primer a proposar una formalització matemàtica del següent escenari: per a cada *unitat* d'anàlisi, com ara una persona, una escola o una parcel·la, es poden definir dos resultats potencials: un si rep un tractament

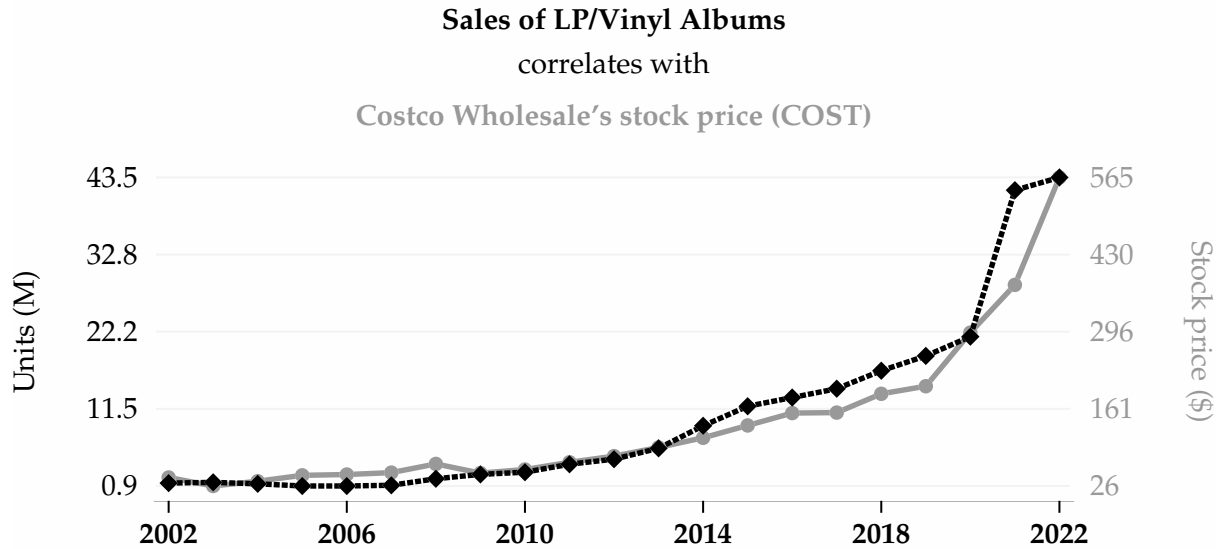


Figura 1: The Groovy Relationship Between Vinyl Sales and Costco Wholesale's Stock Price, [Fig15].

(grup de *tractament*), i un altre si no el rep (grup de *control*). Així doncs, el marc de *resultats potencials* o *model de Neyman-Rubin*, permet concebre la causalitat com la diferència entre el que ha passat i el que hauria pogut passar en una situació alternativa. Aquesta idea planteja la causalitat amb *contrafactuals*, dels quals parlarem més avall. Tot i que no podem observar simultàniament aquests dos mons en una única unitat —el que es coneix com el *problema fonamental de la inferència causal* [Hol86]—, aquest model ens permetrà aproximar el resultat a partir d'un conjunt d'unitats.

Tot i això, abans del segle xx ja podem identificar indicis de la definició de causalitat basada en resultats potencials tant en el pensament d'alguns científics i filòsofs. Per exemple, es pot reconèixer una concepció incipient d'aquest enfocament —encara que aleshores no fos designada amb aquest nom— en aquestes paraules de John Stuart Mill [Mil73, p. 326-327]:

If a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death.

Aquesta idea tan potent és el que fonamenta la causalitat entesa a partir de *contrafactuals*. Ronald Fisher, paral·lelament a la feina de Neyman, va desenvolupar un mètode que buscava neutralitzar automàticament qualsevol factor que pogués distorsionar els resultats en un experiment. Com que és impossible identificar tots els elements que hi influeixen, va proposar confiar en l'assignació aleatòria per repartir-los de manera equitativa entre els grups d'estudi. Aquesta innovació va cristal·litzar en el que avui coneixem com a *assaig clínic amb assignació aleatòria* (RCT). Fins i tot a dia d'avui, aquesta tècnica continua considerant-se el patró d'or per controlar biaixos estadístics. L'aleatorietat del *mecanisme d'assignació* ens permet utilitzar la diferència de les seves mitjanes observades com un estimador no esbiaixat l'efecte causal. Així doncs, en aquest tipus d'assajos avaluem l'efecte mitjà sobre una gran població, però no assegurem què passarà amb cada cas particular; és a dir, encara que es pugui dibuixar una tendència causal global, no estem donant informació sobre la resposta d'un individu concret.

Per tant, ens preguntem: hi ha maneres de parlar de relacions causals més enllà de la comparació de

mitjanes? El model de Neyman-Rubin ens convida a pensar que sí, sempre que trobem la manera de formular el resultat que no hem observat. Però aquesta no és l'única via. El segon gran camí, que serà el punt de partida formal d'aquesta memòria, és representar explícitament els mecanismes causals mitjançant grafs i equacions estructurals.

Aquesta diferència d'ordre és deliberada: la introducció presenta primer el marc de resultats potencials per mostrar una de les grans respostes modernes al problema causal, però el desenvolupament formal començarà pel llenguatge gràfic de Pearl perquè és una aproximació més intuïtiva i, al seu torn, ens permetrà plantejar amb precisió el problema de la identificació d'una manera més orgànica.

LA JERARQUIA CAUSAL DE PEARL

Als anys vuitanta es va impulsar una formalització de la causalitat diferent de la de Neyman-Rubin, que acabaria sent molt influent. Aquest corrent va sorgir de l'interès dels investigadors en Intel·ligència Artificial per crear mètodes capaços d'automatitzar el raonament en situacions d'incertesa. Aviat es va fer palès que les regles per al diagnòstic i altres tasques comunes dels experts gairebé sempre admeten excepcions, i que els sistemes experts necessitaven una manera de gestionar aquesta incertesa sense renunciar a representar dependències entre variables.

Tot i que personatges com Sewall Wright [Wri21] i Hans Reichenbach [Rei56] van posar les bases, va ser Judea Pearl qui va dissenyar un formalisme que facilitava tant la representació i el raonament efectiu amb probabilitats com la inclusió i el tractament de vincles causals [Pea88; Pea93; Pea95]; [Pea09, Secs. 1.2.2–1.3]; [PGJ16; PM18]. Tot i que la seva aportació va començar amb les *xarxes bayesianes*, Pearl va fer el salt decisiu en estendre aquesta idea a les *xarxes bayesianes causals* i els *diagrames causals*. Mentre que les primeres es limiten a descriure associacions i dependències probabilístiques basades en l'observació —*què passa si veiem X*—, les segones ens permeten respondre, per primera vegada, al què passarà si actuem sobre X.

Perquè aquest model funcioni, l'estructura es construeix mitjançant un graf on cada variable es representa com un node i els mecanismes d'influència es codifiquen com a arestes dirigides que van de la causa a l'efecte. El graf es basa en la idea senzilla que, si fem una intervenció quirúrgica per modificar la causa, això pot provocar un canvi en l'efecte. Per exemple, si una aixeta està completament tancada, obrir-la incrementa la probabilitat que un recipient s'ompli d'aigua. Diem que la intervenció és «quirúrgica» perquè es dissenya per variar només un factor, deixant la resta tan inalterats com sigui possible. Evidentment, podria passar que l'obertura de l'aixeta anés acompanyada d'altres canvis: potser, en obrir l'aixeta, també s'acciona automàticament la vàlvula de seguretat per evitar que el recipient es desbordi. És per aquesta raó que és necessari un model que representi les múltiples connexions entre els diferents factors.

El diagrama causal assenyala on una intervenció sobre una variable podria alterar la distribució d'una altra². Mostra no només patrons observats, sinó també una hipòtesi sobre com es generarien les dades

² Curiosament, a [Rei56] Reichenbach va defensar erròniament que la correlació observacional on es dona que $P(Y | X) > P(Y)$ implicava una relació causal (directa o amb causa comuna). [PM18] exposa com això és insuficient i que cal introduir

si canviéssim alguna peça del sistema. Aquesta és la idea que permet organitzar les preguntes causals en tres esgraons que anomenem *jerarquia de Pearl*.

Les capes que hi ha al darrere d'aquesta jerarquia són les següents: $\mathcal{L}_1$, la de les consultes a nivell associatiu, com ara $P(\mathbf{Y} \mid \mathbf{X})$; $\mathcal{L}_2$, la de les consultes intervencionals com $P(\mathbf{Y} \mid \text{do}(\mathbf{X} = \mathbf{x}), \mathbf{Z})$; i $\mathcal{L}_3$, la de les consultes contrafactuals, on poden aparèixer simultàniament variables endògenes i resultats potencials. Es pot demostrar que el conjunt de models on una capa superior col·lapsa a una capa inferior té mesura de Lebesgue zero [Bar+20]. Per tant, en general, la informació d'una capa no determina la informació de les capes superiors sense hipòtesis causals addicionals.

ASSOCIACIÓ

Aquest nivell es caracteritza per l'observació passiva de les dades i es pot resumir amb la pregunta *què succeeix si observo...?* Les consultes en aquest nivell es basen en les *associacions* que trobem en les dades. Aquestes qüestions es poden resoldre mitjançant l'anàlisi de dades i una mica d'estadística tradicional.

Els programes d'aprenentatge automàtic clàssics actuen gairebé exclusivament en un mode associatiu [PM18], ja que estan alimentats per un flux d'observacions al qual intenten ajustar-hi una funció. Les xarxes neuronals profundes clàssiques afegeixen moltes més capes a la complexitat de la funció ajustada, però continuen basant tot el procés d'ajustament en les dades en brut. Actualment, però, hi ha noves disciplines com *Causal Reinforcement Learning* [BZL24] o, més generalment, *Causal AI* [Bar25], on es treballa en incorporar eines causals als models d'intel·ligència artificial.

La recerca pot ser merament passiva, catalogant els fets, o pot adoptar un rol actiu, provocant modificacions com a part de l'estudi. És el que estudiem en el següent nivell de la jerarquia.

INTERVENCIÓ

Estem constantment interactuant amb el món, obrint, tancant i investigant per veure què fa que això o allò passi. Limitar-se a observar i anotar quins fenòmens segueixen a altres és una cosa; però, si volem revelar relacions causals, el pas evident és *intervenir*: introduir un canvi i comprovar-ne l'efecte. De fet, no podem respondre preguntes sobre intervencions amb dades recollides passivament —és a dir, a partir de dades observacionals—, independentment de la mida del conjunt de dades que estiguem usant [PM18].

Naturalment, hi ha una limitació en el mètode experimental. Hi ha alguns factors que no podem controlar però que, al seu torn, poden ser-ne causes. Per exemple, la Lluna, amb la seva gravetat, provoca les marees als oceans i afecta línies costaneres senceres. És un cas tan clar de causalitat com qualsevol altre. I, tanmateix, som completament incapaços d'intervenir físicament per alterar la massa de la Lluna o canviar la seva distància de la Terra.

Les aproximacions intervencionistes defensen, per tant, una noció ampla d'intervenció que no es limiti als canvis que els humans podem generar. Al capdavant, una intervenció pot ser qualsevol modificació en la causa, feta deliberadament per nosaltres o no. Així, si la Lluna es desplaça una mica més lluny, el

intervencions perquè es compleixi: $P(\mathbf{Y} \mid \text{do}(\mathbf{x})) > P(\mathbf{Y})$.

rang de les mareas seria menor; aquesta és una intervenció sobre la Lluna, completament fora del nostre abast. I atès que la variació en l'hora de la marea no altera en absolut la posició de la Lluna, queda palesa l'asimetria pròpia de la causalitat.

Finalment, esmentar que és possible fer prediccions fiables sobre els efectes d'intervencions sense necessitat de dur a terme un experiment. Quan disposem d'un model causal prou robust i precís, podem emprar dades observades —primer nivell— per resoldre preguntes sobre intervencions —segon nivell—. Sense aquest model, és impossible passar de nivell a la jerarquia i, per aquesta mateixa raó, els sistemes que només treballen amb dades de primer nivell i no incorporen cap model causal mai podran respondre preguntes sobre el que succeeix si intercedim activament. Com diu [PM18], hem d'ensenyar als models a trencar les regles de l'entorn en què van ser entrenats.

CONTRAFACTUALS

Suposem que una casa s'està cremant i ens preguntem si la causa és que algú ha deixat una espelma encesa prop d'unes cortines. Què ens autoritza a dir que aquesta espelma encesa és la causa de l'incendi? Una idea plausible és dir simplement que, si no s'hagués deixat l'espelma encesa, l'incendi no hauria tingut lloc. Per què ens sembla plausible aquesta afirmació? Perquè altres aspectes circumstancials —com el color de les parets o el dia concret de l'incident— podrien haver estat diferents, però no haurien fet que el resultat s'hagués vist alterat [MA13]. L'únic fet que realment ha *marcat la diferència* entre que la casa es cremi o no és, sens dubte, l'espelma encesa. Aquest tipus de raonaments, doncs, ens ajuden a explorar relacions de dependència: diem que *A* depèn contrafactualment de *B* si, si *B* no hagués passat, *A* tampoc.

Ja hem fet menció d'aquesta idea més amunt. Introduïda per David Lewis al segle XX [Lew73; Lew81] és un principi extraordinàriament potent i es fa servir habitualment en el dret i en la medicina com a prova de causalitat. Per exemple, si es presenta una demanda contra una empresa per danys i perjudicis, caldria demostrar que l'objecte de la demanda va marcar una diferència. Si es pateix una malaltia, caldria provar que no s'hauria patit si no fos per les accions o negligències de l'acusat. De manera similar, si algú és processat per haver causat un accident, cal demostrar que no hauria tingut lloc si no fos per les accions o omissions del denunciat.

Els contrafactuals tenen una relació especialment complicada amb les dades, perquè aquestes són, per definició, fets [PM18]. Les dades no poden indicar-nos què passaria en un món contrafactual o imaginari en què alguns fets observats quedin directament negats. I, tanmateix, la ment humana realitza de manera repetida —i, en la majoria de casos, fiable— aquest tipus d'inferències.

Però si el lector creu que la ciència és incapaç de respondre a la pregunta *què hauria passat* en situacions imaginàries o contrafactuals, hauria comès un error equiparable al que van fer Galton, Pearson o Fisher en el seu moment, perquè aquesta mena d'enunciats ha estat, des dels seus orígens, part integrant del discurs científic. I, com ja hem comentat, serà part integrant també dels dos marcs teòrics que exposarem en aquest treball: el model de resultats potencials (Neyman-Rubin, cf. [Sp190; Rub74]; [IR15]; [HR20, Cap. 1]) i els models causals estructurals (Pearl, cf. [Pea88]; [Pea09]; [PGJ16]).

Fins i tot les lleis de la física poden interpretar-se com afirmacions sobre escenaris hipotètics, és a dir,

com afirmacions contrafactuals: la tercera llei de Kepler ens diu què passaria si la distància orbital d'un planeta fos diferent. I aquesta predicció no es basa només en quatre càlculs matemàtics, sinó que ha estat confirmada per l'observació acurada del moviment dels planetes durant segles.

Tot i ser una eina molt efectiva per pensar la causalitat, no tot són flors i violes. Primer, perquè no totes les dependències contrafactuals impliquen causalitat. Per exemple, si aquest mes fos juny, el mes vinent seria juliol; tot i que hi ha dependència, no hi ha una relació causal. Perquè una dependència sigui causal, ha de relacionar esdeveniments reals, no només relacions lògiques o matemàtiques. A més, la teoria s'encalla davant fenòmens com la *sobredeterminació*, on múltiples causes podrien haver provocat el mateix efecte: si un capità ordena a dos oficials obrir foc alhora contra el mateix individu, qualsevol dels dos trets hauria estat letal. Per tant, no podem afirmar contrafactualment que la víctima hauria sobreviscut si una d'elles no hagués disparat. Pearl comenta detalladament aquest exemple a [PM18, p. 40] i dona una visió des de cadascun dels graons de la jerarquia causal.

Cal tornar a fer èmfasi en què la jerarquia de Pearl estableix que per respondre preguntes causals no n'hi ha prou amb les dades per si soles: per ascendir al nivell d'intervenció cal afegir suposicions causals explícites. I per anar més enllà, cap al nivell contrafactual, es necessita encara més detall sobre els mecanismes del nostre model. Cada nivell requereix una comprensió més profunda del model i no podem passar d'un nivell a l'altre indistintament.

ESTRUCTURA DEL TREBALL I OBJECTIUS

En aquest punt, convé fer un breu parèntesi per descriure alguns aspectes d'estil adoptats en la present memòria. L'organització dels diferents enunciats matemàtics segueix un sistema de *sorting-by-color*, amb l'objectiu de facilitar-ne la identificació i consulta. Així, cada tipus d'enunciat s'associa a un format i un color específic, fet que permet localitzar ràpidament la informació desitjada.

1. Teoremes, proposicions, lemes, corol·laris, propietats i algorismes presenten una capçalera en gris fosc, i el text de l'enunciat apareix en cursiva:

Teorema. *Compte! L'enunciat del teorema també serà en cursiva! Jove xef, porti whisky amb quinze glaçons d'hidrogen, coi!*

2. Les definicions i notacions utilitzen una capçalera en gris clar i el text *no* apareix en cursiva:

Definició. Aqueix betzol, Jan, comprava whisky de figa.

3. Les remarques, exemples i notes no disposaran d'una capçalera diferenciada:

Observació. Zel de grum: quetxup, whisky, cafè, bon vi; ja!

Per mantenir el ritme visual, **totes les demostracions s'han recollit als annexos**. Val a dir que no es demostraran tots els resultats, però sí els més rellevants; també hem omès les demostracions més llargues. En tals casos, sempre s'indicarà a l'enunciat la font on es pot trobar la demostració.

L'objectiu general del treball és presentar una introducció teòrica a la inferència causal moderna que expliqui com es formulen, s'identifiquen i s'interpreten efectes causals sota hipòtesis explícites.

L'estructura serà la següent:

Models causals estructurals → identificació → Neyman–Rubin → disseny → aplicació i anàlisi

Anirem construint progressivament un llenguatge per respondre preguntes causals: primer, caldrà representar les dependències entre variables mitjançant grafs i models causals estructurals; després, preguntar-se quan una intervenció, escrita formalment amb $\text{do}()$, pot ser identificada a partir de dades observacionals; i, finalment, caldrà connectar aquest llenguatge amb el dels resultats potencials de Neyman–Rubin, on els efectes causals es formulen com comparacions entre mons possibles. D'aquesta manera, el treball es proposa els objectius següents:

1. *Com es representa una hipòtesi causal?* Formalitzar les xarxes bayesianes, la d -separació i els models causals estructurals com un llenguatge matemàtic per representar hipòtesis causals.
2. *Quan pot una pregunta causal ser resposta?* És a dir, estudiar el problema d'identificació: quan una consulta causal, formulada amb $\text{do}()$, es pot escriure només amb quantitats observacionals, mitjançant criteris com *back-door*, *front-door*, el *do-calculus* i els algorismes ID i IDC.
3. Presentar el model de resultats potencials de Neyman–Rubin, el problema fonamental de la inferència causal i els estimands principals que s'obtenen comparant resultats potencials.
4. *Com es relacionen les dues escoles de la inferència causal?* Connectar els llenguatges de Pearl i Neyman–Rubin, mostrant que no són dues teories inconnexes sinó dues maneres diferents de formular i respondre les mateixes preguntes causals.
5. Distingir què aporta el disseny experimental i quines hipòtesis cal defensar en estudis observacionals.
6. Conèixer en què consisteix la paradoxa de Simpson i estudiar el cas de les admissions de Berkeley per mostrar que una conclusió causal depèn tant de la pregunta que es formula com del model causal assumit.

Per tant, l'abast del treball és deliberadament teòric: no es busca oferir un catàleg exhaustiu de mètodes aplicats tampoc, però sí construir el marc conceptual necessari per entendre què vol dir identificar un efecte causal i per què la connexió entre Pearl i Neyman–Rubin és el punt d'arribada més natural.

Com el lector haurà apreciat, no s'ha formalitzat cap dels conceptes vistos al llarg d'aquesta introducció, que pretenia ser una lectura lleugera *per a tots els públics*. A partir del següent capítol, però, es començarà a construir el llenguatge formal de la inferència causal, i a partir d'aquí es mantindrà un to més rigorós i tècnic.

Models causals estructurals

Feliç aquell que ha pogut saber les causes de les coses.

VIRGILI, *Geòrgiques* (37 aC)

Notació 1.1. Sigui, quan no indiquem el contrari, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un *graf dirigit acíclic* (DAG), de manera que $\mathcal{V} = \{V_i\}_{i \in I}$ és el conjunt de variables o vèrtexs, $I = \{1, \dots, n\}$ és el conjunt d'índexs i $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ el conjunt d'arestes del graf. Suposarem el vector aleatori $\mathbf{V} = (V_i)_{i \in I}$. Per raons de formalitat:

1. Per a cada $i \in I$, sigui $(\mathcal{X}_i, \mathcal{B}(\mathcal{X}_i))$ un espai de Borel estàndard. Definim l'espai producte $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$, de manera que $\mathcal{B}(\mathcal{X}) = \bigotimes_{i=1}^n \mathcal{B}(\mathcal{X}_i)$ és la σ -àlgebra producte sobre $\mathcal{X}$.
2. Sigui $(\Omega, \mathcal{F}, \mathbb{P})$ un espai de probabilitat, on Ω és l'espai mostral i $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ és una mesura de probabilitat sobre $(\Omega, \mathcal{F})$ [San99, Sec. 1.2].
3. Les variables aleatòries són aplicacions mesurables $V_i : \Omega \rightarrow (\mathcal{X}_i, \mathcal{B}(\mathcal{X}_i))$; per a $\omega \in \Omega$ i $i \in I$, escriurem $V_i(\omega) = v_i \in \mathcal{X}_i$. El vector aleatori $\mathbf{V} : \Omega \rightarrow (\mathcal{X}, \mathcal{B}(\mathcal{X}))$ és una aplicació mesurable.

Quan treballem amb densitats, fixarem una mesura dominant producte en el sentit de B.10: una mesura σ -finita $\mu = \bigotimes_{i \in I} \mu_i$ sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ tal que la llei conjunta observacional P satisfaci $P \ll \mu$ [Lau96, Sec. 3.1, p. 28]. Aquesta convenció inclou el cas discret, prenent mesures de comptatge, i el cas continu real, prenent mesures de Lebesgue³. Pel teorema de Radon–Nikodym B.11, existeix una funció mesurable $p = \frac{dP}{d\mu} : \mathcal{X} \rightarrow [0, +\infty]$, única μ -gairebé pertot, tal que per a qualsevol $A \in \mathcal{B}(\mathcal{X})$:

$$P(A) = \int_A p \, d\mu.$$

Assumirem que les variables estan indexades respectant un ordre topològic estricte $<$ sobre $\mathcal{V}$ compatible amb $\mathcal{G}$; és a dir, per a tota arista dirigida $(V_i, V_j) \in \mathcal{E}$ es compleix que $V_i < V_j$, la qual cosa implica que $i < j$. D'altra banda, la funció p rep el nom de *densitat conjunta* (*joint density*) de P . Quan algun resultat requereixi que aquesta densitat sigui estrictament positiva, ho indicarem explícitament.

Notació 1.2. Si $\mathbf{S} \subseteq \mathcal{V}$, definim $\mathcal{X}_{\mathbf{S}} := \prod_{V_i \in \mathbf{S}} \mathcal{X}_i$ i:

$$\begin{aligned} \Phi_{\mathbf{S}} : (\Omega, \mathcal{F}) &\longrightarrow (\mathcal{X}_{\mathbf{S}}, \mathcal{B}(\mathcal{X}_{\mathbf{S}})) \\ \omega &\longmapsto (V_i(\omega))_{V_i \in \mathbf{S}} \end{aligned}$$

Aquesta aplicació és mesurable, ja que les seves coordenades són variables aleatòries mesurables i $\mathcal{B}(\mathcal{X}_{\mathbf{S}})$ és la σ -àlgebra producte. Per tant, $\Phi_{\mathbf{S}}$ és el vector aleatori associat a $\mathbf{S}$, i la seva llei és la mesura imatge $P_{\mathbf{S}} := \mathbb{P} \circ \Phi_{\mathbf{S}}^{-1}$ [San99, Sec. 2.2, Def. 2.2]. Quan no hi hagi ambigüitat, escriurem $P(\mathbf{S})$ per aquesta

³ És una hipòtesi de treball per a les expressions amb densitats, no una propietat de qualsevol llei induïda per un SCM [Bon+21, Def. 2.1 i Ex. 2.4].

lleis marginal. La llei conjunta és $P := P(\mathcal{V})$, és a dir, $P(A) := \mathbb{P}(\Phi_{\mathcal{V}} \in A)$ per a $A \in \mathcal{B}(\mathcal{X})$. En canvi, $p_{\mathbf{S}}(\mathbf{s})$ o simplement $p(\mathbf{s})$ designarà la PMF o densitat de $\mathbf{S}$ avaluada en el valor $\mathbf{s} \in \mathcal{X}_{\mathbf{S}}$, i $p(v_i | \mathbf{s})$ designarà una massa o densitat condicional.

A més, escriurem $\mathbf{v}$ per a una assignació conjunta de valors de $\mathcal{V}$, és a dir, $\mathbf{v} = (v_i)_{i \in I} \in \mathcal{X}$, i entendrem $p(\mathbf{v}) = p(v_1, \dots, v_n)$. Si $\mathbf{S} \subseteq \mathcal{V}$, escriurem $\mathbf{s} := \mathbf{v}_{\mathbf{S}} = (v_i)_{i \in \mathbf{S}} \in \mathcal{X}_{\mathbf{S}}$ per a la restricció de l'assignació a $\mathbf{S}$. Més generalment, si $\mathbf{A}, \mathbf{B} \subseteq \mathcal{V}$, $\mathbf{a} := \mathbf{v}_{\mathbf{A}}$ i $\mathbf{b} := \mathbf{v}_{\mathbf{B}}$, l'expressió $\mathbf{a} \setminus \mathbf{b}$ voldrà dir $\mathbf{v}_{\mathbf{A} \setminus \mathbf{B}}$. Quan un valor individual v_i aparegui en una expressió d'aquest tipus, s'entendrà com la restricció $\mathbf{v}_{\{V_i\}}$. Reservarem $\mathbb{P}$ per a la mesura subjacent, P per a les lleis sobre espais observables i p per a les masses o densitats avaluades.

1.1

XARXES BAYESIANES

L'objectiu d'aquesta secció és formalitzar com un graf dirigit pot codificar restriccions probabilístiques. Veurem que una xarxa bayesiana és una distribució que factoritza segons un DAG i que, sota les propietats de Markov, el graf permet llegir independències condicionals.

Les xarxes bayesianes relacionen una estructura gràfica definida sobre un conjunt de variables amb una distribució de probabilitat. Al treballar només en la capa observacional $\mathcal{L}_1$, la interpretació causal requerirà hipòtesis addicionals, que introduïrem més endavant.

Definició 1.3 (Pares Markovians). Sigui $i \in I$ de manera que $\text{Pred}_i = \{V_j \in \mathcal{V} \mid j < i\}$ és el conjunt de predecessors d' V_i ⁴. Anomenem al conjunt de variables PM_i *pares Markovians* d' V_i si $\text{PM}_i \subset \text{Pred}_i$ és el conjunt minimal⁵ de Pred_i tal que:

$$p(v_i | \text{pm}_i) = p(v_i | v_1, \dots, v_{i-1}),$$

on hem fet un lleuger abús de notació: $p_{v_1, \dots, v_{i-1}}(v_i | v_1, \dots, v_{i-1}) = p(v_i | v_1, \dots, v_{i-1})$. En el cas discret, aquesta massa condicional coincideix amb $P(V_i = v_i | V_1 = v_1, \dots, V_{i-1} = v_{i-1})$; en el cas continu, p designa una densitat condicional.

Aquesta és la que es coneix com a densitat o massa condicional de V_i donats $V_1, \dots, V_{i-1}$, i no s'ha de confondre amb la *distribució marginal* d' V_i , que podem denotar per $P(V_i)$.

Propietat 1.4 (Regla de la cadena per a xarxes bayesianes). Diem que si P admet una factorització de l'estil:

$$\prod_{i=1}^n p(v_i | \text{pa}_i^{\mathcal{G}}) = \prod_{i=1}^n p(v_i | v_1, \dots, v_{i-1}) = p(v_1, \dots, v_n) =: p(\mathbf{v})$$

sobre $\mathcal{G}$, aleshores P i $\mathcal{G}$ són Markov-compatibles, o que P factoritza sobre $\mathcal{G}$. Els factors $p(v_i | \text{pa}_i^{\mathcal{G}})$ corresponen als conditional probability distributions (CPDs) de [KF09, Sec. 3.2.3, Defs. 3.4–3.5] avaluats; en

⁴ Cal no confondre predecessors amb no-descendents, ja que Pred_i són els predecessors en l'ordre. Per exemple, en un graf $V_1 \rightarrow V_3 \leftarrow V_2$, V_2 no és predecessor de V_1 ni descendent de V_1 .

⁵ Per definició, no hi haurà un subconjunt propi de PM_i tal que se satisfaci l'equació.

aquest treball els escrivim com a PMFs o densitats condicionals segons el cas.

Observació 1.5. És important distingir els *pares gràfics* dels *pares Markovians* d'una variable V_i en un DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Els primers només depenen de $\mathcal{G}$; els segons depenen de P i d'un ordre topològic $<$.

1. Els *pares* (o *pares gràfics*) de V_i en un graf $\mathcal{G}$ són les variables que tenen una aresta dirigida cap a V_i . Si V_j és un pare de V_i , existeix una aresta $V_j \rightarrow V_i$ en el graf. D'aquí en endavant, $\text{Pa}_{V_i}^{\mathcal{G}} = \{V_j \mid V_j \rightarrow V_i \in \mathcal{E}\}$, que normalment denotarem per $\text{Pa}_i^{\mathcal{G}}$.
2. Donat un ordre topològic $<$ sobre $\mathcal{V}$ i una distribució P , els *pares Markovians* de $V_i \in \mathcal{V}$, PM_i , són el conjunt minimal que fan que V_i sigui condicionalment independent dels altres predecessors. Per això poden existir pares gràfics redundants des del punt de vista probabilístic. A [Pea09, Sec. 1.2.2] es demostra que PM_i és únic, donat l'ordre, sempre que P sigui estrictament positiva.

Teorema 1.6 (Ordered Markov conditioning [Pea88]; [Pea09, Sec. 1.2.3, Th. 1.2.6]). P factoritza sobre $\mathcal{G}$ si, i només si, per a tot $i \in I$ es compleix que V_i és condicionalment independent de $\text{Pred}_i \setminus \text{Pa}_i^{\mathcal{G}}$ donats els seus pares $\text{Pa}_i^{\mathcal{G}}$ respecte d'un cert ordre topològic $<$ sobre les variables, compatible amb $\mathcal{G}$.

El teorema anterior dona un criteri per determinar si P factoritza sobre $\mathcal{G}$, respecte de l'ordre topològic $<$. Ara bé, la noció de factorització no depèn de l'ordre escollit: podem treballar amb els no-descendents i formular la propietat local de Markov, que és purament gràfica.

Definició 1.7 (Propietat local de Markov [Lau96, Sec. 3.2.2, p. 50]). Si per a tot $i \in I$ es compleix que V_i és condicionalment independent de $\text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}}$ condicionat a $\text{Pa}_i^{\mathcal{G}}$, és a dir:

$$(V_i \perp\!\!\!\perp_P \text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}}) \mid \text{Pa}_i^{\mathcal{G}}, \quad (\text{L})$$

on ND_i denota els no-descendents de V_i i $\text{Pa}_i^{\mathcal{G}} \subset \text{ND}_i$ els seus pares gràfics en $\mathcal{G}$, diem que P satisfà la propietat local de Markov respecte $\mathcal{G}$.

Propietat 1.8. Si P és Markov respecte $\mathcal{G}$, aleshores $\text{PM}_i \subseteq \text{Pa}_i^{\mathcal{G}}$ per a tot $i \in I$. Si es compleix 1.53, aleshores $\text{PM}_i = \text{Pa}_i^{\mathcal{G}}$ per a tot $i \in I$.

Teorema 1.9. P satisfà la propietat local de Markov si, i només si, P factoritza sobre $\mathcal{G}$. És a dir, 1.4 $\iff$ 1.7.

Definició 1.10 (Xarxa bayesiana). Una xarxa bayesiana és un parell $(\mathcal{G}, P)$ on P factoritza sobre $\mathcal{G}$.

Exemple 1.11. Considerem la xarxa bayesiana de la figura 1.1. En aquesta, la variable corresponent a l'estació de l'any (V_1) influeix sobre dues altres variables: si l'aspersor del jardí està activat (V_2) i si plou (V_3). I si el terra es troba mullat (V_4) és perquè o bé ha plogut, o bé l'aspersor estava activat. I, en conseqüència, determinen la probabilitat que el terra rellisqui (V_5). Totes les variables, excepte V_1 , poden considerar-se binàries. V_1 , en canvi, pot prendre els quatre valors corresponents a les estacions de l'any. En el cas de V_5 , conèixer l'estat de V_4 és suficient per determinar-ne el comportament: saber si el terra és mullat fa innecessari conèixer la resta de variables (V_1, V_2, V_3). Per tant, V_5 és condicionalment independent de les seves predecessores un cop conegut V_4 . La causa que hagi provocat la humitat o l'estació de l'any deixen de ser rellevants un cop es coneix aquesta informació. Parlarem de configuracions gràfiques com aquesta a la secció 1.4.

Com que $(\mathcal{G}, P)$ és una xarxa bayesiana, P factoritza sobre $\mathcal{G}$. Aplicant 1.9, es compleix la propietat local

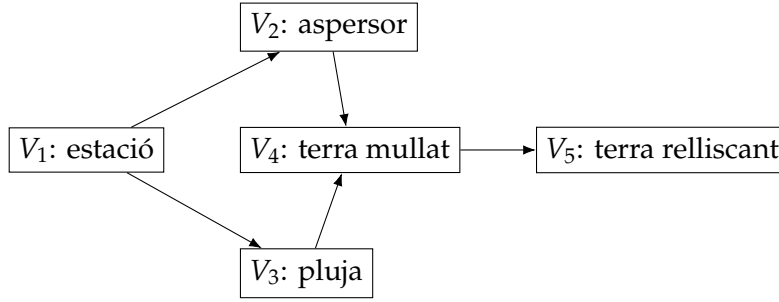


Figura 1.1: Una xarxa bayesiana de cinc variables. Font: [Pea09, Sec. 1.2.2, p. 15, figura 1.2].

de Markov 1.7. Aleshores, podem expressar la seva PMF o densitat per a aquest cas particular com:

$$p(v_1, \dots, v_5) = p(v_1) \cdot p(v_2 | v_1) \cdot p(v_3 | v_1) \cdot p(v_4 | v_2, v_3) \cdot p(v_5 | v_4).$$

1.2

d-SEPARACIÓ, d-CONNEXIÓ I I-MAPES

La d-separació és el criteri gràfic que tradueix, sota la propietat local de Markov, camins blocats en independències condicionals. Entendrem la relació entre d-separació, factorització, I-mapes i equivalència de Markov.

Volem saber quan una independència $X \perp\!\!\!\perp_P Y \mid Z$ es verifica en una distribució associada a $\mathcal{G}$.

Definició 1.12 (Camí i camí dirigit). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un DAG. Un *camí* entre V_{i_1} i V_{i_m} és una seqüència de variables diferents $V_{i_1}, \dots, V_{i_m}$ tal que, per a tot $\ell = 1, \dots, m-1$, existeix una aresta entre V_{i_ℓ} i $V_{i_{\ell+1}}$, independentment de la seva orientació. Un *camí dirigit* de V_{i_1} a V_{i_m} és un camí en què totes les arestes apunten en el sentit del recorregut, és a dir, $V_{i_\ell} \rightarrow V_{i_{\ell+1}}$ per a tot $\ell = 1, \dots, m-1$.

Nota 1.13. Posem, per ara, $\mathbf{X} = \{V_i\}_{i \in A}$, $\mathbf{Y} = \{V_j\}_{j \in B}$ i $\mathbf{Z} = \{V_k\}_{k \in C}$, on $A, B, C \subseteq I$ són disjunts dos a dos. En d-separació, els camins entre $\mathbf{X}$ i $\mathbf{Y}$ s'entenen a partir de camins no dirigits (es poden recórrer ignorant l'orientació de les arestes). Estudiarem les diferents configuracions gràfiques a la secció 1.4.

Per a verificar si $\mathbf{X}$ és independent de $\mathbf{Y}$ condicionat a $\mathbf{Z}$ en qualsevol distribució que factoritzi sobre $\mathcal{G}$, cal comprovar si les variables de $\mathbf{Z}$ *bloquegen* tots els camins que van dels nodes de $\mathbf{X}$ als nodes de $\mathbf{Y}$.

Definició 1.14 (Camí blocat [PJS17, Sec. 6.2, Def. 6.1]). Diem que un camí entre dues variables V_{i_1}, V_{i_m} està *blocat per* $\mathbf{Z} \subseteq \mathcal{V}$ si existeix una variable intermèdia V_{i_k} ($1 < k < m$) tal que es compleix una de les dues condicions:

1. $V_{i_k} \in \mathbf{Z}$ i tenim o bé una cadena (cf. 1.58) o bé un *fork* (cf. 1.60):

$$(V_{i_{k-1}} \rightarrow V_{i_k} \rightarrow V_{i_{k+1}}) \quad \text{o} \quad (V_{i_{k-1}} \leftarrow V_{i_k} \leftarrow V_{i_{k+1}}) \quad \text{o} \quad (V_{i_{k-1}} \leftarrow V_{i_k} \rightarrow V_{i_{k+1}}).$$

2. V_{i_k} és un *collider* (cf. 1.61) en el camí i ni ell ni cap descendent seu pertany a $\mathbf{Z}$:

$$(\{V_{i_k}\} \cup \text{DE}_{i_k}) \cap \mathbf{Z} = \emptyset \quad \text{i} \quad V_{i_{k-1}} \rightarrow V_{i_k} \leftarrow V_{i_{k+1}},$$

on DE_{i_k} correspon al conjunt de descendents de V_{i_k} .

Definició 1.15 (d -connexió). Diem que $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{V}$ (amb $\mathbf{X} \cap \mathbf{Y} = \emptyset$) estan d -connectats respecte de $\mathbf{Z}$ (amb $\mathbf{X} \cap \mathbf{Z} = \mathbf{Y} \cap \mathbf{Z} = \emptyset$) si *existeix* un camí entre alguna variable de $\mathbf{X}$ i alguna variable de $\mathbf{Y}$ que **no** està blocat per $\mathbf{Z}$.

Definició 1.16 (d -separació [Pea88]; [Pea09, Sec. 1.2.3, Def. 1.2.3]). Diem que $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{V}$ (amb $\mathbf{X} \cap \mathbf{Y} = \emptyset$) estan d -separats per $\mathbf{Z}$ (amb $\mathbf{X} \cap \mathbf{Z} = \mathbf{Y} \cap \mathbf{Z} = \emptyset$) si *cada* camí entre les variables de $\mathbf{X}$ i les variables de $\mathbf{Y}$ **sí** està blocat per $\mathbf{Z}$. Ho identificarem per $\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z}$.

Un cop definit el criteri de d -separació, falta precisar quina hipòtesi permet passar d'una separació gràfica a una independència probabilística. Ja hem definit la propietat local de Markov a 1.7; ara introduïm la versió global.

Propietat 1.17 (Global de Markov). Diem que P satisfà la propietat global de Markov respecte el graf $\mathcal{G}$ si per a qualssevol $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts dos a dos tenim que:

$$\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z} \implies \mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z}. \quad (\text{G})$$

Per a DAGs, la propietat global, la propietat local i la factorització de Markov 1.4 són equivalents. A 1.9 ja hem establert que 1.4 $\iff$ 1.7. I ara tenim el següent:

Teorema 1.18 ([PJS17, Sec. 6.5, Th. 6.22] i [Lau96, Sec. 3.2.2, Th. 3.27]). 1.4 $\iff$ 1.7 $\iff$ 1.17.

Podem, doncs, preguntar-nos fins a quin punt aquestes independències imposades pel graf reflecteixen les que realment existeixen en P .

Definició 1.19 (Mapa d'independències condicionals de P [KF09, Def. 3.2]). Definim $I(P)$ com el conjunt d'independències condicionals que es compleixen a P , és a dir, en termes de $\mathcal{V}$; anàlogament, $I(\mathcal{G})$ és el conjunt de relacions de d -separació que es compleixen en $\mathcal{G}$:

$$\begin{aligned} I(P) &= \{(\mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z}) \mid \mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}, \mathbf{X} \cap \mathbf{Y} = \mathbf{X} \cap \mathbf{Z} = \mathbf{Y} \cap \mathbf{Z} = \emptyset\}, \\ I(\mathcal{G}) &= \{(\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z}) \mid \mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}, \mathbf{X} \cap \mathbf{Y} = \mathbf{X} \cap \mathbf{Z} = \mathbf{Y} \cap \mathbf{Z} = \emptyset\}. \end{aligned}$$

És a dir, que $\mathbf{X}, \mathbf{Y}$ i $\mathbf{Z}$ són disjunts dos a dos en ambdós casos.

Definició 1.20 (I -mapa [KF09, Def. 3.3]). Diem que $\mathcal{G}$ és un I -mapa de P si $I(\mathcal{G}) \subseteq I(P)$; és a dir, que si $\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z} \implies \mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z}$. Anàlogament, si $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ són disjunts dos a dos:

$$I(\mathcal{G}) \subseteq I(P) \iff (\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z} \implies \mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z}).$$

El resultat següent formalitza la correspondència bàsica entre ser un I -mapa d'una distribució i la factorització de la distribució conjunta segons aquest graf.

Teorema 1.21 ([KF09, Ths. 3.1, 3.2, 3.3]). Si $\mathcal{G}$ és un I -mapa de P , aleshores P factoritza sobre $\mathcal{G}$. Recíprocament, si P factoritza sobre $\mathcal{G}$, aleshores $\mathcal{G}$ és un I -mapa de P .

Nota 1.22. Perquè $\mathcal{G}$ sigui un $\mathcal{I}$ -mapa de P , cal que qualsevol independència de $\mathcal{G}$ es compleixi també en P . En canvi, P pot contenir independències addicionals que no quedin reflectides en $\mathcal{G}$.

Observació 1.23. Per a gairebé totes (*almost everywhere*, cf. B.17) les parametritzacions CPD compatibles amb $\mathcal{G}$, la d -separació caracteritza exactament les independències que es verifiquen a P [Pea09, Sec. 1.2.3]; [KF09, Sec. 3.3.2, Th. 3.5]. Cal tenir present, però, que per a qualsevol independència que no sigui una conseqüència de la d -separació a $\mathcal{G}$, sempre podem trobar una distribució P que factoritzi sobre $\mathcal{G}$ i que sigui un contraexemple.

Teorema 1.24 ([KF09, Sec. 3.3.2, Th. 3.5]). *Fixat un DAG $\mathcal{G}$, per a gairebé totes (cf. B.17) les distribucions P que factoritzen sobre $\mathcal{G}$, és a dir, totes excepte un conjunt de mesura zero dins l'espai de parametritzacions CPD compatibles amb $\mathcal{G}$, tenim $\mathcal{I}(\mathcal{G}) = \mathcal{I}(P)$.*

Definició 1.25 (Cobertura de Markov o *Markov blanket*, [Pea88]; [PJS17, Sec. 6.5, Def. 6.26]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ i $Y \in \mathcal{V}$ tal que P compleix la propietat local de Markov. La *cobertura de Markov* de Y és el conjunt més petit $\mathbf{M} \subseteq \mathcal{V}$ tal que

$$Y \perp_{\mathcal{G}} \mathcal{V} \setminus (\{Y\} \cup \mathbf{M}) \mid \mathbf{M} \stackrel{1.17}{\implies} Y \perp_P \mathcal{V} \setminus (\{Y\} \cup \mathbf{M}) \mid \mathbf{M}.$$

En altres paraules, condicionat a $\mathbf{M}$, la resta de variables no aporten cap informació nova sobre Y .

Corol·lari 1.26. *La cobertura de Markov $\mathbf{M}$ és el conjunt que d -separa Y de la resta de nodes.*

Proposició 1.27 (*Markov blanket*, [PJS17, Sec. 6.5, Prop. 6.27]). *La cobertura de Markov $\mathbf{M}$ de Y inclou els seus pares $\text{Pa}_Y^{\mathcal{G}}$, els seus fills $\text{Ch}_Y^{\mathcal{G}}$ i els pares dels seus fills $\text{Pa}_{\text{Ch}_Y^{\mathcal{G}}}^{\mathcal{G}}$:*

$$\mathbf{M} = \text{Pa}_Y^{\mathcal{G}} \cup \text{Ch}_Y^{\mathcal{G}} \cup \text{Pa}_{\text{Ch}_Y^{\mathcal{G}}}^{\mathcal{G}} \setminus \{Y\}.$$

Definició 1.28 ($\mathcal{I}$ -equivalència [KF09, Sec. 3.3.4, Def. 3.9]). Dues estructures gràfiques $\mathcal{K}_1$ i $\mathcal{K}_2$ sobre $\mathcal{V}$ són $\mathcal{I}$ -equivalents si $\mathcal{I}(\mathcal{K}_1) = \mathcal{I}(\mathcal{K}_2)$. El conjunt de tots els grafs sobre $\mathcal{V}$ forma una partició de classes d' $\mathcal{I}$ -equivalència, que són el conjunt de classes d'equivalència induïda per la relació d' $\mathcal{I}$ -equivalència.

Si dos grafs codifiquen les mateixes independències, també determinen la mateixa família de distribucions compatibles, i viceversa.

Definició 1.29 (Equivalència de Markov [PJS17, Sec. 6.5]). Denotem per $\mathcal{P}(\mathcal{G})$ el conjunt de distribucions conjuntes que factoritzen sobre $\mathcal{G}$, és a dir, $\mathcal{P}(\mathcal{G}) := \{P \mid P \text{ factoritza sobre } \mathcal{G}\}$. Diem que $\mathcal{G}_1$ i $\mathcal{G}_2$ són *Markov-equivalents* si $\mathcal{P}(\mathcal{G}_1) = \mathcal{P}(\mathcal{G}_2)$.

Nota 1.30. Així, $(X, \mathcal{B}(X), \mathcal{P}(\mathcal{G}))$ és un model estadístic. No s'ha de confondre $\mathcal{P}(\mathcal{G})$ amb la família intervencional $\mathcal{P}_*$ del capítol 2. Reservarem la notació $(\Omega, \mathcal{F}, \mathbb{P})$ de 1.1 per a l'espai de probabilitat i escriurem $(X, \mathcal{A}, \mathcal{P})$ per a un model estadístic (cf. B.70) sobre un espai d'observacions.

Nota 1.31 (Classes d'equivalència de Markov [PJS17, Sec. 6.5]). Els DAGs que són Markov-equivalents a un $\mathcal{G}$ concret formen una *classe d'equivalència de Markov de $\mathcal{G}$* . Es pot representar per un PDAG complet, $\text{CPDAG}(\mathcal{G}) = (\mathcal{V}, \mathcal{E})$, on l'aresta dirigida $X \rightarrow Y \in \mathcal{E}$ si, i només si, tots els membres de la classe d'equivalència de Markov la contenen.

Definició 1.32 (Esquelet [KF09, Sec. 3.3.4, Def. 3.10]). L'esquelet de $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ és un graf *no dirigit* $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ tal que si $X \rightarrow Y \in \mathcal{E}$, aleshores $(X, Y) \in \mathcal{E}'$.

Diem que un patró de tres variables $X, Y, Z \in \mathcal{V}$ és una *v-estructura* si $X \rightarrow Y \in \mathcal{E}$ i $Z \rightarrow Y \in \mathcal{E}$, és a dir, $X \rightarrow Y \leftarrow Z$. També s'anomena *collider* i els veurem en detall a la secció 1.4.

Teorema 1.33 ([KF09, Sec. 3.3.4, Th. 3.7]). Si dos grafs $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E})$ i $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}')$ tenen el mateix esquelet i les mateixes *v-estructures*, aleshores són *I-equivalents*.

Observació 1.34. El recíproc, però, no és cert: existeixen grafs que són *I-equivalents* però no tenen el mateix conjunt de *v-estructures*. Per exemple, tots els DAGs complets⁶ sobre el mateix conjunt de nodes són *I-equivalents*, perquè no codifiquen cap independència condicional (és a dir, $I(\mathcal{G}) = \emptyset$). En aquests grafs no hi ha immoralitats, però les *v-estructures* poden variar segons l'orientació.

Definició 1.35 (Immoralitat [KF09, Sec. 3.3.4, Def. 3.11]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un DAG. Diem que una *v-estructura* $X \rightarrow Y \leftarrow Z$ en $\mathcal{G}$ és una *immoralitat* si els progenitors de Y , X i Z , no són adjacents (és a dir, no hi ha una aresta entre X i Z).

Nota 1.36. Notem que no tota *v-estructura* és una immoralitat, de manera que dos grafs amb les mateixes immoralitats no tenen necessàriament les mateixes *v-estructures*. Per exemple, dos grafs complets sempre tenen les mateixes immoralitats (cap) però poden tenir diferents *v-estructures*.

Teorema 1.37 ([KF09, Th. 3.8]; [PJS17, Lem. 6.25]; [VP90]). Sigui $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E})$ i $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}')$. $\mathcal{G}_1$ i $\mathcal{G}_2$ tenen el mateix esquelet i conjunt d'immoralitats si, i només si, són *I-equivalents*.

La propietat global de Markov només garanteix una direcció: les separacions del graf són independències a la distribució. Si volem que el graf no amagui independències probabilístiques addicionals, com ara les intransitivitats i altres independències «accidentals», cal afegir una hipòtesi més forta. [Fog21, Ex. 3.34] dona un contraexemple molt il·lustratiu. Sota la condició de *d-fidelitat* el graf explica totes les independències condicionals que podem trobar a P (cf. [Bon+21, Def. A.9]).

Definició 1.38 (Fidelitat [PJS17, Sec. 6.5, Def. 6.33]). Diem que P és *d-fidel* (*d-faithful*, o simplement *fidel* o *faithful*) al graf $\mathcal{G}$ si per a tot $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts tenim que $\mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z} \implies \mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z}$. És a dir, $I(P) \subseteq I(\mathcal{G})$.

Observació 1.39. Assumint la propietat global de Markov (cf. 1.17 i 1.18), la *d-fidelitat* es dona si, per a tots els subconjunts disjunts $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ de $\mathcal{V}$,

$$\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{X} \perp_P \mathbf{Y} \mid \mathbf{Z}.$$

Així, una distribució P és fidel a $\mathcal{G}$ si, i només si, els models d'independència $\perp_{\mathcal{G}}$ i $\perp_P$ són isomorfs.

⁶ Un DAG complet és un DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ en què tots els parells de nodes diferents són adjacents. Per tant, no podem tenir cicles i per a cada parell de nodes $V_i, V_j \in \mathcal{V}$ amb $i \neq j$, $V_i \rightarrow V_j$ o bé $V_i \leftarrow V_j$.

MODELS CAUSALS ESTRUCTURALS (SCM)

Aquí passem de l'estructura gràfica com a representació d'independències als models causals estructurals. L'objectiu és definir els SCMs, veure com indueixen distribucions observacionals i preparar la noció d'intervenció com a substitució d'equacions estructurals.

Hi ha diverses maneres de formalitzar els models causals estructurals. [PV20; PV22]; [Par23, Sec. 1.2.1], per exemple, separen explícitament les variables observades, els sorolls exògens i els factors de confusió latents. En aquest treball, definirem $(\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ per obtenir un SCM. Aquesta convenció no és universal; per exemple, [Pea09, Sec. 7.1, Def. 7.1.1] no inclou $P(\mathcal{U})$ dins la definició d'SCM.

Definició 1.40 (Model causal estructural (SCM) [Bar+22, Def. 27.1]; [Pea09, Sec. 1.4.1]). Definim un *model causal estructural* (SCM, *Structural Causal Model*) com $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$, on:

- $\mathcal{U}$ és el conjunt de variables aleatòries *exògenes*. Són variables determinades per factors externs al model. Per a cada $U \in \mathcal{U}$, fixem un espai de Borel estàndard $(\mathcal{U}_U, \mathcal{B}(\mathcal{U}_U))$ de valors possibles.
- $\mathcal{V} = \{V_i\}_{i \in I}$ és el conjunt de variables *endògenes*, determinades per variables en $\mathcal{U} \cup \mathcal{V}$. El valor de les variables de $\mathcal{V}$ queda determinat per les equacions estructurals [Bon+21, Sec. 2]. A la pràctica, però, acostumen a correspondre a les variables que podem observar i mesurar.
- Per a cada $V_i \in \mathcal{V}$, denotem per $\mathcal{U}_i^c \subseteq \mathcal{U}$ el conjunt de factors exògens associats a V_i i per $\text{Pa}_i^{\mathcal{M}} \subseteq \mathcal{V} \setminus \{V_i\}$ el conjunt de variables endògenes que apareixen com a arguments de $f_i \in \mathcal{F}$.
- $\mathcal{F} = \{f_1, \dots, f_n\}$ és un conjunt de *funcions estructurals* tals que, per a cada variable endògena $V_i \in \mathcal{V}$, $i \in I$, existeix una aplicació mesurable:

$$\begin{aligned} f_i : (\mathcal{X}_{\text{Pa}_i^{\mathcal{M}}}, \mathcal{B}(\mathcal{X}_{\text{Pa}_i^{\mathcal{M}}})) \times (\mathcal{U}_i^c, \mathcal{B}(\mathcal{U}_i^c)) &\longrightarrow (\mathcal{X}_i, \mathcal{B}(\mathcal{X}_i)) \\ (\text{pa}_i^{\mathcal{M}}, \mathbf{u}_i^c) &\longmapsto v_i = f_i(\text{pa}_i^{\mathcal{M}}, \mathbf{u}_i^c) \end{aligned}$$

on $\mathcal{X}_{\text{Pa}_i^{\mathcal{M}}} = \prod_{V_j \in \text{Pa}_i^{\mathcal{M}}} \mathcal{X}_j$ i $\mathcal{U}_i^c = \prod_{U \in \mathcal{U}_i^c} \mathcal{U}_U$ és l'espai mesurable producte de valors d' $\mathcal{U}_i^c$. El subconjunt $\text{Pa}_i^{\mathcal{M}}$ rep el nom de *pares estructurals* de V_i dins de $\mathcal{M}$ i el distingim del conjunt de *pares gràfics* $\text{Pa}_i^{\mathcal{G}}$.

- $\mathbb{P}(\mathcal{U})$ és una mesura de probabilitat conjunta sobre l'espai mesurable $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$, on $\mathcal{U} = \prod_{U \in \mathcal{U}} \mathcal{U}_U$. Prenent un vector aleatori $\mathbf{U} : (\Omega, \mathcal{F}) \longrightarrow (\mathcal{U}, \mathcal{B}(\mathcal{U}))$, la llei que usarem serà $P(\mathcal{U}) = \mathbb{P}(\mathcal{U}) \circ \mathbf{U}^{-1}$.

Al llarg del treball considerarem SCMs acíclics: les dependències dirigides entre variables endògenes admeten un ordre topològic $<$.

$$\mathcal{E}_{\mathcal{M}}^{\rightarrow} = \bigcup_{i=1}^n \{X \rightarrow V_i \mid X \in \text{Pa}_i^{\mathcal{M}}\}. \quad (1.1)$$

Anomenarem $\mathcal{G}_{\mathcal{M}}^{\rightarrow} = (\mathcal{V}, \mathcal{E}_{\mathcal{M}}^{\rightarrow})$ el *graf estructural endogen* de $\mathcal{M}$. Aquest graf recull només les dependències dirigides entre variables endògenes especificades per les funcions estructurals. Veurem més endavant que coincideix amb el doagrama causal en el cas Markovià (cf. 1.42), però no és el cas en general.

Observació 1.41.

1. El conjunt complet d'equacions $\mathcal{F}$ forma una aplicació mesurable global de l'espai dels exògens cap a l'espai de les endògenes.

$$\begin{aligned}\Psi_{\mathcal{M}} : (\mathcal{U}, \mathcal{B}(\mathcal{U})) &\longrightarrow (\mathcal{X}, \mathcal{B}(\mathcal{X})) \\ \mathbf{u} &\longmapsto \Psi_{\mathcal{M}}(\mathbf{u}) := (\psi_1(\mathbf{u}), \dots, \psi_n(\mathbf{u})).\end{aligned}$$

Aquí $\Psi_{\mathcal{M}}(\mathbf{u})$ és l'única assignació conjunta $\mathbf{v} = (v_1, \dots, v_n) \in \mathcal{X}$ obtinguda resolent recursivament les equacions estructurals de $\mathcal{M}$ segons un ordre topològic.

2. Quan necessitem mostrar també les variables exògenes, escriurem $\mathcal{G}_{\mathcal{M}}^+$ per al graf augmentat sobre $\mathcal{U} \cup \mathcal{V}$, amb una aresta $X \rightarrow V_i$ per a cada argument $X \in \text{Pa}_i^{\mathcal{M}} \cup \mathcal{U}_i^c$ de f_i . Aquesta convenció correspon al graf augmentat $G^a(\mathcal{M})$ de [Bon+21, Def. 2.7]. En canvi, $\mathcal{G}_{\mathcal{M}}$ queda reservat al diagrama causal sobre $\mathcal{V}$ (els introduïrem al capítol 2).
3. Les equacions estructurals s'han d'entendre com assignacions que especifiquen com uns factors determinen altres variables. En aquest sentit, la literatura clàssica [Pea88]; [Pea09, Sec. 1.4.1]; [Nea20]; [KF09, Sec. 3.2] prefereix utilitzar el terme *equacions estructurals*, mentre que autors com [PJS17, Secs. 2.2 i 6.2, Def. 6.2] opten per parlar d'*assignacions estructurals*.

Definició 1.42 (SCM Markovià [Bar+22, Sec. 27.2]). Un SCM acíclic $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ és *Markovià* si, per a tot $i \neq j$, els conjunts de factors exògens $\mathcal{U}_i^c$ i $\mathcal{U}_j^c$ són independents respecte a $P(\mathcal{U})$ ⁷. En particular, per a cap variable $V_i \in \mathcal{V}$ tenim que $U \in \mathcal{U}_i^c$ sigui argument de dues funcions diferents.

Definició 1.43 (SCM semi-Markovià [Bar+22, Sec. 27.2]). Un SCM acíclic $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ s'anomena *semi-Markovià* si existeixen $i, j \in I$ tals que $\mathcal{U}_i^c$ i $\mathcal{U}_j^c$ no són independents respecte a $P(\mathcal{U})$, o bé comparteixen algun factor exogen.

Nota 1.44. En la literatura d'identificació és habitual treballar amb una forma canònica dels SCMs semi-markovians en què cada variable latent comuna té exactament dues variables observables com a filles [Par23, Sec. 1.2.1]. Això no vol dir que la part inobservable del sistema hagi de ser realment tan simple: una estructura latent més complexa es pot substituir, sense pèrdua de generalitat, per un conjunt de causes comunes latents entre parelles d'observables [VP90; Ver93]. Aquesta convenció és la que més endavant representarem mitjançant arestes bidirigides entre variables endògenes.

Proposició 1.45 (Distribució induïda [PJS17, Prop. 6.3]). Tot SCM acíclic $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ induïx una única distribució de probabilitat conjunta sobre $\mathcal{V}$. Aquesta distribució $P(\mathcal{V})$ s'anomena la distribució observacional induïda per $\mathcal{M}$.

Acabem de veure que aquesta llei és la distribució induïda sobre $\mathcal{V}$ per la distribució exògena $P(\mathcal{U})$ i les funcions estructurals. De fet, $P(\mathcal{V}) = P(\mathcal{U}) \circ \Psi_{\mathcal{M}}^{-1}$, és a dir, $P(\mathcal{V})$ és la imatge de $P(\mathcal{U})$ per l'aplicació mesurable global $\Psi_{\mathcal{M}}$ que hem definit a 1.41.

Per definir el graf observacional induït només ens falta una noció local de C-component al nivell de l'SCM: dues variables endògenes pertanyen a la mateixa component si els seus mecanismes estan

⁷ Aquesta hipòtesi correspon a la suficiència causal; la reprendrem a 2.19.

connectats per una cadena de factors exògens compartits o dependents. La definició gràfica completa es fixa a 2.20.

Nota 1.46 (C-components al nivell de l'SCM [Bar+20, Def. 18]). Si $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ és un SCM, les C-components es poden definir a partir d' $\mathcal{U}$. Siguin $\mathbf{U}_1^c, \dots, \mathbf{U}_\ell^c \subset \mathcal{U}$ subconjunts disjunts i maximals de variables no observables de $\mathcal{M}$ tals que $p(\mathbf{u}) = \prod_{k=1}^\ell p(\mathbf{u}_k^c)$, en cas d'admetre una PMF o densitat. Aleshores $V_i, V_j \in \mathcal{V}$ pertanyen a la mateixa C-component quan f_i i f_j tenen arguments latents dins d'un mateix bloc $\mathbf{U}_k^c$; equivalentment, $|\{\mathbf{U}_k^c \mid \mathbf{U}_k^c \cap \mathbf{U}_i^c \neq \emptyset, \mathbf{U}_k^c \cap \mathbf{U}_j^c \neq \emptyset\}| > 0$.

Definició 1.47 (DAG observacional induït per un SCM [Bar+20, Def. 26]). Sigui $\mathcal{M}$ un SCM acíclic i fixem un ordre topològic $<$ de les seves funcions estructurals, amb una enumeració compatible. Un DAG $\mathcal{G}_M^{\mathcal{L}_1}$ és induït per $\mathcal{M}$ a la capa $\mathcal{L}_1$ si:

1. té un vèrtex per a cada variable endògena de $\mathcal{V}$;
2. conté una aresta $V_j \rightarrow V_i$ sempre que V_j apareix com a argument endogen de f_i ;
3. si V_j i V_i pertanyen a la mateixa C-component i f_j precedeix f_i en l'enumeració topològica, aleshores conté $V_j \rightarrow V_i$ i també $V_k \rightarrow V_i$ per a tot $V_k \in \text{Pa}_j$.

Observació 1.48. El graf $\mathcal{G}_M^{\mathcal{L}_1}$ viu a la capa observacional: serveix per representar una factorització de $P(\mathcal{V})$ com a xarxa bayesiana. No s'ha de confondre ni amb $\mathcal{G}_M^{\rightarrow}$ ni amb el diagrama causal, que introduïrem al capítol següent (cf. 2.4 i 2.5). En el cas Markovià, les C-components són trivials i $\mathcal{G}_M^{\mathcal{L}_1} = \mathcal{G}_M^{\rightarrow}$. En el cas semi-Markovià amb confusió latent, en canvi, $\mathcal{G}_M^{\mathcal{L}_1}$ pot contenir arestes dirigides addicionals introduïdes per la construcció de [Bar+20, Def. 26], provinents d'una possible confusió latent, mentre que el diagrama causal pot contenir arestes bidirigides.

Teorema 1.49 (SCM-BN observacional [Bar+20, App. C, Th. 8]). Sigui $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ un SCM acíclic, sigui $P(\mathcal{V})$ la distribució observacional induïda i sigui $\mathcal{G}_M^{\mathcal{L}_1}$ el DAG observacional induït per $\mathcal{M}$. Aleshores $(\mathcal{G}_M^{\mathcal{L}_1}, P(\mathcal{V}))$ és una xarxa bayesiana; és a dir, $P(\mathcal{V})$ factoritza sobre $\mathcal{G}_M^{\mathcal{L}_1}$. Explícitament:

$$P(\mathcal{V}) = \prod_{i=1}^n P(V_i \mid \text{Pa}_i^{\mathcal{G}_M^{\mathcal{L}_1}}).$$

El recíproc de [Bar+20, Th. 9] diu que tota xarxa bayesiana $(\mathcal{G}, P)$ pot ser induïda per algun SCM. En la nostra formulació, aquest fet queda recollit pel resultat de representabilitat següent.

Proposició 1.50 (Representabilitat [PJS17, Prop. 7.1]). Sigui una distribució conjunta $P(\mathcal{V})$ que admet una densitat respecte de la mesura de Lebesgue a $\mathbb{R}^n$, i sigui $\mathcal{G}$ un DAG tal que $P(\mathcal{V})$ factoritza sobre $\mathcal{G}$. Aleshores existeix un SCM Markovià $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ tal que $\mathcal{G} = \mathcal{G}_M^{\rightarrow}$ i indueix la distribució $P(\mathcal{V})$.

Nota 1.51. La hipòtesi de mesura de Lebesgue és la formulació amb valors reals de [PJS17, Prop. 7.1]. Quan treballem amb densitats, fem servir una mesura dominant producte μ (cf. 1.1 i B.10): mesura de Lebesgue en variables contínues reals i mesura de comptatge en variables discretes.

Tan rellevant com són les propietats de Markov, aquesta només ens indica quines independències condicionals de P són implicades per $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [PGJ16]. De fet, ni tan sols assegura que dues variables

unides per una aresta hagin de ser dependents respecte de P . Per garantir aquesta exigència ho podem fer amb una condició més forta: la *minimalitat*.

Definició 1.52 (Minimalitat causal o de Markov). P satisfà la *minimalitat causal* o *minimalitat de Markov* respecte $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ si hi factoritza, però no sobre cap subgraf $\mathcal{G}' \subset \mathcal{G}$, $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$. En altres paraules, la minimalitat demana que cada aresta aportí informació necessària per reproduir les dependències associades a $\mathcal{G}$.

Proposició 1.53 ([PJS17, Sec. 6.5, Prop. 6.35]). Si P és fidel i factoritza sobre $\mathcal{G}$, llavors P satisfà la *minimalitat causal*.

La minimalitat també es pot formular localment: cap node pot ser condicionalment independent d'un dels seus pares, donada la resta de pares.

Proposició 1.54 (Minimalitat causal [PJS17, Sec. 6.5, Prop. 6.36]). Suposem que P factoritza sobre $\mathcal{G}$ i que admet una densitat respecte de la mesura dominant de 1.1. Llavors, P satisfà la *minimalitat causal* respecte $\mathcal{G}$ si, i només si:

$$V_i \not\perp_P Y \mid \text{Pa}_i^{\mathcal{G}} \setminus \{Y\},$$

per a tot $V_i \in \mathcal{V}$ i per a tot $Y \in \text{Pa}_i^{\mathcal{G}}$.

Corol·lari 1.55. Si P factoritza sobre $\mathcal{G}$, admet una densitat respecte la mesura dominant i, a més, és minimal respecte $\mathcal{G}$, llavors cap aresta de $\mathcal{G}$ és redundant (en el sentit de $\perp_P$); en altres paraules, per a un ordre topològic compatible, $\text{PM}_i = \text{Pa}_i^{\mathcal{G}}$ per a tot $i = 1, \dots, n$. Si posteriorment dotem $\mathcal{G}$ de semàntica causal mitjançant un SCM, $\text{Pa}_i^{\mathcal{G}}$ són les causes directes dins del model.

Definició 1.56 (Submodel causal estructural [Pea09, Def. 7.1.2]). Sigui $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ un SCM, $\mathbf{X} \subset \mathcal{V}$ i $\mathbf{x}$ un valor particular d' $\mathbf{X}$. Un submodel $\mathcal{M}_{\mathbf{x}}$ d' $\mathcal{M}$ és el model $\mathcal{M}_{\mathbf{x}} = (\mathcal{U}, \mathcal{V}, \mathcal{F}_{\mathbf{x}}, P(\mathcal{U}))$ en què $\mathcal{F}_{\mathbf{x}} = \{f_i \mid V_i \notin \mathbf{X}\} \cup \{\mathbf{X} = \mathbf{x}\}$. És a dir, $\mathcal{F}_{\mathbf{x}}$ es forma eliminant de $\mathcal{F}$ totes les funcions f_i corresponents als membres de $\mathbf{X}$ i substituint-les pel conjunt d'assignacions constants $\mathbf{X} = \mathbf{x}$.

La definició de submodel anticipa la semàntica de les intervencions que formalitzarem al capítol següent: intervenir sobre una variable deixa intactes els mecanismes de la resta del sistema.

Definició 1.57 (Modularitat, segons [Nea20]). Sigui $\mathbf{X} \subseteq \mathcal{V}$ un conjunt de variables sobre el qual intervenim, prenent-les com a constants. Per a tot $i \in I$ tenim:

1. Si $V_i \notin \mathbf{X}$, $p(v_i \mid \text{pa}_i^{\mathcal{G}})$ no canvia.
2. Si $V_i \in \mathbf{X}$, aleshores fixem V_i al valor s_i . En el cas discret, això vol dir que $p(v_i \mid \text{pa}_i^{\mathcal{G}}) = 1$ quan $v_i = s_i$ i $p(v_i \mid \text{pa}_i^{\mathcal{G}}) = 0$ altrament.

En aquest punt hem adoptat deliberadament una aproximació menys estricta, atès que encara no hem introduït de manera formal el concepte d'*intervenció*. Tant aquest concepte com la modularitat seran desenvolupades en detall en el capítol següent (cf. 2.9 i 2.13, respectivament). En SCMs, la mateixa intuïció es tradueix en el fet que el submodel intervingut comparteix totes les equacions estructurals amb el model original excepte les de les variables intervingudes [Pea09, Secs. 1.4.3 i 3.2.1]; [Nea20].

Assumim, doncs, una intervenció local és quirúrgica (cf. Introducció) en tant que reemplaça només l'equació estructural de la variable intervinguda, deixant la resta completament inalterades.

CONFIGURACIONS GRÀFIQUES ELEMENTALS

Aquesta secció fa explícites les tres peces que apareixen contínuament en models gràfics: cadenes, bifurcacions i col·lisionadors. L'objectiu és com podem bloquejar o obrir camins d'associació.

Presentem ara les tres configuracions gràfiques bàsiques que apareixen en els grafs dirigits. Seguim amb la notació de 1.1, on ara $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ serà un DAG.

Definició 1.58 (Cadena (chain)). Un patró de tres variables $X, Y, Z \in \mathcal{V}$ és una *cadena* si $X \rightarrow Y \in \mathcal{E}$ i $Y \rightarrow Z \in \mathcal{E}$, és a dir, $X \rightarrow Y \rightarrow Z$. Y és un node intermedi entre X i Z i, quan el graf té semàntica causal, Y es pot interpretar com un *mediador* entre X i Z .

Observació 1.59. Per a tres variables aleatòries $X, Y, Z \in \mathcal{V}$ i el graf elemental format només per aquesta configuració, o quan la resta de camins entre els extrems estan blocats i P és Markov respecte del DAG, tenim:

- En la configuració $X \rightarrow Y \rightarrow Z$, les associacions $X \not\perp_P Y$ i $Y \not\perp_P Z$ es garanteixen sota hipòtesis addicionals com la fidelitat (cf. 1.38); sense aquestes hipòtesis, podríem tenir una cadena *intransitiva* respecte P : $X \not\perp_P Y$ i $Y \not\perp_P Z$, però $X \perp_P Z$.
- Condicionant per Y , X i Z esdevenen independents, i.e. $X \perp_P Z \mid Y$. Aquesta propietat es pot veure com un cas particular de la regla de la cadena de les xarxes bayesianes, suposant que P és Markov respecte del DAG:

$$p(x \mid y, z) = \frac{p(x, y, z)}{p(y, z)} = \frac{p(x) \cdot p(y \mid x) \cdot p(z \mid y)}{p(y) \cdot p(z \mid y)} = p(x \mid y), \quad p(x, y) > 0.$$

Definició 1.60 (Bifurcació (fork)). Un patró de tres variables $X, Y, Z \in \mathcal{V}$ és una *bifurcació* si $Y \rightarrow X \in \mathcal{E}$ i $Y \rightarrow Z \in \mathcal{E}$. En aquest cas, Y és el node central de la bifurcació i actua com a *causa comuna* de X i Z : $X \leftarrow Y \rightarrow Z$. En el graf elemental, les relacions d'independència exposades a 1.59 es mantenen, però la factorització canvia:

$$p(x \mid y, z) = \frac{p(x, y, z)}{p(y, z)} = \frac{p(x \mid y) \cdot p(y) \cdot p(z \mid y)}{p(y) \cdot p(z \mid y)} = p(x \mid y), \quad p(y) > 0.$$

Definició 1.61 (Col·lisionador (collider)). Un patró de tres variables $X, Y, Z \in \mathcal{V}$ és un *collider* si $X \rightarrow Y \in \mathcal{E}$ i $Z \rightarrow Y \in \mathcal{E}$. En aquest cas, el node Y és un *col·lisionador* on conflueixen dues arestes: $X \rightarrow Y \leftarrow Z$. A més, en aquest cas:

$$p(x, z \mid y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x) \cdot p(z) \cdot p(y \mid x, z)}{p(y)}.$$

Com ja hem fet servir més amunt, molts autors (per exemple, [KF09, Sec. 3.3.1]) utilitzen el terme *v-estructura* per referir-se a aquesta configuració gràfica.

Propietat 1.62. Si P factoritza sobre $\mathcal{G} := \{X \rightarrow Y \leftarrow Z\}$ o, més generalment, sobre un graf on tots els altres camins entre X i Z estan blocats sense condicionar, es compleix:

1. X i Z són independents marginalment (i.e. $X \perp_P Z$).
2. Fora dels casos especials en què, per a cada valor fixat de y , existeixen funcions $u(\cdot, y)$ i $v(\cdot, y)$ tals que $p(y | x, z) = c(y) u(x, y) v(z, y)$, condicionar pel col·lisionador obre el camí: $X \not\perp_P Z | Y$.

Definició 1.63 (Blocar per defecte). Els únics camins blocats per defecte són aquells que contenen almenys un *collider*. En la notació de 1.14, diem que un camí està blocat per defecte si $Z = \emptyset$.

Exemple 1.64 (Chain, fork i collider). Considerem la xarxa bayesiana de 1.11. En aquesta xarxa podem trobar les tres configuracions gràfiques que acabem d'introduir:

- *Chains.* $V_1 \rightarrow V_2 \rightarrow V_4$, $V_1 \rightarrow V_3 \rightarrow V_4$, $V_2 \rightarrow V_4 \rightarrow V_5$ i $V_3 \rightarrow V_4 \rightarrow V_5$ són cadenes. Condicionar pel mediador bloca el camí⁸; per exemple, $V_2 \perp_{\mathcal{G}} V_5 | V_4$ i $V_3 \perp_{\mathcal{G}} V_5 | V_4$. Com que P satisfà la propietat global de Markov (cf. 1.17 i 1.18), també tenim $V_2 \perp_P V_5 | V_4$ i $V_3 \perp_P V_5 | V_4$.
- *Fork.* En tenim una, $V_3 \leftarrow V_1 \rightarrow V_2$. Sense condicionar, V_2 i V_3 estan d -connectades. Si condicionem en V_1 , bloquegem aquest camí: $V_2 \perp_{\mathcal{G}} V_3 | V_1$. L'altre camí entre V_2 i V_3 , $V_2 \rightarrow V_4 \leftarrow V_3$, ja està bloquejat perquè passa per un *collider* no condicionat. Per tant, $V_2 \perp_P V_3 | V_1$.
- *Collider.* Tenim $V_2 \rightarrow V_4 \leftarrow V_3$. Si condicionem per V_1 i també pel node central del *collider*, V_4 (o per un descendent seu, com V_5), el camí es desbloqueja i $V_2 \not\perp_{\mathcal{G}} V_3 | V_1, V_4$. Aquest efecte s'anomena *explaining away*.

Nota 1.65. A primera vista, pot semblar paradoxal que condicionar un node que es troba en un camí bloquejat pugui, precisament, desbloquejar-lo. Això passa en situacions de l'estil $X \rightarrow Z \leftarrow Y$ perquè, un cop condicionem per Z , la informació sobre una de les causes (X o Y) altera la probabilitat de l'altra. En l'àmbit de l'estadística, aquest concepte s'anomena *biaix de selecció* o *paradoxa de Berkson*, mentre que en la intel·ligència artificial es coneix com l'efecte *explaining away*.

Exemple 1.66 (Explaining away). Agafem la vinyeta satírica *Semiconcontrolled Demolition* i suposem que Y és la variable aleatòria (binària) corresponent a l'esfondrament de les Torres Bessones. Al seu torn, X i Z corresponen a les seves possibles causes: explosius i atemptat terrorista, respectivament. El graf està format per un únic *collider*, pel que $\mathcal{G} = \{X \rightarrow Y \leftarrow Z\}$. $X \perp_{\mathcal{G}} Z$ perquè Y bloca el camí *per defecte* (i si, a més, hi hagués fidelitat, $X \perp_P Z$). En canvi, si observem l'esfondrament ($Y = 1$), condicionar per Y desbloqueja el camí i obtenim $X \not\perp_P Z | Y$ (cf. 1.62). En aquest exemple, $Y = 1$ no pot ocórrer si $X = 0$ i $Z = 0$. Si, malgrat això, X i Z fossin independents condicionadament a $Y = 1$, tindríem:

$$P(X = 0, Z = 0 | Y = 1) = P(X = 0 | Y = 1) \cdot P(Z = 0 | Y = 1) = 0,$$

D'aquesta igualtat se'n seguiria que $P(X = 0 | Y = 1) = 0$ o bé $P(Z = 0 | Y = 1) = 0$, però per hipòtesi $P(X = 0 | Y = 1), P(Z = 0 | Y = 1) \neq 0$. Per tant, condicionat a $Y = 1$, l'evidència a favor d'una de les causes alternatives redueix la probabilitat de l'altra; així, la *compromise theory* (que totes dues causes hagin actuat alhora) és poc versemblant.

⁸ És el mateix fenomen que hem vist a 1.59. També passa si condicionem per la causa comuna en un *fork*, cf. 1.60. Caldria aplicar la hipòtesi de fidelitat (cf. 1.38) per passar de d -separació a independència probabilística.

Identificació d'efectes causals

La natura no ha produït res perquè fos immòbil.

SÈNECA, *Qüestions naturals*, 6, 12

En aquest capítol passem de descriure relacions causals a preguntar quan es poden calcular a partir de dades observacionals. La noció clau és la intervenció: imposar el valor d'una variable X i estudiar com canvia la distribució d'una resposta Y . Intuïtivament, podem dir que una intervenció ideal modifica X sense obrir altres vies cap a Y [Woo03].

Notació 2.1. Mantenim la convenció del capítol 1, 1.1 i 1.2: treballarem amb grafs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ sobre $\mathcal{V} = \{V_i\}_{i \in I}$, amb $I = \{1, \dots, n\}$. L'espai producte on prenen valors les variables és $\mathcal{X} = \prod_{i \in I} X_i$ i prenem $\mathbf{v} = (v_i)_{i \in I} \in \mathcal{X}$. Si $\mathbf{S} \subseteq \mathcal{V}$, escriurem $\mathbf{s} := \mathbf{v}_{\mathbf{S}} = (v_i)_{V_i \in \mathbf{S}} \in \mathcal{X}_{\mathbf{S}}$. A més, $\mathbf{a} \setminus \mathbf{b} := \mathbf{v}_{\mathbf{A} \setminus \mathbf{B}}$. En aquest capítol, seguint la notació de Pearl i [Bar+22], $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \subseteq \mathcal{V}$ denotaran conjunts de variables endògenes observades i $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$ en denotaran valors.

- Donat $\mathbf{X} \subseteq \mathcal{V}$, denotem per $P(\mathcal{V} \mid \text{do}(\mathbf{X} = \mathbf{x}))$ la distribució *intervinguda* per $\text{do}(\mathbf{X} = \mathbf{x})$ sobre $\mathcal{V}$ (cf. 2.10). Treballarem amb la família:

$$\mathcal{P}_* = \{P(\mathcal{V} \mid \text{do}(\mathbf{X} = \mathbf{x})) \mid \mathbf{X} \subseteq \mathcal{V}, \mathbf{x} \in \mathcal{X}_{\mathbf{X}}\}, \quad (2.1)$$

on la distribució observacional és el cas particular $P(\mathcal{V}) = P(\mathcal{V} \mid \text{do}(\emptyset))$.

- Així doncs, com en 1.1 i 1.2, denotem per $P(\mathcal{V})$ la distribució sobre $\mathcal{V}$ (la *intervenció nul·la*). Quan una distribució en $\mathcal{P}_*$ estigui ben definida i admeti una densitat respecte de la mesura dominant μ , la denotarem per p [Bon+21, Sec. 2.4]. Aleshores, per exemple, tenim $p(\mathbf{v}) = p(v_1, \dots, v_n)$ i $p(\mathbf{v} \mid \text{do}(\mathbf{x})) = p(v_1, \dots, v_n \mid \text{do}(\mathbf{x}))$. Quan ens quedem amb $\mathbf{Y} \subseteq \mathcal{V}$, escriurem $P(\mathbf{Y} \mid \text{do}(\mathbf{x}))$ o $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$.

Escriurem $\mathcal{V}^{(i)} = \{V_1, \dots, V_i\}$, amb $\mathcal{V}^{(0)} = \emptyset$; la restricció corresponent d'una assignació $\mathbf{v}$ serà $\mathbf{v}^{(i)} = \mathbf{v}_{\mathcal{V}^{(i)}}$; també escriurem $\mathbf{s}^{(i)} := \mathbf{v}_{\mathbf{S} \cap \mathcal{V}^{(i)}}$ i $\mathbf{v}^{(i)} \setminus \mathbf{s} := \mathbf{v}_{\mathcal{V}^{(i)} \setminus \mathbf{S}}$. S'aplicarà la mateixa convenció als valors $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$, etcètera.

Per a un subconjunt $\mathbf{S} \subseteq \mathcal{V}$, $\mathcal{G}[\mathbf{S}]$ denotarà el subgraf induït i $\mathbf{C}(\mathcal{G})$ la partició de $\mathcal{G}$ en C-components maximals. Finalment, quan escriguem $\mathcal{G}_{\bar{\mathbf{X}}}$ o $\mathcal{G}_{\bar{\mathbf{X}}, \underline{\mathbf{Z}}}$, seguirem la convenció del capítol C.

Definició 2.2 (Llenguatges de la jerarquia causal, [Bar+22, Sec. 27.3, Def. 27.7]). Fixades les variables endògenes $\mathcal{V}$, distingirem tres capes de consultes:

1. El llenguatge $\mathcal{L}_1$ conté consultes *associacionals*, formades a partir de probabilitats observacionals com $P(\mathbf{Y} = \mathbf{y})$ o, quan estigui ben definida, $P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x})$. És el que hem usat al llarg del capítol anterior.

2. El llenguatge $\mathcal{L}_2$ afegeix consultes *intervencionals*, com $P(Y = y \mid \text{do}(X = x))$, equivalentment escrites com $P(Y_x = y)$ (en llenguatge de resultats potencials).
3. El llenguatge $\mathcal{L}_3$ permet consultes *contrafactuals* conjuntes sobre enunciats intervinguts possiblement incompatibles, com $P(Y_x = y, Z_w = z)$.

La inclusió d'un terme en una capa indica quin tipus d'informació causal pressuposa; en general, les distribucions de $\mathcal{L}_1$ no determinen les de $\mathcal{L}_2$ ni les de $\mathcal{L}_3$.

Teorema 2.3 (Causal Hierarchy Theorem [Bar+22, Sec. 27.3, Th. 27.1]). *Sigui $\Omega_{\mathcal{V}}$ la classe d'SCMs acíclics amb variables endògenes $\mathcal{V}$, amb $|\mathcal{V}| > 1$ i dominis finits no degenerats. Per a $i = 1, 2, 3$, escrivim $M \sim_i M'$ si M i M' assignen el mateix valor a totes les consultes del llenguatge $\mathcal{L}_i$. Direm que la capa $\mathcal{L}_j$ col·lapsa a la capa $\mathcal{L}_i$ en $M^* \in \Omega_{\mathcal{V}}$, amb $i < j$, si $M^* \sim_i M \implies M^* \sim_j M$, per a tot $M \in \Omega_{\mathcal{V}}$. Aleshores, respecte de la mesura de Lebesgue en una codificació adequada del quocient $\Omega_{\mathcal{V}}/\sim_3$, el conjunt d'SCMs M^* per als quals hi ha algun col·lapse $\mathcal{L}_j \rightarrow \mathcal{L}_i$, amb $1 \leq i < j \leq 3$, té mesura zero.*

El que ens diu aquest teorema en particular és que per passar a una capa superior no podem fer-ho automàticament, sinó que cal afegir informació causal sobre l'SCM o restriccions estructurals addicionals.

2.1

DIAGRAMES CAUSALS INDUÏTS PER SCMs

Ara passem del DAG observacional de capa $\mathcal{L}_1$ al graf que codifica els mecanismes causals d'un SCM i que sosté les preguntes intervencionals.

Al capítol anterior hem vist que un SCM indueix una distribució observacional $P(\mathcal{V})$ i que aquesta es pot representar mitjançant un DAG observacional $\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}$ (cf. 1.47 i 1.49). Aquest graf és una eina de capa $\mathcal{L}_1$. En canvi, per formular intervencions i CBNs necessitem el diagrama causal del model, és a dir, el graf que recull les dependències funcionals i la possible confusió latent.

Les definicions 1.42 i 1.43 determinen quin tipus de diagrama causal s'indueix: en el subcas Markovià obtenim un DAG, mentre que en el cas semi-Markovià general obtenim un ADMG, que es redueix a un DAG quan no hi ha confusió latent; al seu torn, anticipen la distinció entre xarxes bayesianes causals Markovianes i semi-Markovianes, que formalitzarem després.

Definició 2.4 (Diagrama causal, cas Markovià [Bar+22, Sec. 27.4, Def. 27.10]). *Sigui un model causal Markovià $M = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$. El diagrama causal Markovià induït per M és el DAG $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}_M^{\rightarrow})$ construït de la següent manera:*

1. Els nodes de $\mathcal{G}_M$ són les variables endògenes $\mathcal{V}$.
2. Les arestes dirigides $\mathcal{E}_M^{\rightarrow}$ són les de (1.1): $V_j \rightarrow V_i \in \mathcal{E}_M^{\rightarrow} \iff V_j \in \text{Pa}_i^M$.
3. Coincideix amb el graf $\mathcal{G}_M^{\rightarrow}$ de 1.41, $\mathcal{G}_M := \mathcal{G}_M^{\rightarrow}$.

Definició 2.5 (Diagrama causal, cas semi-Markovià [Bar+22, Sec. 27.4, Def. 27.12]). *Sigui M un model causal semi-Markovià. El diagrama causal semi-Markovià induït per M és l'ADMG $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}_M^{\rightarrow} \cup \mathcal{E}_M^{\leftrightarrow})$, és a dir, el graf acíclic dirigit mixt construït de la següent manera:*

1. Els nodes de $\mathcal{G}_M$ són les variables endògenes $\mathcal{V}$.
2. Les arestes dirigides $\mathcal{E}_M^{\rightarrow}$ són les de (1.1): $V_j \rightarrow V_i \in \mathcal{E}_M^{\rightarrow} \iff V_j \in \text{Pa}_i^M$.
3. El conjunt d'arestes bidirigides $\mathcal{E}_M^{\leftrightarrow}$ està format per les arestes $V_j \leftrightarrow V_i$ si V_i i V_j comparteixen algun factor exogen o, més generalment, si els conjunts de factors exògens $\mathbf{U}_i^c$ i $\mathbf{U}_j^c$ no són independents respecte a $P(\mathcal{U})$.

Proposició 2.6 (Condicció causal de Markov [PJS17, Prop. 6.31]). *Sigui M un SCM Markovià acíclic i sigui $\mathcal{G}_M$ el diagrama causal Markovià induït per M . Aleshores la distribució observacional induïda $P(\mathcal{V})$ factoritza respecte de $\mathcal{G}_M$. Quan $P(\mathcal{V})$ admet una densitat o PMF, aquesta es pot escriure com:*

$$p(\mathbf{v}) = \prod_{i=1}^n p(v_i \mid \text{pa}_i^{\mathcal{G}_M}).$$

Nota 2.7. Com ja hem comentat a 1.48, en el cas Markovià es compleix que $\mathcal{G}_M^{\mathcal{L}_1} = \mathcal{G}_M^{\rightarrow} = \mathcal{G}_M$. En el cas semi-Markovià amb confusió latent, en canvi, $\mathcal{G}_M^{\mathcal{L}_1}$ i $\mathcal{G}_M^{\rightarrow}$ poden no coincidir: el primer pot afegir arestes dirigides per representar la factorització observacional i el segon no, mentre que el diagrama causal $\mathcal{G}_M$ afegeix, a més, les arestes bidirigides de confusió latent.

2.2

XARXES BAYESIANES CAUSALS (CBNs)

Les CBNs afegeixen intervencions al diagrama causal i separen observació d'acció.

El graf amb què treballarem no és només una xarxa bayesiana amb arestes interpretades informalment com a causes. La idea fonamental que hi ha al darrere d'una xarxa bayesiana causal és la combinació d'un diagrama causal que prové d'un SCM (cf. 1.40 i 1.41) i una família de distribucions intervingudes induïda pel model. El diagrama passa a codificar quins mecanismes es mantenen invariants i quan una intervenció es pot substituir per una observació.

Exemple 2.8. Considerem dues variables: el nivell real de glucosa a la sang d'un pacient, G , i la lectura mostrada per un glucòmetre, L . Si no hi ha errors de calibració en el glucòmetre, L depèn de G i les dues variables són estadísticament dependents. Aquesta dependència observacional, però, és compatible amb més d'una estructura causal:

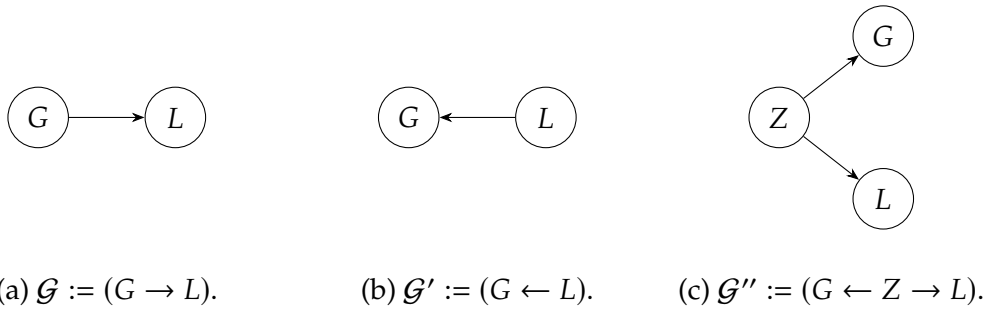


Figura 2.1: Diagrames causals possibles per a l'exemple de la glucosa.

Z representa una causa comuna de G i L . Els tres diagrames poden ser observacionalment equivalents, tot i que impliquen efectes d'intervenció diferents [Par23, Sec. 1.1.2, Fig. 1.3].

Definició 2.9 (Operador $\text{do}()$ [Pea09, Sec. 3.2.1, Def. 3.2.1]). Sigui $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ un SCM, $\mathbf{X} \subseteq \mathcal{V}$ i $\mathbf{x} = (x_i)_{V_i \in \mathbf{X}} \in \mathcal{X}_{\mathbf{X}}$. Denotem per $\text{do}(\mathbf{X} = \mathbf{x})$, abreujadament $\text{do}(\mathbf{x})$, la intervenció que imposa a les variables de $\mathbf{X}$ els valors $\mathbf{x}$. Aquesta intervenció produeix el submodel $\mathcal{M}_{\mathbf{x}} = (\mathcal{U}, \mathcal{V}, \mathcal{F}_{\mathbf{x}}, P(\mathcal{U}))$, on $\mathcal{F}_{\mathbf{x}}$ s'obté substituint, per a cada $V_i \in \mathbf{X}$, l'equació estructural $V_i = f_i(\text{pa}_i, \mathbf{u}_{V_i})$ per l'assignació constant $V_i := x_i$, i deixant inalterades totes les equacions f_j amb $V_j \notin \mathbf{X}$.

Definició 2.10 (Distribució intervinguda). Sigui $\mathbf{X} \subseteq \mathcal{V}$ el conjunt de variables intervingudes. La distribució conjunta intervinguda del sistema és $P(\mathcal{V} \mid \text{do}(\mathbf{X} = \mathbf{x}))$, abreujadament $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$. Per a qualsevol subconjunt de variables d'interès $\mathbf{Y} \subseteq \mathcal{V} \setminus \mathbf{X}$, escriurem $P(\mathbf{Y} \mid \text{do}(\mathbf{X} = \mathbf{x}))$, o simplement $P(\mathbf{Y} \mid \text{do}(\mathbf{x}))$.

Nota 2.11 (Efectivitat i graf mutilat). Per a tota variable $V_i \in \mathbf{X}$, si x_i és el valor imposat per $\mathbf{x}$ sobre V_i , la llei intervinguda de V_i és la mesura de Dirac δ_{x_i} : per a tot $A \in \mathcal{B}(\mathcal{X}_i)$, $P(V_i \in A \mid \text{do}(\mathbf{x})) = \delta_{x_i}(A)$. En el cas discret, això equival a $P(V_i = v_i \mid \text{do}(\mathbf{x})) = \mathbb{1}_{\{v_i = x_i\}}$. Aquesta propietat s'anomena *efectivitat*. Gràficament, la intervenció elimina totes les arestes que entren als nodes intervinguts; el graf resultant és $\mathcal{G}_{\overline{\mathbf{X}}}$ (cf. C.24). Aquesta intervenció també rep el nom d'*intervenció perfecta* [Bon+21, Sec. 2.4, Def. 2.12] o *intervenció atòmica* [PJS17, Sec. 6.3, Def. 6.8].

Al llarg d'aquest treball ens restringirem essencialment a l'escenari acíclic. Els models semi-Markovians que considerarem més endavant incorporen factors de confusió latents mitjançant arestes bidirigides, però no introdueixen cicles causals dirigits.

Exemple 2.12 (2.8, continuació). Suposem que intervenim sobre L i li imposem una lectura de glucosa l , escrit $\text{do}(L = l)$. En el graf 2.1(a), tenim $P(G \mid \text{do}(L = l)) = P(G)$, ja que manipular el lector no altera el nivell real de glucosa. En canvi, si intervenim sobre G i imposem un nivell de glucosa g (per exemple, amb una injecció d'insulina), la distribució de la lectura sí que canvia; en aquest graf, $P(L \mid \text{do}(G = g)) = P(L \mid G = g)$.

Nota 2.13. Arribats aquí, podem revisar la modularitat de 1.57: és una restricció sobre com canvien els mecanismes locals sota intervencions. En les CBNs Markovianes, això es formalitza exigint que la família de distribucions intervingudes sigui compatible amb el diagrama causal i satisfaci les condicions de *missing-link* i *parents do/see*. El problema pràctic serà expressar una distribució intervinguda en termes de $P(\mathcal{V})$ quan això sigui possible. La definició següent estableix el cas Markovià.

Definició 2.14 (Xarxa bayesiana causal Markoviana [Bar+22, Def. 27.11, Th. 27.2]).

Una *xarxa bayesiana causal* (CBN, de l'anglès *Causal Bayesian Network*) Markoviana és un parell $(\mathcal{G}, \mathcal{P}_*)$, on $\mathcal{G}$ és un diagrama causal Markovià sobre $\mathcal{V}$ i $\mathcal{P}_*$ és la família de distribucions de (2.1), tal que per a tot subconjunt $\mathbf{X} \subseteq \mathcal{V}$ i tota intervenció $\text{do}(\mathbf{x})$, la distribució intervinguda $P(\mathcal{V} \mid \text{do}(\mathbf{x})) \in \mathcal{P}_*$ satisfà aquestes condicions respecte de $\mathcal{G}$ (cf. [Pea09, Def. 1.3.1]):

1. *Markovianitat*: $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$ és Markov respecte de $\mathcal{G}$.

2. *Parents do/see*: Per a tota variable no intervinguda $V_i \notin \mathbf{X}$, tenim:

$$p(v_i \mid \text{do}(\mathbf{x}), \text{do}(\text{pa}_i^{\mathcal{G}})) = p(v_i \mid \text{do}(\mathbf{x}), \text{pa}_i^{\mathcal{G}}).$$

3. *Missing link*: si ens restringim, a més, a aquelles $V_i \notin \mathbf{X}$ tals que no hi hagi cap aresta de $\mathbf{X}$ a V_i , llavors:

$$p(v_i \mid \text{do}(\mathbf{x}), \text{do}(\text{pa}_i^{\mathcal{G}})) = p(v_i \mid \text{do}(\text{pa}_i^{\mathcal{G}})).$$

La condició *parents do/see* diu que, per a una variable no intervinguda, intervenir sobre els seus pares és equivalent a observar-los. La condició *missing link* formalitza el fenomen de 2.12: si ja intervenim sobre els pares d'una variable, una intervenció addicional en una variable sense aresta cap a ella no en canvia la distribució.

Nota 2.15 ([Bar+22, Sec. 27.4]). Les dues darreres condicions corresponen a la *modularitat* (cf. 1.57 i 2.13) i, en cap cas, venen implicades per la Markovianitat.

Teorema 2.16 (SCM-CBN Markoviana [Bar+22, Sec. 27.4, Th. 27.2]). *Sigui un model causal Markovià $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ i sigui $\mathcal{G}$ el diagrama causal Markovià induït per $\mathcal{M}$ (cf. 2.4). Llavors $(\mathcal{G}, \mathcal{P}_*^{\mathcal{M}})$ és una CBN Markoviana.*

Aquest resultat està demostrat a [Bar+20, App. D].

Definició 2.17 (Efecte d'una acció [Pea09, Sec. 3.2.1, Def. 3.2.1]). *Sigui $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ un SCM i $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{V}$. L'efecte de $\text{do}(\mathbf{x})$ en $\mathbf{Y}$ és la distribució $P(\mathbf{Y} \mid \text{do}(\mathbf{x}))$.*

L'efecte causal $P(\mathbf{Y} \mid \text{do}(\mathbf{x}))$ depèn de la CBN que descriu el sistema, no només de la distribució observacional P . En l'exemple de la glucosa, l'efecte d'intervenir sobre el glucòmetre depèn de quin diagrama causal triem entre els observacionalment compatibles. En 2.1(a) i 2.1(c), manipular el lector no altera la glucosa real i $P(G \mid \text{do}(L = l)) = P(G)$. En canvi, en 2.1(b), L és causa directa de G i, en absència de confusió, la regla *do/see* dona $P(G \mid \text{do}(L = l)) = P(G \mid L = l)$.

Teorema 2.18 (Factorització truncada per a CBNs Markovianes [Bar+22, Sec. 27.4, Th. 27.3]). *Sigui una CBN Markoviana $(\mathcal{G}, \mathcal{P}_*)$. Per a tot $\mathbf{X} \subseteq \mathcal{V}$ i tot $\mathbf{x} \in \mathcal{X}_{\mathbf{X}}$, la distribució intervinguda $P(\mathcal{V} \mid \text{do}(\mathbf{x})) \in \mathcal{P}_*$ és identifiable mitjançant informació purament observacional. En el cas discret, la PMF truncada completa és:*

$$p(\mathbf{v} \mid \text{do}(\mathbf{x})) = \prod_{i \mid V_i \notin \mathbf{X}} p(v_i \mid \text{pa}_i^{\mathcal{G}}) \Big|_{\mathbf{x}=\mathbf{x}} \cdot \prod_{i \mid V_i \in \mathbf{X}} \mathbb{1}_{\{v_i=x_i\}}.$$

Equivalentment, si ens quedem només amb la llei sobre $\mathcal{V} \setminus \mathbf{X}$ i aquesta admet una densitat o PMF, podem escriure:

$$p(\mathbf{v} \setminus \mathbf{x} \mid \text{do}(\mathbf{x})) = \prod_{i \mid V_i \notin \mathbf{X}} p(v_i \mid \text{pa}_i^{\mathcal{G}}) \Big|_{\mathbf{x}=\mathbf{x}}.$$

$\cdot|_{\mathbf{x}=\mathbf{x}}$ indica que, en aquests factors observacionals, qualsevol pare que pertanyi a $\mathbf{X}$ s'avalua a $\mathbf{x}$. Les coordenades intervingudes segueixen les mesures de Dirac δ_{x_j} ⁹.

⁹ En aquest context, δ_{x_j} és la mesura de probabilitat concentrada al punt x_j , definida com $\delta_{x_j}(A) = 1$ si $x_j \in A$ i 0 si $x_j \notin A$.

Sota suficiència causal les distribucions intervencionals es poden obtenir mitjançant la factorització truncada. Per aquest motiu, en el cas continu és habitual escriure la densitat només sobre les variables no intervingudes.

Nota 2.19 (Suficiència causal i confusió). Quan s'exigeix *suficiència causal*, no existeix en el model cap variable latent inobservable que actui com a causa comuna de *dues* variables endògenes (cf. 1.42, 1.43 i 1.44). Si prenem $X, Y \in \mathcal{V}$ i $U \in \mathcal{U}$, una confusió latent entre X i Y es pot representar com:

- com un *fork* de l'estil $X \leftarrow U \rightarrow Y$ (explícitament), o bé
- mitjançant arestes bidirigides discontinues $X \leftrightarrow Y$ (implícitament).

El primer cas correspon a $\mathcal{G}_{\mathcal{M}}^+$, on les variables latents apareixen explícitament (cf. 1.41). El segon és un graf acíclic mixt, o ADMG, que anomenem *projecció latent*: prové de $\mathcal{G}_{\mathcal{M}}^+$, però només treballa sobre variables observades. Aquesta és la representació ADMG que, en la construcció de 2.5, anomenem diagrama causal semi-Markovià induït per l'SCM, i precisament és el diagrama causal $\mathcal{G}_{\mathcal{M}} = (\mathcal{V}, \mathcal{E}_{\mathcal{M}}^{\rightarrow} \cup \mathcal{E}_{\mathcal{M}}^{\leftrightarrow})$ que utilitzarem per a les CBNs semi-Markovianes.

Amb tot, assumir que tots els factors rellevants estan mesurats és poc habitual en el món real. Per això el cas semi-Markovià amb confusió latent (cf. 1.43 i 2.19) és el més interessant, però també altera les condicions de factorització. Per formular aquesta factorització introduïm ara les C-components i els pares estesos.

Definició 2.20 (C-component, [TP02; SP06]). Sigui $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana i $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ un subgraf de $\mathcal{G}$, de manera que $\mathcal{V}_{\mathcal{H}} \subseteq \mathcal{V}$. Direm que $\mathcal{H}$ és una **C-component** si existeix un subconjunt $\mathcal{E}' \subseteq \mathcal{E}_{\mathcal{H}}$ format només per arestes bidirigides tal que $(\mathcal{V}_{\mathcal{H}}, \mathcal{E}')$ és connex.

Tot i que la literatura acostuma a reservar el nom C-component per a les components maximals de la partició, nosaltres distingim explícitament entre C-component i C-component maximal.

Definició 2.21 (C-component maximal [SP06; SP08]). Sigui $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana i sigui $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ una C-component de $\mathcal{G}$. Direm que $\mathcal{H}$ és una **C-component maximal** si, per a tot camí bidirigit en $\mathcal{G}$ que contingui algun node de $\mathcal{V}_{\mathcal{H}}$, aquell ja és un camí en $\mathcal{H}$.

Lema 2.22. Sigui una CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*)$ amb $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. El conjunt de variables $\mathcal{V}$ es pot dividir en subconjunts disjunts $\mathbf{S}_1, \dots, \mathbf{S}_k$ de tal manera que el conjunt de subgrafs induïts $C(\mathcal{G}) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$ és una partició de C-components maximals de $\mathcal{G}$.

Per a cada variable endògena V_i necessitem condicionar respecte als seus pares, les variables de la mateixa C-component que la precedeixen, i els pares d'aquestes últimes, amb l'objectiu d'aïllar V_i de la resta de variables no descendents en el diagrama causal. Aquest conjunt maximal es defineix formalment com Pa_i^+ de la següent manera.

Definició 2.23 (Pares estesos [Bar+22, Sec. 27.4]). Sigui $<$ un ordre topològic sobre les variables de $\mathcal{G}$, amb una enumeració compatible, i sigui $\mathcal{G}^{(i)}$ el subgraf de $\mathcal{G}$ induït per les variables $\mathcal{V}^{(i)} := \{V_j \in \mathcal{V} \mid j \leq i\}$. Donat $\mathbf{S} \subseteq \mathcal{V}$, definim $\text{Pa}_1^{\mathcal{G}}(\mathbf{S})$ com $\text{Pa}_1^{\mathcal{G}}(\mathbf{S}) = \mathbf{S} \cup \bigcup_{S \in \mathbf{S}} \text{Pa}_S^{\mathcal{G}}$. Sigui $C^{(i)}$ la C-component a la

per a tot conjunt mesurable A .

qual pertany V_i restringida al subgraf $\mathcal{G}^{(i)}$. Llavors, el conjunt de **pares estesos** de V_i es defineix com:

$$\text{Pa}_i^+ = \text{Pa}_1^{\mathcal{G}} \left(\{V_j \in \mathcal{C}^{(i)} \mid j \leq i\} \right) \setminus \{V_i\}.$$

Per a les CBNs semi-Markovianes farem servir la formulació finita discreta de [Bar+22, Sec. 27.1.2 i 27.4], en què p denota una PMF, amb una restricció de factorització anàloga a 1.4 i a 2.18.

Definició 2.24 (Distribució semi-Markoviana [Bar+22, Sec. 27.4, Def. 27.14]). Es diu que una distribució P és semi-Markoviana respecte (*semi-Markov relative to*) un graf $\mathcal{G}$ si, per a qualsevol ordre topològic $<$ sobre les variables, compatible amb $\mathcal{G}$, i amb l'enumeració corresponent, P factoritza com:

$$p(\mathbf{v}) = p(v_1, \dots, v_n) = \prod_{V_i \in \mathcal{V}} p(v_i \mid \text{pa}_i^+),$$

on pa_i^+ són els *pares estesos* definits a 2.23.

Observació 2.25 ([Bar+22, Sec. 27.4, Defs. 27.14–27.15]). La mateixa relació s'aplica a distribucions intervingudes $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$. En aquest cas, el graf esdevé el graf mutilat $\mathcal{G}_{\bar{\mathbf{x}}}$ (cf. 2.11) i la factorització és:

$$p(\mathbf{v} \mid \text{do}(\mathbf{x})) = p(v_1, \dots, v_n \mid \text{do}(\mathbf{x})) = \prod_{V_i \in \mathcal{V}} p(v_i \mid \text{pa}_i^{x+}, \text{do}(\mathbf{x})),$$

on pa_i^{x+} es construeix exactament igual que pa_i^+ , però ara sobre el graf mutilat $\mathcal{G}_{\bar{\mathbf{x}}}$.

Amb això podem definir les CBNs semi-Markovianes: els conceptes de *missing-link* i *do/see* es reformulen tenint en compte que ara hi ha confusió no observada.

Definició 2.26 (CBN semi-Markoviana [Bar+22, Sec. 27.4, Def. 27.15 i Th. 27.4]). Sigui $\mathcal{P}_*$ la col·lecció (2.1) de distribucions $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$, per a qualsevol $\mathbf{X} \subseteq \mathcal{V}$ i $\mathbf{x} \in \mathcal{X}_{\mathbf{X}}$. Una CBN semi-Markoviana és un parell $(\mathcal{G}, \mathcal{P}_*)$, on $\mathcal{G}$ és un diagrama causal semi-Markovià sobre $\mathcal{V}$, tal que, per a qualsevol intervenció $\text{do}(\mathbf{x})$, es compleixen aquestes condicions respecte de $\mathcal{G}$:

1. La distribució $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$ és semi-Markov (cf. 2.24) respecte al graf mutilat $\mathcal{G}_{\bar{\mathbf{x}}}$.
2. *Aresta dirigida absent (missing directed-link)*: Per a tota variable no intervinguda $V_i \in \mathcal{V} \setminus \mathbf{X}$, i qualsevol subconjunt $\mathbf{W} \subseteq \mathcal{V} \setminus (\text{Pa}_i^{x+} \cup \mathbf{X} \cup \{V_i\})$:

$$p(v_i \mid \text{do}(\mathbf{x}), \text{do}(\mathbf{w}), \text{pa}_i^{x+}) = p(v_i \mid \text{do}(\mathbf{x}), \text{pa}_i^{x+}).$$

3. *Aresta bidirigida absent (missing bidirected-link)*: Per a tota variable $V_i \in \mathcal{V} \setminus \mathbf{X}$, prenem els seus pares estesos en $\mathcal{G}_{\bar{\mathbf{x}}}$ i els separem en pares confosos (Pa_i^c) i pares no confosos (Pa_i^u). Els pares confosos són els $P \in \text{Pa}_i^{x+}$ connectats amb V_i per un camí bidirigit en $\mathcal{G}_{\bar{\mathbf{x}}}$, i els pares no confosos són la resta. Llavors:

$$p(v_i \mid \text{do}(\mathbf{x}), \text{do}(\text{pa}_i^u), \text{pa}_i^c) = p(v_i \mid \text{do}(\mathbf{x}), \text{pa}_i^c, \text{pa}_i^u).$$

Teorema 2.27 (Connexió SCM-CBN semi-Markoviana [Bar+22, Th. 27.4]). Sigui un model causal semi-Markovià $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ i sigui $\mathcal{G}$ el diagrama causal semi-Markovià induït per $\mathcal{M}$ (cf. 2.5). Llavors $(\mathcal{G}, \mathcal{P}_*^{\mathcal{M}})$ és una CBN semi-Markoviana, on $\mathcal{P}_*^{\mathcal{M}}$ és la col·lecció (2.1) induïda per $\mathcal{M}$.

Nota 2.28 (Interpretació de la CBN Semi-Markoviana). Totes les condicions són relaxacions de les propietats que es compleixen en les CBNs Markovianes:

- La segona condició diu que, sota $\text{do}(\mathbf{x})$, si condicionem pels pares estesos en el graf intervingut $(\text{Pa}_i^{\mathbf{x}^+})$, la variable V_i és invariant a intervencions sobre altres variables $\mathbf{W}$.
- La tercera condició relaxa la propietat *do/see*: amb confusió no observada ja no podem igualar sempre intervenció i observació, però sí per als pares estesos no confosos $(\text{Pa}_i^{\mathbf{u}})$.

En models semi-Markovians, la identificació es troba en una zona grisa. Les restriccions de la capa $\mathcal{L}_1$ (la primera capa de la jerarquia de Pearl) d'una xarxa bayesiana no determinen, en general, els efectes causals. A l'altre extrem trobem els models Markovians amb la hipòtesi de suficiència causal, que ens garanteix que l'efecte serà **sempre** identificable. El repte de les CBNs semi-Markovianes és que la resposta no és ni un sí ni un no rotund, sinó que només ho podrem fer en casos concrets, però quins?

2.3

IDENTIFICABILITAT I DO-CALCULUS

Quan podem calcular l'efecte d'una intervenció només amb dades observacionals? Aquesta secció introdueix els criteris back-door i front-door, les regles del do-calculus i la noció formal d'identificabilitat.

Tornem a una pregunta fonamental del capítol i, per extensió, del treball. Donada una CBN i una consulta causal, volem saber si $P(\mathbf{Y} \mid \text{do}(\mathbf{x}))$ es pot expressar només amb termes observacionals. Aquest és el problema de la **identificabilitat**. El conjunt de regles sintàctiques per resoldre'l, desenvolupades per Judea Pearl [Pea95], permeten transformar expressions amb intervencions sota certes condicions.

Definició 2.29 (Criteri de la porta enrere [Pea93]; [Par23, Sec. 1.1.4]). Donada $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana, amb $P(\mathcal{V}) \in \mathcal{P}_*$, i tres subconjunts de variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts dos a dos, diem que $\mathbf{Z}$ satisfà el criteri de la porta enrere (*Back-door criterion*) respecte a $(\mathbf{X}, \mathbf{Y})$ si:

- Cap variable de $\mathbf{Z}$ és descendent de $\mathbf{X}$, per a tot $X \in \mathbf{X}$ (i.e. $\mathbf{Z} \cap \text{De}_{\mathbf{X}}^{\mathcal{G}} = \emptyset$).
- Per a tot $X \in \mathbf{X}$ i $Y \in \mathbf{Y}$, $\mathbf{Z}$ *d*-separa tots els *back-door paths* entre X i Y ¹⁰.

Nota 2.30. Una manera pràctica de comprovar la segona condició és eliminar les arestes dirigides que surten de $\mathbf{X}$ i mirar si $\mathbf{Z}$ bloqueja tots els camins restants entre $\mathbf{X}$ i $\mathbf{Y}$. Aquesta idea tornarà a aparèixer al *do-calculus* (cf. 2.37).

Teorema 2.31 (*Back-door adjustment* [Pea93]; [Par23, Sec. 1.1.4, Th. 1]). Donada una CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*)$, sigui $P(\mathcal{V}) \in \mathcal{P}_*$ la distribució observacional. Considerem tres subconjunts de variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts dos a dos. Si $\mathbf{Z}$ satisfà el criteri back-door respecte a $(\mathbf{X}, \mathbf{Y})$, llavors l'efecte causal de $\mathbf{X}$ en $\mathbf{Y}$ és:

$$P(\mathbf{Y} \mid \text{do}(\mathbf{x})) = \mathbb{E}_{\mathbf{Z}}[P(\mathbf{Y} \mid \mathbf{x}, \mathbf{Z})] = \sum_{\mathbf{z}} p(\mathbf{z}) \cdot P(\mathbf{Y} \mid \mathbf{x}, \mathbf{z}). \quad (2.2)$$

¹⁰ Segons [Pea93], un *back-door path* o *camí de porta enrere* és un camí entre X i Y que conté una aresta entrant a X .

¹¹ L'expressió amb $\mathbb{E}_{\mathbf{Z}}[P(\mathbf{Y} \mid \mathbf{x}, \mathbf{Z})]$ unifica la notació dels casos discret i continu. L'última igualtat mostra l'expansió discreta. Si $\mathbf{Z}$ fos contínua, caldria escriure $\int P(\mathbf{Y} \mid \mathbf{x}, \mathbf{z}) dP_{\mathbf{Z}}(\mathbf{z})$ respecte de la llei observacional marginal de $\mathbf{Z}$. Fixem-nos també que

Observació 2.32. Cal destacar que, en CBNs Markovianes (és a dir, sense confusors latents), la distribució $P(Y \mid \text{do}(x))$ és identificable per a qualsevol tractament X i resultat Y mitjançant la factorització truncada de 2.18. En el cas particular d'un tractament $X = \{X\}$ i d'un resultat $Y \subseteq \text{De}_X^{\mathcal{G}} \setminus \{X\}$, també es pot obtenir per (2.2) prenent $Z = \mathcal{V} \setminus (\text{De}_X^{\mathcal{G}} \cup \{X\})$, el conjunt de variables no descendents de X . Aquest conjunt bloqueja tots els camins de porta enrere de X cap a Y , perquè aquests camins comencen per un pare no descendent de X .

Nota 2.33. Aquest fet enllaça amb el model de resultats potencials del capítol següent. En particular, [Pea09, Sec. 11.3.2] presenta el criteri *back-door* com una manera gràfica de justificar l'admissibilitat que, en el llenguatge de resultats potencials, s'expressa mitjançant la *ignorabilitat forta* o la no-confusió condicional. Així, quan un conjunt Z satisfà 2.29, la fórmula (2.2) és l'estimand d'ajustament associat a aquesta ignorabilitat, que en [RR83] i [IR15, Sec. 12.2] s'ha de formular com una hipòtesi sobre el mecanisme d'assignació i les covariants de pretractament (cf. 3.45).

Exemple 2.34 (Tabac i càncer de pulmó). Als anys cinquanta, diversos estudis (per exemple, [DH50; WG50]) van començar a mostrar una correlació massiva entre fumar tabac (T) i el càncer de pulmó (C). La majoria de metges van assumir, doncs, que $T \rightarrow C$. Ronald Fisher, el pare de l'estadística actual, era un fumador empedreït i feia de consultor per a les tabaqueres: va proposar que un cert factor genètic G feia de *confounder*, és a dir, que era el responsable tant de l'addicció al tabac com del càncer de pulmó [Fis58] (per exemple, G podria ser un gen que afectés el sistema de recompensa del cervell, fent que es fos més propens a fumar i també a desenvolupar càncer de pulmó [PM18]).

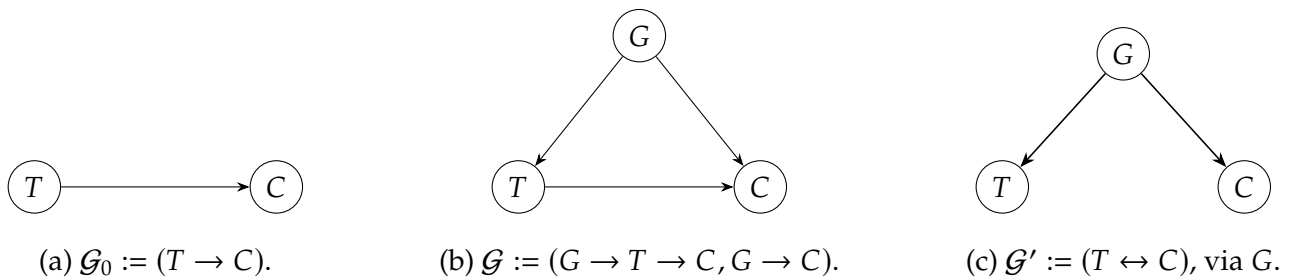


Figura 2.2: Xarxes bayesianes on es mostren possibles relacions entre el gen G , el tabaquisme T i el càncer de pulmó C . En (a), T té un efecte causal directe sobre C , mentre que en (b) no.

per aïllar l'efecte d'una intervenció i eliminar la influència de variables de confusió, la comunitat científica recorre als assajos controlats aleatoritzats (*Randomized Controlled Trials*, RCTs)¹². Amb les eines de què disposava Fisher, doncs, l'única manera de fer un experiment completament rigorós era agafar una mostra de persones a l'atzar i obligar-los a fumar durant dècades per observar-ne conseqüències, la qual cosa és impossible, i observacionalment no hi ha cap manera de distingir entre els dos models¹³, ja que G és una variable latent.

necessitem $p(x, z) > 0$ per a tot $z \in \mathcal{Z}$ tal que $p(z) > 0$.

¹² Els tractarem en profunditat més endavant, però la idea és senzilla: es divideixen els participants a l'atzar en diferents grups, cadascun sotmès a unes condicions o tractaments diferents. La robustesa d'aquest mètode recau en el seu disseny: gràcies a l'assignació aleatòria, les característiques individuals i els factors externs (tant els coneguts com els ocults) es distribueixen de manera equitativa entre tots els grups.

¹³ Aquesta idea no és pas nova, ja l'hem vist a 2.8: les dues xarxes bayesianes de la figura 2.2 són observacionalment equivalents.

Ara bé, amb el teorema que acabem d'introduir, això encara no esdevé possible: no podem ajustar per G i, per tant, no podem aplicar *back-door*. Ho resoldrem amb el següent criteri.

Definició 2.35 (Criteri de la porta endavant (*front-door*) [Par23, Sec. 1.1.4]). Donada $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana, sigui $P(\mathcal{V}) \in \mathcal{P}_*$ la distribució observacional. Considerem $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts dos a dos. Diem que $\mathbf{Z}$ satisfà el criteri de la porta endavant respecte a $(\mathbf{X}, \mathbf{Y})$ si, per a tot $X \in \mathbf{X}$ i $Y \in \mathbf{Y}$, es compleixen les següents condicions:

1. $\mathbf{Z}$ intercepta tots els camins dirigits de X a Y .
2. Tots els *back-door paths* de X a $\mathbf{Z}$ estan bloquejats per $\emptyset$.
3. Tots els *back-door paths* de $\mathbf{Z}$ a Y estan bloquejats per X .

Teorema 2.36 (*Front-door adjustment*, [Pea93]; [Par23, Sec. 1.1.4, Th. 2]). Donada $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana, amb distribució observacional $P(\mathcal{V}) \in \mathcal{P}_*$, i tres subconjunts de variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ disjunts dos a dos, si $\mathbf{Z}$ satisfà el criteri de la porta endavant si per a tot $\mathbf{x}$ tal que $p(\mathbf{x}) > 0$ i $p(\mathbf{x}', \mathbf{z}) > 0$ i tot $\mathbf{z}$ tal que $p(\mathbf{z} | \mathbf{x}) > 0$, llavors:

$$P(\mathbf{Y} | \text{do}(\mathbf{x})) = \mathbb{E}_{\mathbf{Z}|\mathbf{x}} [\mathbb{E}_{\mathbf{X}}[P(\mathbf{Y} | \mathbf{X}, \mathbf{Z})]] \stackrel{\text{disc.}}{=} \sum_{\mathbf{z}} p(\mathbf{z} | \mathbf{x}) \sum_{\mathbf{x}'} P(\mathbf{Y} | \mathbf{x}', \mathbf{z}) p(\mathbf{x}').$$

Teorema 2.37 (*do-calculus*, [Pea09, Sec. 3.4.2, Th. 3.4.1]; [Par23, Sec. 1.1.4, Th. 3]). Sigui $(\mathcal{G}, \mathcal{P}_*)$ una CBN, Markoviana o semi-Markoviana, sobre $\mathcal{V}$, i siguin $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \subseteq \mathcal{V}$ conjunts disjunts de variables, amb $\mathbf{X}$ i $\mathbf{W}$ possiblement buits. Aleshores, les distribucions de $\mathcal{P}_*$ satisfan les regles següents respecte de $\mathcal{G}$:

(R1) *Inserció i eliminació d'observacions*: si $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{X}}}} \mathbf{Z} | \mathbf{X}, \mathbf{W}$, aleshores:

$$P(\mathbf{Y} | \text{do}(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{Y} | \text{do}(\mathbf{x}), \mathbf{w}).$$

(R2) *Intercanvi entre intervenció i observació*: si $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{X}}, \mathbf{Z}}} \mathbf{Z} | \mathbf{X}, \mathbf{W}$, aleshores:

$$P(\mathbf{Y} | \text{do}(\mathbf{x}), \text{do}(\mathbf{z}), \mathbf{w}) = P(\mathbf{Y} | \text{do}(\mathbf{x}), \mathbf{z}, \mathbf{w}).$$

(R3) *Inserció/eliminació d'intervencions*: si $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{X}}, \mathbf{Z}}(\mathbf{W})} \mathbf{Z} | \mathbf{X}, \mathbf{W}$, aleshores:

$$P(\mathbf{Y} | \text{do}(\mathbf{x}), \text{do}(\mathbf{z}), \mathbf{w}) = P(\mathbf{Y} | \text{do}(\mathbf{x}), \mathbf{w}),$$

on $\mathbf{Z}(\mathbf{W}) := \mathbf{Z} \setminus \text{An}_{\mathbf{W}}^{\mathcal{G}_{\bar{\mathbf{X}}}}$. És a dir, el subconjunt de $\mathbf{Z}$ que no són avantpassats de cap node de $\mathbf{W}$ en el graf $\mathcal{G}_{\bar{\mathbf{X}}}$.

Nota 2.38. Aquestes són les regles sintàctiques del *do-calculus*. En models causals semi-Markovians recursius amb variables de domini finit, [SP06] prova que aquestes regles, juntament amb manipulacions probabilístiques estàndard, són completes per identificar efectes del tipus $p(\mathbf{y} | \text{do}(\mathbf{x}))$. [SP08, Sec. 3] reprèn aquest resultat i l'estén als efectes condicionals $p(\mathbf{y} | \mathbf{z}, \text{do}(\mathbf{x}))$ mitjançant l'algorisme IDC.

Exemple 2.39 (Tabac i càncer de pulmó, continuació). Per aplicar el criteri de la porta endavant, introduïm una variable observable Q que representa el quitrà acumulat als pulmons. Considerem un SCM semi-Markovià acíclic $\mathcal{M}$ amb $\mathcal{V} = \{T, Q, C\}$ i un confusor latent $G \in \mathcal{U}$ de T i C . El graf estructural endogen és $\mathcal{G}_{\mathcal{M}}^{\rightarrow} = (T \rightarrow Q \rightarrow C)$; el graf augmentat $\mathcal{G}_{\mathcal{M}}^+$ afegeix $G \rightarrow T$ i $G \rightarrow C$; i el diagrama causal

semi-Markovià $\mathcal{G}_M$, la projecció latent sobre $\mathcal{V}$, substitueix G per l'aresta $T \leftrightarrow C$. Treballem amb la CBN $(\mathcal{G}_M, \mathcal{P}_*)$; per a les comprovacions gràfiques escriurem simplement $\mathcal{G} := \mathcal{G}_M^+$.

Assumim que no hi ha efecte directe $T \rightarrow C$, que no hi ha confusió latent entre T i Q , i que la confusió entre Q i C queda bloquejada per T . Aleshores $\{Q\}$ satisfà el criteri *front-door* respecte de (T, C) :

- intercepta l'únic camí dirigit de T a C en $\mathcal{G}_M$,
- no hi ha cap camí *back-door* obert entre T i Q ; en $\mathcal{G}$, el camí $T \leftarrow G \rightarrow C \leftarrow Q$ queda bloquejat pel col·lisionador C , i
- tots els camins *back-door* entre Q i C queden bloquejats per T ; en $\mathcal{G}$, el camí $Q \leftarrow T \leftarrow G \rightarrow C$ queda bloquejat en condicionar per T .

Assumim també la positivitats necessària: en el cas discret, $p(t) > 0$ i $p(t', q) > 0$ per als q amb $p(q | t) > 0$ i els t' amb $p(t') > 0$. Per 2.36, l'efecte causal de fumar tabac sobre el càncer de pulmó és identificable dins d'aquesta CBN. Derivem ara $p(c | \text{do}(t))$: ho farem mitjançant el do-calculus, seguint [PM18, p. 236]. La figura 2.3 mostra els dos grafs causals que farem servir i la figura 2.4, els grafs mutilats.

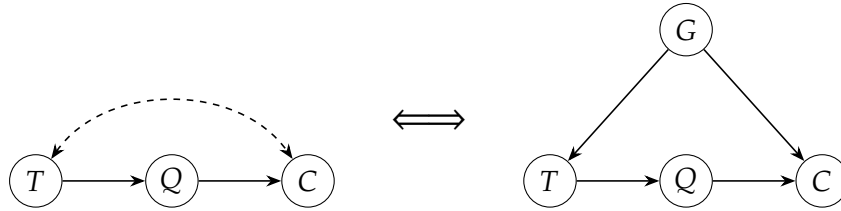


Figura 2.3: Diagrama causal semi-Markovià $\mathcal{G}_M$ sobre $\{T, Q, C\}$ (esquerra) i graf augmentat $\mathcal{G}_M^+$ restringit a G, T, Q, C (dreta) per al mateix SCM.

Comprovarem les d -separacions a $\mathcal{G}$, que és un DAG, per evitar introduir separació en ADMGs. Per fer explícites les marginalitzacions, desenvolupem la derivació en el cas discret; la formulació general és l'esperança iterada de 2.36.

Comencem aplicant la llei de la probabilitat total sobre Q , mantenint la intervenció $\text{do}(t)$ com a condició de fons:

$$p(c | \text{do}(t)) = \sum_q p(c, q | \text{do}(t)) = \sum_q p(c | \text{do}(t), q) p(q | \text{do}(t)). \quad (2.3)$$

Intercanvi d'observació per intervenció al primer factor. Per al terme $p(c | \text{do}(t), q)$, (R2) permet substituir q per $\text{do}(q)$ si $C \perp_{\mathcal{G}_{T, Q}} Q | T$. Aquesta condició es compleix en el graf de la figura 2.4a (hem eliminat les arestes que entren a T i les que surten de Q); per tant,

$$p(c | \text{do}(t), q) = p(c | \text{do}(t), \text{do}(q)). \quad (2.4)$$

Intercanvi d'intervenció per observació al segon factor. Per al terme $p(q | \text{do}(t))$, tornem a aplicar (R2), aquesta vegada per substituir la intervenció sobre T per l'observació de t . Cal comprovar que $Q \perp_{\mathcal{G}_T} T$. En el graf de la figura 2.4b, l'aresta $T \rightarrow Q$ ha estat eliminada i el camí $T \leftarrow G \rightarrow C \leftarrow Q$ queda bloquejat pel col·lisionador no observat C . Així, $p(q | \text{do}(t)) = p(q | t)$, i substituint això i (2.4) a (2.3),

arribem a

$$p(c \mid \text{do}(t)) = \sum_q p(c \mid \text{do}(t), \text{do}(q)) p(q \mid t). \quad (2.5)$$

Eliminació de la intervenció sobre T. Un cop imposat Q per intervenció, la intervenció sobre T deixa de ser rellevant per a C . Per (R3), cal veure que $C \perp_{\mathcal{G}_{\overline{Q}, T}} T \mid Q$. La figura 2.4c mostra que T queda desconnectat de C en eliminar les arestes entrants a Q i a T . Per tant, $p(c \mid \text{do}(t), \text{do}(q)) = p(c \mid \text{do}(q))$. Introduint aquest resultat a (2.5), obtenim

$$p(c \mid \text{do}(t)) = \sum_q p(c \mid \text{do}(q)) p(q \mid t). \quad (2.6)$$

Desenvolupament del terme $p(c \mid \text{do}(q))$. Apliquem la llei de la probabilitat total sobre T :

$$p(c \mid \text{do}(q)) = \sum_{t'} p(c, t' \mid \text{do}(q)) = \sum_{t'} p(c \mid \text{do}(q), t') p(t' \mid \text{do}(q)). \quad (2.7)$$

Substitució del primer factor per observació. Per al terme $p(c \mid \text{do}(q), t')$, la regla (R2) permet substituir $\text{do}(q)$ per l'observació q si $C \perp_{\mathcal{G}_Q} Q \mid T$. En el graf $\mathcal{G}_Q$ de la figura 2.4d, $Q \rightarrow C$ desapareix i el camí alternatiu $Q \leftarrow T \leftarrow G \rightarrow C$ queda bloquejat perquè condicionem per T . Així,

$$p(c \mid \text{do}(q), t') = p(c \mid q, t'). \quad (2.8)$$

Eliminació del segon factor intervingut. Per al terme $p(t' \mid \text{do}(q))$, apliquem (R3). Hi ha prou amb comprovar que $T \perp_{\mathcal{G}_Q} Q$. En el graf de la figura 2.4e, l'aresta $T \rightarrow Q$ ha estat eliminada i Q queda d -separat de T . Concloem que

$$p(t' \mid \text{do}(q)) = p(t'). \quad (2.9)$$

Substituint (2.8) i (2.9) a (2.7), obtenim

$$p(c \mid \text{do}(q)) = \sum_{t'} p(c \mid q, t') p(t'). \quad (2.10)$$

Finalment, introduint (2.10) a (2.6), arribem a la fórmula *front-door*:

$$p(c \mid \text{do}(t)) = \sum_q p(q \mid t) \sum_{t'} p(c \mid q, t') p(t').$$

Aquesta expressió identifica l'efecte causal dins d'aquesta CBN. La conclusió descansa en hipòtesis fortes: l'efecte del tabac sobre el càncer queda completament mediat pel quitrà, no hi ha confusió latent entre el tabac i el quitrà, els factors no observats que afecten el càncer no afecten també el quitrà excepte indirectament a través del tabaquisme, i les probabilitats condicionals anteriors estan ben definides en el suport rellevant. L'exemple s'ha de llegir, més que com una demostració absoluta que Fisher s'equivocava, com una il·lustració del criteri *front-door*.

Nota 2.40. Les tres regles generalitzen idees que ja han aparegut: d -separació en grafs mutilats, la regla *do/see* de les CBNs Markovianes i la ignorabilitat condicional (cf. 2.33).

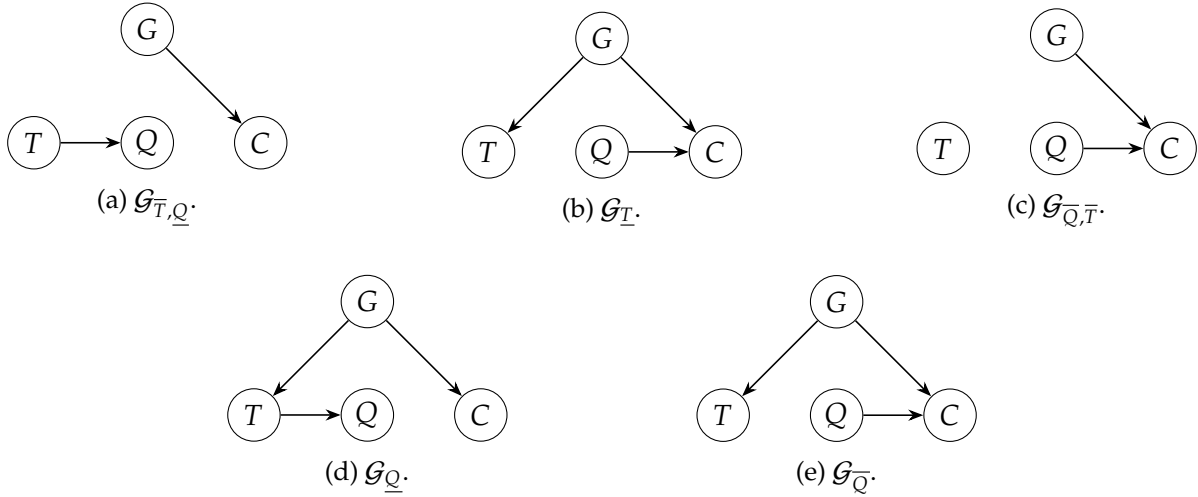


Figura 2.4: Subgrafs mutilats de $\mathcal{G} := \mathcal{G}_{\mathcal{M}}^+$ que justifiquen les aplicacions successives de les regles (R2) i (R3) en la derivació de l'ajustament *front-door* en l'exemple 2.39.

Definició 2.41 (Identificabilitat d'efectes causals [SP06, Def. 2]; [Bar+22, Sec. 27.4, Def. 27.16]).

Si $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V}$ són conjunts disjunts de variables, direm que l'efecte causal de $\text{do}(\mathbf{x})$ en $\mathbf{Y}$ donat $\mathbf{Z} = \mathbf{z}$, és a dir, la distribució $P(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{z})$, és *identificable* a partir de $P(\mathcal{V})$ en el diagrama causal $\mathcal{G}$ si, per a qualsevol parell de models causals $\mathcal{M}_1$ i $\mathcal{M}_2$ que indueixen CBNs amb el mateix diagrama causal $\mathcal{G}$ i amb la mateixa distribució observacional $P_{\mathcal{M}_1}(\mathcal{V}) = P_{\mathcal{M}_2}(\mathcal{V}) > 0$, tenim:

$$P_{\mathcal{M}_1}(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{z}) = P_{\mathcal{M}_2}(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{z}).$$

Així, una consulta causal és identificable si només depèn de l'estructura causal representada pel diagrama $\mathcal{G}$ i de la distribució $P(\mathcal{V})$. Quan això passa, existeix un *estimand* que expressa la consulta causal exclusivament en termes observacionals.

Nota 2.42. No s'ha de confondre amb la *identificabilitat* de l'estadística clàssica. Si $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, on Θ és l'espai de paràmetres, la família parametritzada de lleis $\mathcal{P}$ és identificable si per a qualsevol $\theta_1, \theta_2 \in \Theta$ amb $\theta_1 \neq \theta_2$, tenim $P_{\theta_1} \neq P_{\theta_2}$. És a dir, l'aplicació $\theta \mapsto P_\theta$ és injectiva.

III

Model de resultats potencials

*Per la manca d'un clau, es va perdre la ferradura
per la manca d'una ferradura, es va perdre el cavall
per la manca d'un cavall, es va perdre el genet
per la manca d'un genet, es va perdre la batalla
i així com la batalla, es va perdre un regne
i tot per la manca d'un clau de ferradura.*

George HERBERT, *Jacula Prudentum* (1651)

En aquest capítol presentem el model de resultats potencials per estudiar l'efecte causal d'un tractament sobre una variable Y . Començarem amb una població finita de n unitats, en el sentit de Neyman [Sp190], i després introduïrem el nivell de superpoblació quan calgui parlar d'esperances poblacionals. Escriurem $I_u = \{1, \dots, n\}$ per al conjunt d'unitats.

Notació 3.1 (Convenció notacional del capítol). Ara l'índex $i \in I_u$ recorre unitats i no variables d'un graf. El conjunt $I_u = \{1, \dots, n\}$ és una població o mostra finita d'unitats, i no s'ha de confondre amb l'índex I de les variables $\mathcal{V} = \{V_i\}_{i \in I}$. També reservarem $\mathbf{X}$ per a la matriu de covariants de les unitats; en aquest capítol, doncs, $\mathbf{X}$ no denota un conjunt de variables intervingudes.

3.1

EL PROBLEMA FONAMENTAL DE LA INFERÈNCIA CAUSAL

Fixem el llenguatge dels resultats potencials i el problema que el motiva: per a cada unitat només observem un dels resultats possibles.

Notació 3.2 (Tractaments i resultats potencials en població finita). En el nivell de població finita de Neyman, les covariants i els resultats potencials de les n unitats són fixos. L'aleatorietat prové només del mecanisme d'assignació, que introduïrem més endavant. Sigui $\mathcal{W}$ el conjunt de possibles tractaments, sigui $\mathcal{A}_n := \mathcal{W}^n$ l'espai d'assignacions per a les n unitats i sigui $\mathcal{Y} \subseteq \mathbb{R}$ un borelià de possibles resultats numèrics:

1. Un vector $\mathbf{w} = (w_1, \dots, w_n) \in \mathcal{A}_n$ és una assignació possible de tractaments a les n unitats.
2. Per a cada unitat $i \in I_u$ i cada assignació $\mathbf{w} \in \mathcal{A}_n$, $Y_i(\mathbf{w}) \in \mathcal{Y}$ és el *resultat potencial*: el resultat que obtindria la unitat i sota l'assignació $\mathbf{w}$. Escriurem $\mathbf{Y}(\mathbf{w}) := (Y_1(\mathbf{w}), \dots, Y_n(\mathbf{w}))$ per al vector de resultats potencials sota $\mathbf{w}$.
3. En el cas binari, $\mathcal{W} = \{0, 1\}$ i, per tant, $\mathcal{A}_n = \{0, 1\}^n$.

La llei del vector aleatori $\mathbf{W} = (W_1, \dots, W_n)$ quedarà determinada especificant una PMF sobre $\mathcal{A}_n$.

Abans d'estimar res, cal que els resultats potencials estiguin ben definits per a cada unitat i per a cada assignació.

Definició 3.3 (No-interferència entre unitats). Per a tota unitat $i \in I_u$ i tota parella d'assignacions de tractaments $\mathbf{w}, \mathbf{w}' \in \mathcal{A}_n$, si $w_i = w'_i$, aleshores $Y_i(\mathbf{w}) = Y_i(\mathbf{w}')$.

Exemple 3.4 (No-interferència). En un programa de vacunació dins d'una classe, els alumnes vacunats poden reduir la transmissió i protegir també els no vacunats (ja que disminueix la transmissió del virus o bacteri en qüestió). Això constitueix una violació de l'assumpció de no-interferència: l'estat resultant d'un alumne pot dependre del tractament rebut pels seus companys i no només del seu propi tractament.

Nota 3.5 (Spillover). Aquest tipus d'efecte indirecte (*immunitat de grup* o *herd immunity*) és un cas d'*spillover*. En l'exemple anterior és positiu: els no vacunats poden semblar més sans del que serien en una classe sense vacunació.

Definició 3.6 (Absència de múltiples versions). Suposem que cada nivell de tractament $a \in \mathcal{W}$ pot ser implementat de diverses maneres, recollides en un espai $\mathcal{W}_a$ d'implementacions del tractament a . Aleshores, el resultat de la unitat i sota el nivell a i la versió $v \in \mathcal{W}_a$ es pot escriure com $Y_i(a, v) \in \mathcal{Y}$. Aquesta hipòtesi estableix que $Y_i(a, v) = Y_i(a, v')$, per a tot $v, v' \in \mathcal{W}_a$.

Exemple 3.7. Suposem que el tractament és una operació quirúrgica. Si dues tècniques quirúrgiques poden produir resultats diferents, no haurien de quedar «amagades» sota el mateix tractament. A la pràctica, qualsevol tractament presenta variacions en alguns aspectes; la qüestió rellevant és, però, decidir quines variacions són prou rellevants per definir tractaments diferents.

Aquestes dues assumpcions constitueixen el que Rubin va anomenar *Stable Unit Treatment Value Assumption* (SUTVA) a [Rub74].

Definició 3.8 (SUTVA). Se satisfà SUTVA si es compleixen simultàniament 3.3 i 3.6. Aleshores, podem escriure $Y_i(\mathbf{w}) = Y_i(w_1, \dots, w_n) = Y_i(w_i)$. SUTVA garanteix, doncs, que els $Y_i(w_i)$ estan ben definits i no depenen de l'assignació de tractament de les altres unitats.

Un cop dut a terme l'estudi, denotem el resultat observat per $Y_i^{\text{obs}} \in \mathcal{Y}$. La igualtat amb el resultat potencial corresponent és l'assumpció de *consistència* [IR15, Sec. 1.6].

Definició 3.9 (Consistència). Un cop fixat un mecanisme d'assignació, sigui $\mathbf{W} = (W_1, \dots, W_n)$ el vector aleatori de tractaments rebuts. Se satisfà la consistència si, per a tota assignació $\mathbf{w} \in \mathcal{A}_n$, el fet que $\mathbf{W} = \mathbf{w}$ implica $Y_i^{\text{obs}} = Y_i(\mathbf{w})$ per a tot $i \in I_u$. És a dir, si el vector d'assignacions observat és $\mathbf{w}$, el resultat observat de cada unitat coincideix amb el seu resultat potencial sota aquella assignació.

Nota 3.10. Sota SUTVA, si el tractament és binari, la consistència permet escriure $Y_i^{\text{obs}} = Y_i(W_i)$. Així doncs, d'ara en endavant suposarem el tractament binari, tret que s'indiqui el contrari. En el nivell finit només necessitem l'espai d'assignacions $\mathcal{A}_n$ i, quan estudiem el mecanisme d'assignació, una PMF sobre aquest espai.

Definició 3.11 (Resultats potencials observats i no observats [IR15, Secs. 1.7 i 2.6]). Sota tractament binari, SUTVA i consistència, el vector $\mathbf{Y}^{\text{obs}} = (Y_i^{\text{obs}})_{i \in I_u}$ és el vector de resultats observats. Component a

component, tenim:

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0), & \text{si } W_i = 0; \\ Y_i(1), & \text{si } W_i = 1. \end{cases} \iff Y_i^{\text{obs}} = W_i Y_i(1) + (1 - W_i) Y_i(0).$$

De manera anàloga, $\mathbf{Y}^{\text{mis}} = (Y_i^{\text{mis}})_{i \in I_u}$ representa els resultats potencials que falten:

$$Y_i^{\text{mis}} = Y_i(1 - W_i) \iff \begin{cases} Y_i(1), & \text{si } W_i = 0; \\ Y_i(0), & \text{si } W_i = 1. \end{cases} \iff Y_i^{\text{mis}} = (1 - W_i) Y_i(1) + W_i Y_i(0).$$

Nota 3.12 (Convenció per als resultats observats). Quan no hi hagi risc d'ambigüitat, escriurem $Y_i := Y_i^{\text{obs}}$ per al resultat observat de la unitat i , i $\mathbf{Y} := \mathbf{Y}^{\text{obs}} = (Y_1, \dots, Y_n)$ per al vector de resultats observats.

Observació 3.13 ([IR15, Secs. 1.7 i 2.6]). Podem invertir aquestes relacions i caracteritzar els resultats potencials en termes dels resultats observats i no observats:

$$Y_i(0) = \begin{cases} Y_i^{\text{mis}}, & \text{si } W_i = 1; \\ Y_i^{\text{obs}}, & \text{si } W_i = 0. \end{cases} \quad Y_i(1) = \begin{cases} Y_i^{\text{mis}}, & \text{si } W_i = 0; \\ Y_i^{\text{obs}}, & \text{si } W_i = 1. \end{cases}$$

Exemple 3.14 (de l'aspirina). Suposem que tenim una població formada per un conjunt d' n individus amb mal de cap.

1. El tractament w consisteix en prendre una aspirina o no en un cert interval de temps; per tant, el tractament és binari, i.e. $\mathcal{W} = \{0, 1\}$.
2. $Y_i(1)$ i $Y_i(0)$ representen els resultats potencials sota tractament amb aspirina o control de la unitat i , respectivament.
3. $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$ tenen per components el mal de cap de cada pacient sota el control i sota l'efecte de l'aspirina, respectivament.

El nostre objectiu és determinar si l'efecte de l'aspirina sobre el conjunt de pacients és causal o no; l'efecte causal del tractament, doncs, serà una comparació entre $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$.

Nota 3.15 (Problema fonamental de la inferència causal, [Hol86]). El problema fonamental de la inferència causal succeeix perquè, independentment de com definim l'efecte causal, solament podem observar un dels dos resultats potencials per a cada unitat de tractament.

$\mathbf{X}$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\dots$	$\mathbf{X}_n$
$\mathbf{Y}(0)$	$Y_1(0)$	?	?	$Y_4(0)$	?	?	$Y_7(0)$	?	$\dots$	?
$\mathbf{Y}(1)$	?	$Y_2(1)$	$Y_3(1)$	?	$Y_5(1)$	$Y_6(1)$	?	$Y_8(1)$	$\dots$	$Y_n(1)$
$\mathbf{Y}(0)$ vs. $\mathbf{Y}(1)$	?	?	?	?	?	?	?	?	$\dots$	?

Taula 3.1: Resultats potencials: observats i contrafactuals. Font: [Rub05].

Per tant, almenys la meitat¹⁴ de tots els resultats potencials estaran no observats o absents. Per estimar l'efecte causal d'una unitat concreta, caldria conèixer el resultat potencial que no s'ha observat (cf. 3.11). Com que això no és possible directament, la inferència depèn d'hipòtesis sobre l'assignació del tractament i sobre la comparabilitat entre unitats. Aquí entren les característiques prèvies de les unitats, o *covariants*, que corresponen a la primera fila de la taula 3.1.

Definició 3.16 (Covariant [IR15, Sec. 1.8]). Sigui I_u el conjunt d'unitats. Una *covariant* (o *covariable*) per a la unitat $i \in I_u$, denotada X_i , és un vector de variables observades abans de l'assignació del tractament que descriu característiques prèvies de la unitat i i que no estan afectades pel tractament. Les covariants serveixen per formular hipòtesis sobre el mecanisme d'assignació, estudiar efectes en subgrups i millorar la precisió de les estimacions.

Exemple 3.17 (de l'aspirina, continuació). En l'estudi de l'aspirina, una covariant podria ser la intensitat del dolor abans de decidir si es pren el tractament. Com que aquesta dada es recull *a priori*, aquesta dada no està afectada per l'aspirina i pot servir per:

1. comparar individus amb nivells inicials de dolor semblants, per reduir les diferències de base entre els pacients;
2. estudiar si l'efecte varia segons la intensitat inicial, utilitzant aquesta intensitat com a covariant;
3. reduir el biaix que apareix amb les persones amb més dolor, ja que també són les més propenses a prendre aspirina.

Notació 3.18 (Matriu de covariants, [IR15, Secs. 1.8 i 1.10]). Suposem que tenim $K \in \mathbb{N}$ covariants per a cada unitat. Per a $j = 1, \dots, K$, sigui $\mathcal{X}_{\text{cov},j}$ l'espai de valors possibles de la j -èsima covariant. Definim l'espai de covariants com el producte $\mathcal{X}_{\text{cov}} := \prod_{j=1}^K \mathcal{X}_{\text{cov},j}$. Així, per a cada unitat $i \in I_u$, el seu vector de covariants és $X_i = (X_{i1}, \dots, X_{iK}) \in \mathcal{X}_{\text{cov}}$, on $X_{ij} \in \mathcal{X}_{\text{cov},j}$ és el valor de la j -èsima covariant de la unitat i . Denotem per $\mathbf{X} := (X_{ij})_{i \in I_u, j=1, \dots, K}$ la matriu $n \times K$ de covariants de la mostra finita. Equivalentment, podem escriure $\mathbf{X} = (X_i)_{i \in I_u}$, entenent que cada fila X_i és el vector de covariants de la unitat i -èsima.

EFECTE CAUSAL I ESTIMANDS

La secció distingeix l'efecte individual dels efectes mitjans i condicionals, i mostra que cada estimand causal és una comparació entre resultats potencials.

Abans d'escollir la metodologia d'estimació que utilitzarem, cal fixar quina magnitud causal es vol estimar. Aquestes magnituds són els *estimands*¹⁵: funcions dels resultats potencials de la taula 3.1.

Nota 3.19 (Efecte causal, [IR15, Secs. 1.3 i 1.10]). Per poder parlar d'*estimand*, la comparació causal s'ha de fer entre resultats potencials de les mateixes unitats: $\{Y_i(0) \mid i \in I_u\}$ i $\{Y_i(1) \mid i \in I_u\}$, no entre subconjunts diferents¹⁶.

¹⁴ Si hi ha ℓ tractaments possibles, la proporció de no observats és de $\frac{\ell-1}{\ell} \geq \frac{1}{2}$ si $\ell \geq 2$.

¹⁵ Aquest terme no existeix en català [07] i traduccions com *estimacions* o *estimadors* són imprecises. El mantindrem en anglès.

¹⁶ Per exemple, comparar directament la nota d'un alumne que va a classes de reforç amb la d'un alumne excel·lent que no hi va no identifica, per si sol, l'efecte causal de les classes de reforç. Hi ha una sèrie de suposicions injustificades al darrere.

La forma més directa d'efecte causal és la diferència entre els dos resultats potencials d'una unitat.

Definició 3.20 (Efecte causal individual, [IR15, Sec. 1.10]). Assumim que $\mathcal{W} = \{0, 1\}$ (cf. 3.10). Podem definir l'efecte causal individual (o l'efecte unitari del tractament) com:

$$\tau_i = Y_i(1) - Y_i(0).$$

Nota 3.21. A causa del problema fonamental de la inferència causal (cf. 3.15), la taula $(Y_i(0), Y_i(1))_{i \in I_u}$ dels resultats potencials no es pot inferir directament a partir de les dades. Anàlogament, la distribució empírica dels efectes individuals τ_i , 3.20, no és identificable sense assumpcions addicionals.

L'efecte mitjà del tractament (ATE, *Average Treatment Effect*) es pot identificar en certes situacions. Existeixen diversos tipus d'ATE segons el conjunt d'unitats sobre el qual es fa la mitjana.

Definició 3.22 (SATE, [Ima13, Sec. 1.3, p. 3]). En el nivell finit, l'efecte mitjà del tractament en la mostra (SATE, *sample average treatment effect*) és la mitjana dels efectes causals individuals de les n unitats observades:

$$\text{SATE} = \frac{1}{n} \sum_{i=1}^n (Y_i(1) - Y_i(0)).$$

Definició 3.23 (SATT, [Ima13, Sec. 1.3, p. 3]). Un cop feta l'assignació del tractament, l'efecte mitjà del tractament sobre els tractats en la mostra (SATT, *Sample Average Treatment Effect for the Treated*) és la mitjana dels efectes causals individuals restringida a les unitats que van rebre el tractament:

$$\text{SATT} = \frac{1}{\sum_{i=1}^n W_i} \sum_{i=1}^n W_i (Y_i(1) - Y_i(0)), \quad \text{si } \sum_{i=1}^n W_i > 0.$$

Aquesta quantitat és d'especial interès en l'avaluació de programes o polítiques, ja que sovint resulta més rellevant analitzar l'impacte sobre les unitats que s'hi han inscrit o participat en lloc de la mostra completa (especialment si els grups de tractament i control ja presenten prèviament diferències significatives).

Per definir estimands poblacionals introduïm un segon nivell probabilístic. En el nivell de superpoblació, les unitats observades es veuen com una mostra i.i.d. d'una distribució conjunta sobre els resultats potencials i les covariants.

Notació 3.24 (Superpoblació). Sigui a partir d'ara els espais de Borel estàndard $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, $(\mathcal{X}_{\text{cov}}, \mathcal{B}(\mathcal{X}_{\text{cov}}))$ i $(\mathcal{W}, 2^{\{0,1\}})$, on $2^{\{0,1\}}$ és la σ -àlgebra discreta de $\mathcal{W} = \{0, 1\}$.

Considerem un espai de probabilitat $(\Omega_{\text{sp}}, \mathcal{F}_{\text{sp}}, \mathbb{P}_{\text{sp}})$ i una unitat genèrica descrita pel vector aleatori

$$\Phi_{\text{sp}} := (Y(0), Y(1), \mathbf{X}_{\text{sp}}) : (\Omega_{\text{sp}}, \mathcal{F}_{\text{sp}}) \longrightarrow (\mathcal{Z}_{\text{sp}}, \mathcal{B}(\mathcal{Z}_{\text{sp}})),$$

on

$$\mathcal{Z}_{\text{sp}} := \mathcal{Y}^2 \times \mathcal{X}_{\text{cov}}, \quad \mathcal{B}(\mathcal{Z}_{\text{sp}}) := \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{X}_{\text{cov}}).$$

Aquest requisit és més subtil del que podria semblar a primera vista.

Aquí $\mathbf{X}_{\text{sp}}$ és el vector de covariants d'una unitat genèrica; en canvi, $\mathbf{X} = (\mathbf{X}_i)_{i \in I_u}$ continua denotant la matriu finita de covariants quan parlem de les n unitats observades.

Denotem per $P_{\text{sp}} := \mathbb{P}_{\text{sp}} \circ \Phi_{\text{sp}}^{-1}$ la llei conjunta d'aquest vector sobre $(\mathcal{Z}_{\text{sp}}, \mathcal{B}(\mathcal{Z}_{\text{sp}}))$ i per $\mathbb{E}_{\text{sp}}$ l'esperança corresponent. Per a lleis marginals escriurem, per exemple, $P(\mathbf{X}_{\text{sp}})$ o $P(Y(0), Y(1), \mathbf{X}_{\text{sp}})$. Escriurem

$$\mathcal{S}_{\mathbf{X}} := \{\mathbf{x} \in \mathcal{X}_{\text{cov}} \mid P_{\mathbf{X}_{\text{sp}}}(\mathcal{U}) > 0\}, \quad \text{i} \quad \mathcal{S}_{Y(0), Y(1), \mathbf{X}} := \{(y_0, y_1, \mathbf{x}) \in \mathcal{Y}^2 \times \mathcal{X}_{\text{cov}} \mid P_{Y(0), Y(1), \mathbf{X}_{\text{sp}}}(\mathcal{V}) > 0\},$$

per a tot obert $\mathcal{U} \subseteq \mathcal{X}_{\text{cov}}$ amb $\mathbf{x} \in \mathcal{U}$ i per a tot obert $\mathcal{V} \subseteq \mathcal{Y}^2 \times \mathcal{X}_{\text{cov}}$ amb $(y_0, y_1, \mathbf{x}) \in \mathcal{V}$ (és a dir, els suports de les lleis marginals corresponents). En el cas discret, aquestes definicions es redueixen als valors amb probabilitat positiva. Quan considerem una mostra aleatòria de n unitats d'aquesta superpoblació, escriurem $(Y_i(0), Y_i(1), \mathbf{X}_i)$, $i \in I_u$, per a còpies i.i.d. de Φ_{sp} ; la llei de la mostra completa és $P_{\text{sp}}^{\otimes n}$. Un cop obtinguda la mostra finita, el vector d'assignacions es genera separatament mitjançant $\mathbf{W} \sim \pi(\cdot \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$. Si pel context queda clar, ometrem el subíndex sp i escriurem simplement P i $\mathbb{E}$.

Així doncs, en la superpoblació, podem interpretar els resultats potencials i les covariants com a variables aleatòries, definir lleis imatge i treballar amb distribucions condicionals regulars.

Definició 3.25 (PATE i PATT). L'efecte poblacional mitjà del tractament (PATE, *Population Average Treatment Effect*) es defineix com l'esperança matemàtica incondicional de l'efecte causal individual:

$$\text{PATE} = \mathbb{E}_{\text{sp}}[Y(1) - Y(0)].$$

Un cop especificat un règim d'assignació o observació que defineix W i estén la llei de la unitat genèrica, l'efecte poblacional mitjà del tractament sobre els tractats (PATT, *Population Average Treatment Effect for the Treated*) restringeix l'esperança a la subpoblació amb $W = 1$:

$$\text{PATT} = \mathbb{E}_{\text{sp}}[Y(1) - Y(0) \mid W = 1].$$

Nota 3.26 (Mostra finita i població). El SATE o el SATT són estimands de mostra finita i fan referència a les unitats concretes observades en l'estudi. A diferència del SATE, el SATT depèn també del vector d'assignacions realitzat, perquè es restringeix a les unitats tractades.

Introduïm ara els estimands poblacionals, com el PATE, el PATT o el CATE, que s'expressen mitjançant esperances superpoblacionals.

Definició 3.27 (CATE). L'efecte mitjà condicional del tractament (CATE, *Conditional Average Treatment Effect*) descriu l'heterogeneïtat de l'efecte causal en funció de $\mathbf{X}_{\text{sp}}$:

$$\text{CATE}(\mathbf{x}) = \mathbb{E}_{\text{sp}}[Y(1) - Y(0) \mid \mathbf{X}_{\text{sp}} = \mathbf{x}].$$

El CATE permet determinar de quina manera característiques prèvies dels individus alteren la intensitat de l'efecte causat per la intervenció.

Els efectes mitjans no sempre capturen tota la informació rellevant de les intervencions. Per analitzar l'impacte d'un tractament en certs extrems de la distribució (com, per exemple, els salaris més baixos

d'una població), es poden definir altres quantitats d'interès, com els efectes causals sobre els quantils:

Definició 3.28 (QTE, [Ima13, Sec. 1.3, p. 3]). L'efecte causal sobre el quantil $q \in (0, 1)$ (QTE, *Quantile Treatment Effect*) es defineix com:

$$\text{QTE}_q = q_1(q) - q_0(q), \quad q_w(q) = \inf\{y \in \mathbb{R} \mid P_{\text{sp}}(Y(w) \leq y) \geq q\},$$

on $q_w(q)$ és el quantil q de la distribució del resultat potencial $Y(w)$ per a $w \in \{0, 1\}$. Aquesta quantitat mesura com el tractament afecta la posició d'un cert quantil q en la distribució dels resultats.

Observació 3.29. Alternativament, podem expressar el QTE com:

$$\text{QTE}_q = F_{Y(1)}^{-1}(q) - F_{Y(0)}^{-1}(q),$$

on $F_{Y(1)}$ i $F_{Y(0)}$ són les funcions inverses generalitzades de les funcions de distribució acumulada marginals dels resultats potencials $Y(1)$ i $Y(0)$.

3.3

MECANISME D'ASSIGNACIÓ DE TRACTAMENTS

En aquesta secció estudiarem com es decideix qui rep el tractament. L'objectiu és formalitzar el mecanisme d'assignació, connectar-lo amb els resultats observats i introduir la puntuació de propensió.

Nota 3.30 (Mostra de superpoblació i assignació finita). La notació de 3.24 descriu una mostra i.i.d. de la llei d'una unitat genèrica abans de l'assignació. En aquesta secció tornem al nivell de població finita: condicionats a $\mathbf{X}$, $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$, el vector d'assignacions $\mathbf{W} = (W_i)_{i \in I_u}$ pot tenir una llei conjunta sobre $\mathcal{A}_n$. Per tant, les components W_i no han de ser independents; per exemple, un disseny completament aleatoritzat amb un nombre fix de tractats imposa restriccions globals sobre $\mathbf{W}$ [IR15, Secs. 3.3 i 3.6].

Un cop fixat un mecanisme d'assignació, escriurem $n_c = \sum_{i=1}^n (1 - W_i)$ i $n_t = \sum_{i=1}^n W_i$ per al nombre d'unitats de control i tractades de l'assignació realitzada, respectivament, de manera que $n_c + n_t = n$. La notació de resultats observats i no observats ja ha quedat fixada a 3.11. Formalitzem ara l'assignació de tractament com la funció que assigna probabilitats als 2^n valors que pot prendre $\mathbf{W}$.

Definició 3.31 (Mecanisme d'assignació [IR15, Sec. 3.3, Def. 3.1]). Donada la matriu finita de covariants $\mathbf{X}$ i els vectors fixos de resultats potencials $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$, el mecanisme d'assignació és la PMF condicionada sobre l'espai d'assignacions $\mathcal{A}_n = \{0, 1\}^n$:

$$\begin{aligned} \pi(\cdot \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) : \mathcal{A}_n &\longrightarrow [0, 1] \\ \mathbf{w} &\longmapsto \pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \end{aligned}$$

per a tot $\mathbf{X}$, $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$. Per definició de PMF:

$$\sum_{\mathbf{w} \in \mathcal{A}_n} \pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1.$$

Per a valors fixats de $\mathbf{X}$, $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$, aquesta PMF és la distribució de disseny de l'assignació sobre $(\mathcal{A}_n, 2^{\mathcal{A}_n})$. Respecte d'aquesta mesura, podem identificar $\mathbf{W}$ amb el vector aleatori $\mathbf{W}(\mathbf{w}) = \mathbf{w}$.

Seguint Imbens i Rubin [IR15, Sec. 3.3], imposarem també que el mecanisme sigui invariant sota permutacions simultànies de les components de $\mathbf{w}$, $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$ i de les files de $\mathbf{X}$.

Observació 3.32.

- $\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ és la probabilitat que el vector aleatori d'assignacions $\mathbf{W}$ prengui el valor concret $\mathbf{w}$, no la probabilitat que una unitat concreta rebi el tractament.
- En superpoblació teníem còpies i.i.d. de $(Y(0), Y(1), \mathbf{X}_{\text{sp}})$. Aquí, $\mathbf{W} = (W_i)_{i \in I_u}$ pot tenir una llei conjunta $\pi(\cdot \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ que no és producte de les marginals.
- El model permet que determinats vectors d'assignació tinguin una probabilitat de zero¹⁷. També és comú restringir el disseny perquè, si n és parell, exactament la meitat de la mostra estigui tractada, és a dir, que $\sum_{i=1}^n W_i = \frac{n}{2}$.

Definició 3.33 (Probabilitat d'assignació individual, [IR15, Sec. 3.3, Def. 3.2]). La probabilitat d'assignació individual d'una unitat i -èsima és la probabilitat marginal que aquesta unitat sigui assignada al tractament, condicionada a les covariants i als resultats potencials de totes les unitats:

$$\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \sum_{\substack{\mathbf{w} \in \mathcal{A}_n \\ w_i = 1}} \pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)).$$

Sumem les probabilitats de tots els vectors d'assignació $\mathbf{w}$ on $w_i = 1$. De tots els 2^n vectors possibles, 2^{n-1} (la meitat) compleixen aquesta condició. La probabilitat de ser assignada al control és $1 - \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$.

Nota 3.34. Imbens–Rubin escriuen habitualment p_i per a aquesta probabilitat individual. Aquí adaptem la convenció a π_i per no confondre-la amb la notació p que, des de 1.2, queda reservada per a masses o densitats avaluades.

A la pràctica, imposarem restriccions sobre aquesta distribució. Aquestes restriccions determinen quan les comparacions condicionals poden rebre una interpretació causal. La primera és la individualitat: la probabilitat marginal d'assignació d'una unitat s'ha de poder descriure amb la informació d'aquella unitat, encara que la PMF conjunta incorpori restriccions globals del disseny.

Definició 3.35 (Mecanisme individualístic, [IR15, Sec. 3.4, Def. 3.4]). El mecanisme d'assignació és *individualístic* si existeix una funció $q : \mathcal{X}_{\text{cov}} \times \mathcal{Y}^2 \rightarrow [0, 1]$ i un conjunt admissible $\mathcal{D} \subseteq \mathcal{A}_n \times \mathcal{X}_{\text{cov}}^n \times \mathcal{Y}^n \times \mathcal{Y}^n$ tals que, posant $q_i := q(\mathbf{X}_i, Y_i(0), Y_i(1))$, tenim $\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q_i$ per a tot $i \in I_u$ i, a més,

$$\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} c(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \prod_{i=1}^n q_i^{w_i} (1 - q_i)^{1-w_i}, & (\mathbf{w}, \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \in \mathcal{D}, \\ 0, & \text{altrament.} \end{cases}$$

¹⁷ Per exemple, en un assaig clínic, podríem voler evitar escenaris on ningú rebi el medicament. En aquest cas, assignariem una probabilitat de 0 al vector on tots els valors de W_i siguin zero.

La constant de normalització queda determinada per:

$$c(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))^{-1} = \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{A}_n \\ (\tilde{\mathbf{w}}, \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \in \mathcal{D}}} \prod_{i=1}^n q_i^{\tilde{w}_i} (1 - q_i)^{1 - \tilde{w}_i},$$

sempre que aquesta suma sigui estrictament positiva. El conjunt $\mathcal{D}$ recull les restriccions admissibles del disseny.

Observació 3.36 (Conjunt admissible). Quan les restriccions només afecten el vector d'assignació, podem identificar $\mathcal{D}$ amb un subconjunt $\mathbf{D} \subseteq \mathcal{A}_n$. En un assaig de Bernoulli, $\mathbf{D} = \mathcal{A}_n$, $c = 1$ i les assignacions són independents [IR15, Sec. 4.3]. En un experiment completament aleatoritzat amb n_t tractats,

$$\mathbf{D} = \left\{ \mathbf{w} \in \mathcal{A}_n : \sum_{i=1}^n w_i = n_t \right\}.$$

Si $q_i = \frac{n_t}{n}$ per a tot i , el producte anterior és constant sobre $\mathbf{D}$ i s'obté $\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \binom{n}{n_t}^{-1}$ per a $\mathbf{w} \in \mathbf{D}$ i zero fora de $\mathbf{D}$ [IR15, Secs. 3.6 i 4.4]. En aquest cas les variables W_i no són independents, tot i que $\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \frac{n_t}{n}$ per a tota unitat. Si els q_i són heterogenis i es trunca un producte Bernoulli a un conjunt $\mathbf{D}$ amb restriccions globals, les marginals resultants no coincideixen necessàriament amb els q_i ; això només és individualístic en el sentit estricte de Imbens i Rubin [IR15, Def. 3.4] si també es compleix $\pi_i = q_i$.

Nota 3.37. Els experiments *seqüencials* violen aquest supòsit perquè, en aquests dissenys, l'assignació de les unitats tractades en una etapa posterior pot dependre dels resultats observats de les unitats assignades anteriorment [IR15, Sec. 3.6, p. 41].

La segona condició és que cada unitat tingui probabilitat estrictament positiva de ser assignada tant al tractament com al control. En cap cas, doncs, la probabilitat ha de ser 0 o 1, ja que això implicaria que en algunes unitats no es podria comparar el fet de rebre tractament o no.

Definició 3.38 (Mecanisme probabilístic). Un mecanisme d'assignació és probabilístic si la probabilitat d'assignació de cada unitat està estrictament entre 0 i 1, és a dir, si per a tot $i \in I_u$ i per a tot $\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)$ es compleix que $0 < \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) < 1$.

La tercera restricció és que el mecanisme conjunt no depengui dels resultats potencials.

Definició 3.39 (Assignació no confosa en població finita, [IR15, Sec. 3.4]). El mecanisme d'assignació és *no-confós* si:

$$\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \pi(\mathbf{w} \mid \mathbf{X}, \tilde{\mathbf{Y}}(0), \tilde{\mathbf{Y}}(1)),$$

per a tot $\mathbf{w} \in \mathcal{A}_n$, per a tot $\mathbf{X}$ i per a tots els valors possibles de $\mathbf{Y}(0), \mathbf{Y}(1), \tilde{\mathbf{Y}}(0)$ i $\tilde{\mathbf{Y}}(1)$. Quan es compleix, denotarem el valor comú per $\pi(\mathbf{w} \mid \mathbf{X})$ i escriurem $\pi_i(\mathbf{X})$ per a la probabilitat marginal d'assignació de la unitat i .

Les probabilitats marginals d'assignació també es poden resumir per valors de covariants. Aquesta funció és el que coneixem com a *valor de propensió*.

Definició 3.40 (Valor de propensió per a poblacions finites [IR15, Sec. 3.3]). El *valor de propensió* (*propensity score*) en $\mathbf{x}$ és:

$$e_{\text{fin}}(\mathbf{x}) = \frac{1}{N(\mathbf{x})} \sum_{i|\mathbf{X}_i=\mathbf{x}} \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)),$$

on $N(\mathbf{x}) = \#\{i \in \{1, \dots, n\} \mid \mathbf{X}_i = \mathbf{x}\}$. Per als valors de $\mathbf{x}$ tals que $N(\mathbf{x}) = 0$, definim $e_{\text{fin}}(\mathbf{x}) = 0$.

Nota 3.41. Imbens i Rubin [IR15, Def. 3.3] emprà la notació $e(\mathbf{x})$ per a aquesta quantitat. Aquí escrivim $e_{\text{fin}}(\mathbf{x})$ per distingir-la del valor de propensió de superpoblació.

Proposició 3.42. Si el mecanisme d'assignació és individualístic i no-confós en el sentit de 3.39, aleshores el valor de propensió finit coincideix amb la probabilitat individual d'assignació per a qualsevol unitat amb aquell valor de covariants. És a dir, si $\mathbf{X}_i = \mathbf{x}$, $e_{\text{fin}}(\mathbf{x}) = \pi_i(\mathbf{X})$.

Quan considerem un mecanisme d'assignació unitari de superpoblació, aquest afegeix un tractament W a la unitat genèrica i estén la llei P_{sp} . Com que treballem amb espais de Borel estàndard, fixem una versió regular de la distribució condicional de W donat $(Y(0), Y(1), \mathbf{X}_{\text{sp}})$ i escrivim:

$$\pi_{\text{sp}}(w \mid y_0, y_1, \mathbf{x}) := P_{\text{sp}}(W = w \mid Y(0) = y_0, Y(1) = y_1, \mathbf{X}_{\text{sp}} = \mathbf{x}).$$

Com és habitual, aquesta probabilitat condicional només queda determinada llevat de conjunts nuls respecte de la llei marginal $P_{Y(0), Y(1), \mathbf{X}_{\text{sp}}}$ (és a dir, dues versions poden diferir en un conjunt $N \subseteq \mathcal{Y}^2 \times \mathcal{X}_{\text{cov}}$ tal que $P_{Y(0), Y(1), \mathbf{X}_{\text{sp}}}(N) = 0$). Aquesta notació no imposa que, en una mostra finita, les assignacions W_i siguin independents. Quan existeixen les densitats o masses corresponents respecte d'una mesura dominant producte, Imbens i Rubin [IR15, Sec. 3.5] escriuen la llei d'una unitat genèrica amb assignació en la forma:

$$f_{Y(0), Y(1), W, \mathbf{X}_{\text{sp}}}(y_0, y_1, w, \mathbf{x}; \phi, \theta, \psi) = f_{W|Y(0), Y(1), \mathbf{X}_{\text{sp}}}(w \mid y_0, y_1, \mathbf{x}; \phi) f_{Y(0), Y(1)|\mathbf{X}_{\text{sp}}}(y_0, y_1 \mid \mathbf{x}; \theta) f_{\mathbf{X}_{\text{sp}}}(\mathbf{x}; \psi),$$

on ϕ, θ, ψ són els paràmetres del mecanisme d'assignació, de la distribució dels resultats potencials condicionada a les covariants i de la distribució de les covariants, respectivament. Com que W és binari, el primer factor de la dreta és una massa condicional i, amb la versió regular fixada més amunt,

$$\pi_{\text{sp}}(w \mid y_0, y_1, \mathbf{x}) = f_{W|Y(0), Y(1), \mathbf{X}_{\text{sp}}}(w \mid y_0, y_1, \mathbf{x}; \phi), \quad w \in \{0, 1\}.$$

Quan no sigui rellevant, suprimirem aquests paràmetres de la notació. En aquesta formulació de superpoblació, un cop fixats els paràmetres del model, el mecanisme és individualístic: la probabilitat d'assignació d'una unitat depèn només de les seves variables $(Y(0), Y(1), \mathbf{X}_{\text{sp}})$.

Definició 3.43 (Valor de propensió de superpoblació, [IR15, Sec. 3.5, Def. 3.7]). El *valor de propensió de superpoblació* en $\mathbf{x}$ és la mitjana poblacional de la probabilitat d'assignació unitària entre les unitats amb covariants $\mathbf{X}_{\text{sp}} = \mathbf{x}$:

$$e_{\text{sp}}(\mathbf{x}) := \mathbb{E}_{\text{sp}}[\pi_{\text{sp}}(1 \mid Y(0), Y(1), \mathbf{X}_{\text{sp}}) \mid \mathbf{X}_{\text{sp}} = \mathbf{x}] = P_{\text{sp}}(W = 1 \mid \mathbf{X}_{\text{sp}} = \mathbf{x}).$$

El terme $\pi_{\text{sp}}(1 \mid Y(0), Y(1), \mathbf{X}_{\text{sp}})$ és el mateix que $f_{W|Y(0), Y(1), \mathbf{X}_{\text{sp}}}(1 \mid Y(0), Y(1), \mathbf{X}_{\text{sp}}; \phi)$, amb el paràmetre

ϕ implícit. Quan no hi hagi risc de confusió, escriurem simplement $e(\mathbf{x})$ en lloc d' $e_{\text{sp}}(\mathbf{x})$.

Definició 3.44 (Assignació probabilística en superpoblació, [IR15, Sec. 3.5, Def. 3.8]). És *probabilístic* si

$$0 < \pi_{\text{sp}}(1 \mid y_0, y_1, \mathbf{x}) < 1$$

per a $P_{Y(0), Y(1), \mathbf{X}_{\text{sp}}}$ -gairebé tots els valors de $(y_0, y_1, \mathbf{x})$. Si es fixa una versió concreta del nucli condicional definida en tot $\mathcal{S}_{Y(0), Y(1), \mathbf{X}}$, la lectura punt a punt d'aquesta desigualtat s'ha d'entendre només per als $(y_0, y_1, \mathbf{x}) \in \mathcal{S}_{Y(0), Y(1), \mathbf{X}}$; fora d'aquest suport, la versió del condicional és arbitrària.

Definició 3.45 (Assignació no confosa en superpoblació, [IR15, Sec. 3.5, Def. 3.9]). El mecanisme d'assignació de superpoblació és *no-confós* si:

$$(Y(0), Y(1)) \perp\!\!\!\perp_{P_{\text{sp}}} W \mid \mathbf{X}_{\text{sp}}.$$

Equivalentment, per a una versió regular de les probabilitats condicionades, la probabilitat d'assignació només depèn de les covariants:

$$\pi_{\text{sp}}(w \mid y_0, y_1, \mathbf{x}) = \pi_{\text{sp}}(w \mid \mathbf{x})$$

per a $w \in \{0, 1\}$ i per a $P(Y(0), Y(1), \mathbf{X}_{\text{sp}})$ -gairebé tot $(y_0, y_1, \mathbf{x})$.

Definició 3.46 (Positivitat o superposició). La condició de *positivitat* o *superposició* exigeix que $0 < e_{\text{sp}}(\mathbf{x}) < 1$ per a $P_{\mathbf{X}_{\text{sp}}}$ -gairebé tot $\mathbf{x}$. Si es formula punt a punt, la restricció s'entén sobre $\mathcal{S}_{\mathbf{X}}$, el suport de la llei marginal de les covariants. En el cas discret, això equival a exigir-ho per a tot valor de covariant amb probabilitat positiva. Sota assignació no confosa en superpoblació, aquesta condició coincideix amb l'assignació probabilística de superpoblació.

Definició 3.47 (Assignació fortament ignorable, [RR83, Sec. 1.3]). L'assignació donades les covariants observades $\mathbf{X}_{\text{sp}}$ és *fortament ignorable* si es compleixen simultàniament:

$$(Y(0), Y(1)) \perp\!\!\!\perp_{P_{\text{sp}}} W \mid \mathbf{X}_{\text{sp}} \quad \text{i} \quad 0 < P_{\text{sp}}(W = 1 \mid \mathbf{X}_{\text{sp}} = \mathbf{x}) < 1$$

per a $P_{\mathbf{X}_{\text{sp}}}$ -gairebé tot $\mathbf{x}$; en una formulació punt a punt, per a tot $\mathbf{x} \in \mathcal{S}_{\mathbf{X}}$.

Nota 3.48. 3.39 és una propietat de la PMF d'assignació sobre les n unitats fixes. En canvi, 3.45 és la versió unitària de superpoblació de Imbens i Rubin [IR15, Sec. 3.5]. A Imbens i Rubin [IR15, Sec. 12.2.2], aquesta segona formulació s'escriu també com una independència condicional. El pas d'una formulació a l'altra requereix les hipòtesis de mostreig i d'assignació corresponents.

Observació 3.49. Aquesta condició és molt potent per dues raons diferents:

- Exigeix que el conjunt de covariants $\mathbf{X}$ sigui suficient per controlar la confusió rellevant per a l'assignació.
- D'una banda, en el marc dels resultats potencials, la no-confusió s'exigeix com *una hipòtesi* sobre el mecanisme d'assignació del tractament. D'altra banda, en el marc de Pearl, la independència condicional corresponent queda garantida (cf. 2.33) si $\mathbf{X}$ satisfà el criteri de la porta del darrere respecte del tractament i el resultat (*back-door criterion*, cf. 2.29 i 2.31).

Nota 3.50. La no-confusió no és *contrastable* directament a partir de les dades observades, perquè imposa una restricció sobre resultats potencials que, per definició, no observem simultàniament. Depèn del coneixement del problema, del disseny i del fet d’haver mesurat les covariants que poden influir tant en l’assignació del tractament com en el resultat.

Tot i que la hipòtesi d’ignorabilitat forta és molt convenient per a l’anàlisi contrafactual, Pearl [Pea09, Sec. 11.3] observa que, per si sola, no determina per quines covariants cal ajustar. Afegir covariants sense analitzar-ne el paper causal pot introduir nous biaixos.

Definició 3.51 (Regularitat en població finita). Un mecanisme d’assignació de població finita és regular si:

1. és individualístic,
2. és probabilístic i
3. és no-confós en el sentit de 3.39.

Si el mecanisme no és regular, direm que és *irregular*. Nosaltres ens centrem en mecanismes regulars.

En els experiments aleatoritzats clàssics (i, en particular, en els experiments completament aleatoritzats), aquestes condicions provenen del disseny: l’investigador controla l’assignació i pot garantir que el mecanisme sigui probabilístic, individualístic i no-confós [HR20, Cap. 2]; [IR15, Secs. 3.6 i 4.4]. En estudis observacionals, les hipòtesis de superpoblació s’han de justificar a partir de les covariants disponibles i del coneixement que tinguem del problema [HR20, Cap. 3]; [IR15, Sec. 12.2].

3.4

IDENTIFICABILITAT PER AJUSTAMENT

Quan consistència, ignorabilitat i positivitat es compleixen, alguns estimands causals es poden escriure amb quantitats observacionals.

Definició 3.52 (Identificabilitat, [HR20, Fine Point 3.1, p. 29]). Direm que un estimand causal és *identificable* si, un cop imposades aquestes assumpcions, la distribució de les dades observades determina un únic valor possible per a l’estimand. Equivalentment, l’estimand pot escriure’s com una funció de la distribució observacional.

Proposició 3.53 (Identificació per ajustament o estandardització, [HR20, Sec. 2.3 i Technical Point 4.1]). Definim, per a $w \in \{0, 1\}$,

$$m_w(\mathbf{x}) = \mathbb{E}_{\text{sp}} [Y^{\text{obs}} \mid W = w, \mathbf{X}_{\text{sp}} = \mathbf{x}] .$$

Aquesta quantitat s’entén definida on el condicionament és ben definit; en particular, quan $P_{\text{sp}}(W = w \mid \mathbf{X}_{\text{sp}} = \mathbf{x}) > 0$. Sota consistència, ignorabilitat (o intercanviabilitat) condicional i positivitat, el PATE i el PATT són identificables a partir de la distribució observacional mitjançant:

$$\text{PATE} = \mathbb{E}_{\text{sp}} [m_1(\mathbf{X}_{\text{sp}}) - m_0(\mathbf{X}_{\text{sp}})] , \quad \text{PATT} = \mathbb{E}_{\text{sp}} [Y^{\text{obs}} - m_0(\mathbf{X}_{\text{sp}}) \mid W = 1] .$$

Sota aquestes hipòtesis, les comparacions entre unitats tractades i no tractades amb covariants semblants

poden rebre una interpretació causal. En experiments aleatoritzats, la validesa ve imposada pel mecanisme d'assignació; en estudis observacionals, en canvi, cal defensar-la mitjançant covariants, ignorabilitat i positivitat. La construcció d'estimadors i variàncies es desenvolupa al capítol 4.

3.5

PEARL I NEYMAN–RUBIN

Connectem els resultats potencials amb les intervencions en SCMs i amb els axiomes contrafactuals de Galles–Pearl.

En aquesta secció, $\mathcal{V}$ tornarà a denotar el conjunt de variables endògenes d'un SCM $\mathcal{M}$. Quan calgui tractar els contrafactuals com a variables aleatòries, farem servir l'espai probabilístic subjacent $(\Omega, \mathcal{F}, \mathbb{P})$. Les intervencions s'escriuran $\text{do}(\mathbf{A} = \mathbf{a})$.

En la formulació de població finita de Neyman (*design-based inference*), els resultats potencials $Y_i(w)$ són fixos i la incertesa prové exclusivament del fet que l'investigador (o la natura) assigna els tractaments aleatòriament.

El marc causal estructural de Pearl aborda l'aleatorietat des del determinisme latent (*model-based inference*, capítol 4). En un SCM, les funcions estructurals són fixes i la incertesa neix de la llei $P(\mathcal{U})$ sobre els contextos exògens. Si prenem un vector aleatori $\mathbf{U} : (\Omega, \mathcal{F}) \rightarrow (\mathcal{U}, \mathcal{B}(\mathcal{U}))$, aleshores $\mathbf{u} = \mathbf{U}(\omega)$, $\omega \in \Omega$ és la realització exògena corresponent.

Per connectar rigorosament ambdós marcs, podem representar una «unitat física» $i \in I_u$ del model de Neyman mitjançant un estat $\omega \in \Omega$ o, equivalentment, mitjançant $\mathbf{u} = \mathbf{U}(\omega)$. Un cop fixat l'estat ω , el resultat potencial fix $Y_i(w)$ s'identifica amb la solució de l'equació estructural per a $\mathbf{u}$, que denotarem per $Y_w(\omega)$.

Dit això, els axiomes introduïts per [GP98, Secs. 2–4] descriuen el comportament contrafactual dels models causals estructurals. En el cas recursiu (quan el graf associat és un DAG), permeten connectar la semàntica intervencional dels SCMs amb la notació de resultats potencials de Neyman–Rubin. Ens interessaran sobretot tres propietats: *composició*, *efectivitat* i *reversibilitat*.

Notació 3.54 (Notació global). Sigui $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$ un SCM i $\mathcal{G}_{\mathcal{M}}$ el diagrama causal induït per $\mathcal{M}$ en el sentit de 2.4 i 2.5. El conjunt $\mathcal{U}$ conté els factors exògens, el conjunt $\mathcal{V} = \{V_i\}_{i \in I}$, amb $I = \{1, \dots, m\}$, conté les variables endògenes i per a cada $V_i \in \mathcal{V}$ tenim una funció mesurable:

$$f_i : \mathcal{X}_{\text{Pa}_i^{\mathcal{G}_{\mathcal{M}}}} \times \mathcal{U}_i^c \rightarrow \mathcal{X}_i, \quad v_i = f_i(\text{pa}_i, \mathbf{u}_i^c).$$

Posem $\mathcal{U} := \prod_{U \in \mathcal{U}} \mathcal{U}_U$. Prendrem $P(\mathcal{U})$ sobre $(\Omega, \mathcal{F}, \mathbb{P})$, escriurem $\mathbf{U} = (U)_{U \in \mathcal{U}}$ per al vector aleatori exogen amb llei $P(\mathcal{U})$ i, per a cada estat $\omega \in \Omega$, $\mathbf{u} := \mathbf{U}(\omega)$. Anàlogament, $\mathbf{U}_i^c$ denota el vector de factors exògens associats a V_i i $\mathbf{u}_i^c := \mathbf{U}_i^c(\omega)$. Aleshores, $\mathcal{M}$ induïx una distribució $P(\mathcal{V})$ sobre $\mathcal{V}$ (cf. 1.45).

Notació 3.55 (Submodel i resultat potencial, [Pea09, Def. 7.1.2]; [GP98, Defs. 2.2–2.4]). Donats $\mathbf{A} \subseteq \mathcal{V}$ i un valor $\mathbf{a} \in \mathcal{X}_{\mathbf{A}}$, el submodel $\mathcal{M}_{\mathbf{a}}$ s'obté substituint les equacions de les variables de $\mathbf{A}$ per les equacions

constants $\mathbf{A} = \mathbf{a}$ (cf. 2.9). Per a qualsevol $Y \in \mathcal{V}$, definim primer la resposta estructural

$$\tilde{Y}_{\mathbf{a}} : \mathcal{U} \longrightarrow \mathcal{X}_Y,$$

on $\tilde{Y}_{\mathbf{a}}(\mathbf{u})$ és la component corresponent a Y de l'única solució del submodel $\mathcal{M}_{\mathbf{a}}$ quan $\mathbf{u} \in \mathcal{U}$. La variable aleatòria que farem servir nosaltres és la composició següent:

$$Y_{\mathbf{a}} := \tilde{Y}_{\mathbf{a}} \circ \mathbf{U} \iff Y_{\mathbf{a}}(\omega) = \tilde{Y}_{\mathbf{a}}(\mathbf{U}(\omega)).$$

Quan no fem cap intervenció, escriurem simplement Y , és a dir, $Y_{\emptyset} = Y$.

En aquest treball mantindrem ω com a argument dels contrafactuals. [GP98] escriu sovint $Y_{\mathbf{a}}(\mathbf{u})$; amb la convenció anterior, aquesta notació correspon a $\tilde{Y}_{\mathbf{a}}(\mathbf{u})$, i només suprimirem la titlla quan reprenguem explícitament la sintaxi formal de la font.

Definició 3.56 (Enunciat contrafactual i conseqüència semàntica). Un enunciat contrafactual elemental és una fórmula de la forma $Y_{\mathbf{a}}(\omega) = y$, on $Y \in \mathcal{V}$, $\mathbf{A} \subseteq \mathcal{V}$, $\mathbf{a} \in \mathcal{X}_{\mathbf{A}}$, $\omega \in \Omega$ i $y \in \mathcal{X}_Y$. Equivalentment, $\tilde{Y}_{\mathbf{a}}(\mathbf{U}(\omega)) = y$. Prenent $\mathbf{u} \in \mathcal{U}$, podem escriure l'enunciat estructural com $\tilde{Y}_{\mathbf{a}}(\mathbf{u}) = y$, o simplement $Y_{\mathbf{a}}(\mathbf{u}) = y$ quan recuperem la notació de [GP98].

- Donat un conjunt $\mathbf{S}$ d'enunciats contrafactuals, denotem per $\mathcal{M}_{\mathbf{S}}$ el conjunt de tots els SCMs que satisfan $\mathbf{S}$. Direm que $\mathbf{S}$ *deriva semànticament* un enunciat σ , i ho escriurem $\mathbf{S} \models \sigma$, si σ se satisfà per tots els models de $\mathcal{M}_{\mathbf{S}}$.
- Donat un conjunt d'axiomes $\mathbf{A}$, direm que $\mathbf{S}$ *deriva sintàcticament* σ , i ho escriurem $\mathbf{S} \vdash_{\mathbf{A}} \sigma$, si σ s'obté de $\mathbf{S}$ mitjançant aplicacions repetides dels axiomes de $\mathbf{A}$.

Nota 3.57 (Lectura estructural dels contrafactuals). Hi ha dos casos que convé tenir presents:

1. Si $Y \in \mathcal{V}$ i $\mathbf{A} = \mathcal{V} \setminus \{Y\}$, aleshores $Y_{\mathbf{a}}(\omega) = f_Y(\mathbf{a}_{\text{Pa}_Y}, \mathbf{U}_Y(\omega))$, on $\mathbf{a}_{\text{Pa}_Y}$ indica que els pares de Y , Pa_Y , han rebut els valors indicats per $\mathbf{a}$. En aquest sentit, cada funció estructural pot llegir-se com un *contrafactual* (el terme que utilitza Pearl per a referir-se als resultats potencials).
2. Si $Y \in \mathbf{A}$ i la intervenció $\mathbf{A} = \mathbf{a}$ imposa $Y = y$, aleshores $Y_{\mathbf{a}}(\omega) = y$ per a tot $\omega \in \Omega$. El contrafactual coincideix, per tant, amb el valor imposat per la intervenció.

Observació 3.58 (Connexió amb els resultats potencials). Quan $\mathbf{W}$ recull les variables de tractament d'un SCM i Y és el resultat d'una unitat, $Y_{\mathbf{w}}$ és l'anàleg de $Y_i(\mathbf{w})$ en la notació de Neyman–Rubin. Aquí $\mathbf{W}$ ja no és el vector finit d'assignacions, sinó el bloc de variables sobre el qual intervenim dins de l'SCM.

COMPOSICIÓ, EFECTIVITAT I REVERSIBILITAT

Els axiomes de composició, efectivitat i reversibilitat descriuen què es conserva quan manipulem intervencions dins d'un SCM.

Un cop interpretat $Y_{\mathbf{a}}$ en un SCM, podem preguntar-nos quines lleis satisfan aquests termes contrafactuals. Introduïrem tres axiomes que són fonamentals, i veurem que són satisfets per qualsevol SCM amb solució única.

Enunciarem les propietats següents en la nostra notació. La forma estructural s'obté substituint $Y_a(\omega)$ per $\tilde{Y}_a(\mathbf{U}(\omega))$; si es fixa directament un context exogen $\mathbf{u}$, coincideixen amb la notació habitual de Galles–Pearl basada en expressions del tipus $Y_a(\mathbf{u})$.

Propietat 3.59 (Composició, [GP98, Sec. 3]). *Siguin $Y, W \in \mathcal{V}$ i sigui $\mathbf{A} \subseteq \mathcal{V} \setminus \{W\}$. Aleshores,*

$$W_a(\omega) = w \implies Y_{a,w}(\omega) = Y_a(\omega).$$

Nota 3.60 (Interpretació de la composició). La composició diu que si, sota la intervenció $\text{do}(\mathbf{A} = \mathbf{a})$, la variable W ja pren el valor w , aleshores afegir la intervenció $\text{do}(W = w)$ no altera la resposta de Y . En termes d'equacions estructurals, passar de $\mathcal{M}_a$ a $\mathcal{M}_{a,w}$ només substitueix l'equació de W per una constant que ja era el seu valor en $\mathcal{M}_a$.

Corol·lari 3.61 (Consistència). *Per a qualsevol $Y, X \in \mathcal{V}$ i qualsevol valor $x \in \mathcal{X}_X$,*

$$X(\omega) = x \implies Y(\omega) = Y_x(\omega).$$

És la versió contrafactual de la consistència introduïda a 3.9: si el vector d'assignacions observat és $\mathbf{w} \in \mathcal{A}_n$, el resultat observat coincideix amb el resultat potencial corresponent, $Y_i^{\text{obs}} = Y_i(\mathbf{w})$.

En efecte, és el cas particular de la composició amb $\mathbf{A} = \emptyset$ i $W = X$.

Propietat 3.62 (Efectivitat, [GP98, Sec. 3]). *Per a tot $X \in \mathcal{V}$, tot $W \in \mathcal{V} \setminus \{X\}$ i tots els valors $x \in \mathcal{X}_X$ i $w \in \mathcal{X}_W$,*

$$X_{x,w}(\omega) = x.$$

L'efectivitat garanteix que la variable manipulada pren el valor imposat per la intervenció: si forcem $\text{do}(X = x)$, aleshores $X = x$ en el submodel intervingut.

Observació 3.63. La formulació anterior és la forma singular de [GP98, Sec. 3]. La definició de submodel de [GP98, Sec. 2, Defs. 2 i 4] dona immediatament la versió per a intervencions sobre un conjunt: si $\mathbf{A} \subseteq \mathcal{V}$, $\mathbf{a} \in \mathcal{X}_\mathbf{A}$ i $A_j \in \mathbf{A}$, aleshores

$$(A_j)_\mathbf{a}(\omega) = a_j.$$

Farem servir aquesta forma quan la intervenció imposi valors a més d'una variable.

Propietat 3.64 (Reversibilitat, [GP98, Sec. 3]). *Siguin $Y, W \in \mathcal{V}$ dues variables diferents i sigui $\mathbf{A} \subseteq \mathcal{V} \setminus \{Y, W\}$. Aleshores,*

$$(Y_{a,w}(\omega) = y) \wedge (W_{a,y}(\omega) = w) \implies Y_a(\omega) = y.$$

En SCMs recursius, composició i efectivitat, juntament amb definitud, unicitat i ordenació causal, donen una axiomatització completa del comportament contrafactual.

Teorema 3.65. *La propietat 3.59 es compleix en tot SCM en què els submodels $\mathcal{M}_a$ i $\mathcal{M}_{a,w}$ que intervenen en l'enunciat admeten una única solució.*

Teorema 3.66. *La propietat 3.62 es compleix en tot SCM.*

Teorema 3.67. La propietat 3.64 es compleix en tot SCM en què els submodels $\mathcal{M}_a$, $\mathcal{M}_{a,w}$, $\mathcal{M}_{a,y}$ i $\mathcal{M}_{a,w,y}$ admeten una única solució.

Introduïm ara els conceptes de definitud, unicitat i ordenació causal, necessaris per caracteritzar el comportament contrafactual.

Definició 3.68 (Ordenació causal). Un *ordenament causal* de $\mathcal{V}$ és un ordre estricte $<$ sobre les variables, compatible amb la part dirigida del diagrama causal $\mathcal{G}_M$: si $V_j \in \text{Pa}_i^{\mathcal{G}_M}$, aleshores $V_j < V_i$. Quan les variables estan enumerades d'acord amb aquest ordre, això implica $j < i$.

Nota 3.69 ([GP98, Sec. 3, Def. 3.2]). Si $<$ és un ordre topològic sobre les variables i l'enumeració és compatible amb $<$, aleshores, per a $i < k$, intervenir sobre V_k no afecta V_i , ja que V_k no pot ser avantpassat de V_i :

$$(V_i)_{v_k, \mathbf{z}}(\omega) = (V_i)_{\mathbf{z}}(\omega),$$

per a tot $\mathbf{Z} \subseteq \mathcal{V} \setminus \{V_i, V_k\}$, tot $\mathbf{z} \in X_{\mathbf{Z}}$ i tot $v_k \in X_{V_k}$. Denotarem aquest conjunt de $\frac{m(m-1)}{2}$ enuncisats per CO (de *Causal Ordering*).

Propietat 3.70 (Definitud). Per a tota variable $Y \in \mathcal{V}$, $\mathbf{A} \subseteq \mathcal{V}$ i $\omega \in \Omega$, existeix un $y \in X_Y$ tal que $Y_{\mathbf{a}}(\omega) = y$.

Propietat 3.71 (Unicitat). Per a tota variable $Y \in \mathcal{V}$ i tot $\mathbf{A} \subseteq \mathcal{V}$, $Y_{\mathbf{a}}(\omega) = y \wedge Y_{\mathbf{a}'}(\omega) = y' \implies y = y'$.

Per exemple, no podem llançar una moneda, intervenir perquè caigui de cara i que el resultat sigui cara i creu simultàniament.

El resultat següent precisa fins a quin punt aquests axiomes capturen el comportament contrafactual dels SCMs recursius. Com és habitual en la formulació axiomàtica de Galles–Pearl i Halpern, el resultat s'entén sobre un llenguatge finit: hi intervenen finitament moltes variables endògenes, cada variable té un conjunt finit de valors possibles, i els contextos exògens que apareixen als enuncisats són els del fragment considerat.

Teorema 3.72. $A_C = \{(C), (E), \text{definitud}, \text{unicitat}, \text{CO}\}$ forma un conjunt complet d'axiomes per als sistemes causalment ordenats, respecte de conjunts finits d'enuncisats contrafactuals. És a dir,

$$\mathbf{S} \models \sigma \implies \mathbf{S} \vdash_{A_C} \sigma,$$

per a tot conjunt finit $\mathbf{S}$ d'enuncisats contrafactuals i tot enunciat σ sobre un sistema causalment ordenat.

Nota 3.73. [Hal00, Th. 3.2] va donar posteriorment una axiomatització completa per a la classe dels SCMs amb solució única (inclosos els no recursius). En particular, això precisa el resultat de Galles–Pearl: la composició, l'efectivitat i la reversibilitat formen part del sistema, però la completeness també fa servir els axiomes lògics, definitud i unicitat.

Nota 3.74 (Eliminació de contrafactuals niats [Bar+20, Th. 7]). També podem pensar en expressions on una intervenció depèn del valor que una altra variable hauria pres en un món intervingut. Aquests contrafactuals es poden reescriure sense intervencions niades: en el cas simple, $P(Y_{X_w} = y) = \sum_x P(Y_x = y, X_w = x)$. Més generalment, si $\mathbf{X}^*$ i $\mathbf{Z}^*$ són «intervencions funcionals naturals» sobre conjunts disjunts,

i $\mathbf{Y}$ és disjunt de les variables intervingudes, aleshores:

$$P(\mathbf{Y}_{\mathbf{Z}^*, \mathbf{X}^*} = \mathbf{y}) = \sum_{\mathbf{x}} P(\mathbf{Y}_{\mathbf{Z}^*, \mathbf{x}} = \mathbf{y}, \mathbf{X}^* = \mathbf{x}).$$

FRONT-DOOR PER AXIOMES CONTRAFACTUALS

Reconstruïm la fórmula front-door manipulant resultats potencials. L'objectiu és mostrar que els axiomes contrafactuals i el do-calculus arriben al mateix estimand.

Recuperem el cas del tabac, el quitrà i el càncer de pulmó en llenguatge contrafactual. Al capítol d'identificació vam obtenir la fórmula *front-door* amb el do-calculus; l'objectiu ara és veure que la mateixa expressió es pot reconstruir manipulant directament resultats potencials.

Exemple 3.75 (Tabac i càncer de pulmó, continuació). Continuem l'exemple del tabac introduït a 2.34 i desenvolupat amb el criteri *front-door* a 2.39, seguint ara [GP98, Sec. 5]. Introduïm Q per representar el quitrà acumulat als pulmons i considerem l'SCM $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$, amb $\mathcal{V} = \{T, Q, C\}$ i $\mathcal{U} = \{U_1, U_2\}$. Les funcions estructurals són:

$$T = f_T(U_1), \quad Q = f_Q(T, U_2), \quad C = f_C(Q, U_1), \quad U_1 \perp_{P(\mathcal{U})} U_2.$$

El graf augmentat $\mathcal{G}_{\mathcal{M}}^+$ conté les arestes $U_1 \rightarrow T$, $U_1 \rightarrow C$, $U_2 \rightarrow Q$, $T \rightarrow Q$ i $Q \rightarrow C$. En projectar sobre les variables endògenes, el diagrama causal semi-Markovià $\mathcal{G} = \mathcal{G}_{\mathcal{M}}$ satisfà $\text{Pa}_T^{\mathcal{G}} = \emptyset$, $\text{Pa}_Q^{\mathcal{G}} = \{T\}$ i $\text{Pa}_C^{\mathcal{G}} = \{Q\}$, i conté l'aresta bidirigida $T \leftrightarrow C$ induïda per U_1 . És el mateix escenari del criteri *front-door*: el tabac T afecta el càncer C a través del mediador Q , mentre que la confusió latent entre T i C queda governada per U_1 .

Que $T \rightarrow C \notin \mathcal{E}$ vol dir que, dins d'aquest model, no hi ha efecte directe de T sobre C un cop controlat Q : queda completament mediat pel quitrà Q . D'altra banda, $U_1 \perp_{P(\mathcal{U})} U_2$ ens diu que, encara que U_1 agreugi el risc de C , no hauria d'afectar Q un cop imposat T per intervenció. Així, U_2 recull la variació de Q que no prové d' U_1 .

L'objectiu amb contrafactuals no és descriure l'associació entre tabac i càncer, sinó comparar el risc de càncer fumant o sense fumar. Volem escriure el risc de càncer sota la intervenció $T = t$,

$$P(C = c \mid \text{do}(t)) = P(C_t = c)$$

com una expressió que només depengui de $P(\mathcal{V}) = P(T, Q, C)$. Farem servir dues regles de [GP98, Sec. 5].

1. **Restricció d'exclusió:** Per a tot $Y \in \mathcal{V}$, tot $\mathbf{Z} \subseteq \mathcal{V} \setminus (\{Y\} \cup \text{Pa}_Y^{\mathcal{G}})$ i tot $\mathbf{z} \in \mathcal{X}_{\mathbf{Z}}$, tenim $Y_{\text{pa}_Y} = Y_{\text{pa}_Y, \mathbf{z}}$, amb la convenció que, si $\text{Pa}_Y^{\mathcal{G}} = \emptyset$, aleshores $Y_{\text{pa}_Y} = Y$. Un cop imposats els valors dels pares causals de Y , intervenir sobre variables que no en són pares no en canvia el resultat.
2. **Restricció d'independència:** Siguin $Y \in \mathcal{V}$ i $\mathbf{Z} = \{Z_1, \dots, Z_k\}$ tals que no pertanyen a la mateixa

C-component¹⁸. Aleshores:

$$Y_{\text{pa}_Y} \perp\!\!\!\perp_P \{Z_{1\text{pa}_{Z_1}}, \dots, Z_{k\text{pa}_{Z_k}}\}. \quad (3.1)$$

Assumim que les probabilitats condicionals que apareixen en la derivació estan ben definides en el suport rellevant. En aquest model, les regles donen $Q_t(\omega) = Q_{c,t}(\omega)$, $T_c(\omega) = T_{q,c}(\omega) = T_q(\omega) = T(\omega)$, i:

$$C_q(\omega) = C_{q,t}(\omega), \quad (3.2)$$

$$Q_t \perp\!\!\!\perp_P \{C_q, T\}. \quad (3.3)$$

Les igualtats reflecteixen que Q només respon causalment a T , que T no té pares endògens i que C només respon causalment a Q , respectivament. La independència $Q_t \perp\!\!\!\perp_P \{C_q, T\}$ és la traducció contrafactual de $U_1 \perp\!\!\!\perp_{P(\mathcal{U})} U_2$: un cop imposat $T = t$ per a Q_t i $Q = q$ per a C_q , la restricció d'independència separa Q_t tant de C_q com de T ¹⁹.

Efecte de T sobre Q . Per (3.3), en particular (per reducció de $\perp\!\!\!\perp_P$), $Q_t \perp\!\!\!\perp_P T$. Per tant,

$$P(Q = q \mid \text{do}(t)) = P(Q_t = q) = P(Q_t = q \mid T = t) = P(Q = q \mid T = t) = p(q \mid t). \quad (3.4)$$

La tercera igualtat és consistència, 3.9: si $T = t$, Q coincideix amb Q_t (el quitrà que hauríem observat sota la intervenció $\text{do}(T = t)$). *Efecte de Q sobre C .* Per fer explícites les marginalitzacions, desenvolupem la resta de la derivació en el cas discret; la versió general s'escriu amb esperances. Comencem amb la llei de la probabilitat total sobre T . Així, per a cada valor t' i per als valors q amb probabilitat positiva condicionalment a $T = t'$,

$$P(C_q = c) = \sum_{t'} P(C_q = c \mid T = t')p(t').$$

De (3.3), amb t substituït per t' , i per simetria (S) i unió feble (UF) de $\perp\!\!\!\perp_P$:

$$Q_{t'} \perp\!\!\!\perp_P \{C_q, T\} \xrightarrow{\text{(UF)}} Q_{t'} \perp\!\!\!\perp_P C_q \mid T \xrightarrow{\text{(S)}} C_q \perp\!\!\!\perp_P Q_{t'} \mid T.$$

Així, per a cada valor t' i per als valors on els condicionals estan ben definits,

$$\begin{aligned} P(C_q = c \mid T = t') &= P(C_q = c \mid T = t', Q_{t'} = q) = P(C_q = c \mid T = t', Q = q) \\ &= P(C = c \mid T = t', Q = q) = p(c \mid q, t'). \end{aligned} \quad (3.5)$$

El primer pas usa la independència condicional anterior. El segon pas usa consistència per substituir $Q_{t'}$ per Q quan $T = t'$, i el tercer torna a usar consistència per substituir C_q per C quan $Q = q$. Substituint (3.5) en la suma anterior, l'efecte causal de Q sobre C és:

$$P(C = c \mid \text{do}(q)) = P(C_q = c) = \sum_{t'} p(c \mid q, t')p(t'). \quad (3.6)$$

Efecte de T sobre C . Finalment combinem els dos efectes. Fixem un estat $\omega \in \Omega$ i posem $q = Q_t(\omega)$.

¹⁸ Introduïm aquest concepte a 2.20. La idea és que en la projecció latent de la CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*)$ no existeixi cap camí format per arestes bidireccionals entre Y i Z_i , $i \in \{1, \dots, k\}$.

¹⁹ Qualsevol camí entre Q i C format per variables de $\mathcal{U}$ queda exclòs per (3.1). El mateix passa amb T : $T = T_0 = f_T(U_1)$, mentre que $Q_t = f_Q(t, U_2)$; com $U_1 \perp\!\!\!\perp_{P(\mathcal{U})} U_2$, obtenim la independència corresponent.

Per composició, si sota $\text{do}(T = t)$ el quitrà pren el valor q , afegir la intervenció $\text{do}(Q = q)$ no altera el valor de C : $C_t(\omega) = C_{t,q}(\omega)$. Per (3.2), a més, $C_{t,q}(\omega) = C_q(\omega)$. Com que aquest q és precisament $Q_t(\omega)$, podem escriure $C_t(\omega) = C_{Q_t}(\omega)$. Aquesta abreviació significa que $C_{Q_t}(\omega) = C_q(\omega)$ amb $q = Q_t(\omega)$; amb la convenció estructural anterior, equival a $\tilde{C}_{Q_t(\omega)}(\mathbf{U}(\omega))$. Per tant,

$$\begin{aligned} P(C = c \mid \text{do}(t)) &= P(C_t = c) = P(C_{Q_t} = c) = \sum_q P(C_{Q_t} = c \mid Q_t = q)P(Q_t = q) \\ &= \sum_q P(C_q = c \mid Q_t = q)P(Q_t = q) = \sum_q P(C_q = c)P(Q_t = q). \end{aligned} \quad (3.7)$$

En l'últim pas hem usat que (3.3) implica $Q_t \perp\!\!\!\perp_P C_q$. Substituint (3.4) i (3.6) a (3.7), obtenim exactament l'ajustament *front-door* de 2.36:

$$P(C = c \mid \text{do}(t)) = P(C_t = c) = \sum_q p(q \mid t) \sum_{t'} p(c \mid q, t')p(t').$$

Nota 3.76. En l'últim pas no fa falta que passem per la condicional. En efecte, podem escriure:

$$P(C_t = c) = \sum_q P(C_t = c, Q_t = q) \stackrel{\text{Comp.}}{=} \sum_q P(C_q = c, Q_t = q) \stackrel{(3.3)}{=} \sum_q P(C_q = c)P(Q_t = q).$$

Estudis experimentals i observacionals

L'home, servidor i intèrpret de la natura, tant fa i entén com ha observat de l'ordre de la natura per l'experiència o per la reflexió; ni sap ni pot res més.

Francis BACON, *Novum Organum* (1620)

Al capítol anterior hem fixat l'estimand causal, les hipòtesis que el fan identificable i els estadístics observables. Ara els lliguem al disseny de l'estudi. Una mateixa diferència de mitjanes pot estar justificada pel disseny d'un experiment aleatoritzat o bé descansar en hipòtesis fortes si prové d'un estudi observacional [HR20, Cap. 2–3].

Neyman i Fisher van donar una sortida neta al problema fonamental de la inferència causal: assignar els tractaments a l'atzar (cf. [Sp190] i [IR15, Secs. 2.3 i 2.5]). L'aleatorització no resol el problema fonamental de la inferència causal, però fa aleatori quin resultat potencial observem i evita que aquest patró depengui del pronòstic de cada unitat [IR15, Secs. 6.3–6.5]. Però no sempre podem aleatoritzar. Imaginem, per exemple, que volem estudiar l'impacte del tabac en el càncer de pulmó. D'una banda, factors de confusió com l'estil de vida i la salut prèvia dels fumadors alteren els resultats. De l'altra, assignar persones a fumar durant anys les exposaria a riscos previsibles per a la seva salut; per això, un assaig d'aquest tipus difícilment superaria una revisió ètica moderna sota els principis de la Declaració d'Hèlsinki [Wor25, pars. 16–18 i 25–26], encara que disposéssim del consentiment explícit i informat dels pacients.

Davant d'aquestes limitacions, veurem que hi ha vies menys lesives per arribar a un bon resultat. En particular, es distingeixen dos tipus d'estudis: els experimentals, en què sí podem controlar el mecanisme d'assignació, i els observacionals, en què els grups ja venen donats, o bé no tenim cap control sobre com s'ha decidit qui rep què. Aquests darrers plantegen un repte seriós per a la inferència causal, ja que el grup de tractament i el de control poden diferir de moltes altres maneres, més enllà del tractament en si. Comparar-los directament pot portar a conclusions enganyoses.

Notació 4.1. Mantindrem la notació del capítol anterior, 3.1, per a unitats i resultats potencials: $I_u = \{1, \dots, n\}$, $W_i \in \{0, 1\}$, $\mathbf{W} = (W_1, \dots, W_n)$, $\mathbf{w} \in \mathcal{A}_n$, $\mathbf{Y}(0)$, $\mathbf{Y}(1)$, $\mathbf{Y}^{\text{obs}}$, covariants $\mathbf{X}_i$, etc.

- Escriurem $n_t = \sum_{i=1}^n W_i$ i $n_c = \sum_{i=1}^n (1 - W_i)$ per al nombre d'unitats tractades i de control. Quan el disseny els fixa, les esperances i variàncies amb subíndex $\mathbf{W}$, com $E_{\mathbf{W}}$ i $\text{Var}_{\mathbf{W}}$, es prenen respecte de l'aleatorització de $\mathbf{W}$, condicionant implícitament a $\mathbf{Y}(0)$, $\mathbf{Y}(1)$ i $\mathbf{X}$ si s'escau.
- En població finita, $\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ denota el mecanisme d'assignació de 3.31 i $\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ la probabilitat marginal d'assignació de la unitat i .
- Els estimadors d'efectes causals s'escriuran amb barret, per exemple $\hat{\tau}_{\text{dif}}$, $\hat{\tau}_{\text{sub}}$ i $\hat{\tau}_{\text{IPW}}$. Les quantitats sense barret, com SATE o PATE, són estimands.

- En estudis observacionals, P i $\mathbb{E}$ designaran la llei i l'esperança de superpoblació de 3.24 (ometrem el subíndex sp). Quan definim funcions poblacionals, W , $\mathbf{X}$ i Y^{obs} denoten la unitat genèrica; quan les avaluem a la mostra, escriurem W_i , $\mathbf{X}_i$ i Y_i^{obs} . Així, el valor de propensió poblacional serà $e(\mathbf{x}) = P(W = 1 \mid \mathbf{X} = \mathbf{x})$ i, a la pràctica, una estimació seva a la unitat i s'escriurà $\hat{e}(\mathbf{X}_i)$.
- Els pesos de ponderació inversa s'escriuran λ_i . Les mitjanes condicionals observacionals que apareixen en l'ajustament s'escriuran $m_w(\mathbf{x}) = \mathbb{E}[Y^{\text{obs}} \mid W = w, \mathbf{X} = \mathbf{x}]$, per a $w \in \{0, 1\}$.

4.1

ESTUDIS EXPERIMENTALS

Aquesta secció explica què guanyem quan controlem el mecanisme d'assignació del tractament. L'objectiu és veure per què l'aleatorització converteix comparacions en estimadors causals (sota condicions adequades).

En un estudi experimental, el disseny fixa el mecanisme d'assignació: quins vectors d'assignació són admissibles i amb quines probabilitats. En un assaig controlat aleatoritzat (*randomized controlled trial*, RCT), aquest mecanisme incorpora un dispositiu aleatori; per tant, abans d'observar els resultats, l'assignació no queda determinada pel pronòstic de la unitat [HR20, Cap. 2]. En la terminologia del capítol anterior, el disseny imposa una forma de no-confusió. En els experiments clàssics de [IR15, Secs. 3.6 i 4.4], a més, el mecanisme és conegut i controlat per l'investigador.

En l'experiment completament aleatoritzat que tractarem tot seguit, si també es compleixen consistència i SUTVA (cf. 3.8 i 3.9), la diferència de mitjanes observades és un estimador sense biaix del SATE respecte de l'aleatorització [IR15, Sec. 6.3]. Aquesta conclusió és únicament aplicable al conjunt d'unitats estudiades: per llegir el SATE com a PATE cal alguna cosa més, com ara que les unitats de l'estudi representin la població objectiu o assumir un model de superpoblació [IR15, Sec. 6.7].

Comencem pel cas més simple: l'experiment completament aleatoritzat. Podem imaginar una urna amb una bola etiquetada per cada unitat; n'extraiem n_t sense reemplaçament i assignem aquestes unitats al tractament, mentre que les $n_c = n - n_t$ restants queden al control [IR15, Sec. 4.4].

Definició 4.2 (Experiment completament aleatoritzat [IR15, Def. 4.2 i Sec. 4.4]). Fixem $0 < n_t < n$ i posem $n_c = n - n_t$. Un experiment és completament aleatoritzat si exactament n_t unitats són assignades al tractament i tots els vectors d'assignació amb aquest nombre de tractats són igualment probables:

$$\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \binom{n}{n_t}^{-1}, & \text{si } \sum_{i=1}^n w_i = n_t, \\ 0, & \text{altrament.} \end{cases}$$

Proposició 4.3 (Aleatorització i no-confusió [IR15, Sec. 4.7]). En un experiment completament aleatoritzat, el mecanisme d'assignació no depèn ni de les covariants ni dels resultats potencials. En particular,

$$\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \pi(\mathbf{w}) \implies \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \frac{n_t}{n}, \text{ i } 1 - \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \frac{n_c}{n}.$$

Per tant, l'assignació és probabilística i no-confosa en el sentit finit de 3.39.

Nota 4.4. La proposició anterior no diu que els indicadors W_i siguin independents entre ells. De fet, la restricció $\sum_i W_i = n_t$ els fa dependents. El punt clau és un altre: l'assignació del tractament no depèn dels resultats potencials, sinó de l'aleatorietat.

MOSTREIG REPETIT DE NEYMAN

Un cop fixat el disseny experimental, podem estudiar l'estimador com una variable aleatòria induïda per l'assignació. S'hi recullen la diferència de mitjanes, la seva esperança sota aleatorització, la variància de Neyman i l'estimador conservador de la variància.

Ara passem de l'estimand als estimadors, és a dir, als estadístics que depenen només de les dades observades $\mathbf{W}$, $\mathbf{Y}^{\text{obs}}$ i, si escau, de les covariants $\mathbf{X}$ o de funcions conegudes d'aquestes covariants (cf. 3.11 i annexos). La diferència de mitjanes observades depèn només de $\mathbf{W}$ i $\mathbf{Y}^{\text{obs}}$, i ja és present en Neyman [Sp190]. En un experiment completament aleatoritzat podem calcular-ne l'esperança i la variància respecte de l'assignació [IR15, Secs. 6.3–6.5].

Definició 4.5 (Estimador de diferència de mitjanes [IR15, Sec. 6.3]). Suposem $0 < n_t < n$, on $n_t = \sum_{i=1}^n W_i$ i $n_c = \sum_{i=1}^n (1 - W_i)$. Definim l'estimador de diferència de mitjanes observades per:

$$\hat{\tau}_{\text{dif}} = \frac{1}{n_t} \sum_{i=1}^n W_i Y_i^{\text{obs}} - \frac{1}{n_c} \sum_{i=1}^n (1 - W_i) Y_i^{\text{obs}}. \quad (4.1)$$

Nota 4.6. Quan el disseny fixa n_t i n_c , l'única font d'aleatorietat de $\hat{\tau}_{\text{dif}}$ és el vector d'assignació $\mathbf{W}$. Per això escriurem $\mathbb{E}_{\mathbf{W}}$ i $\text{Var}_{\mathbf{W}}$ per indicar esperança i variància respecte de la distribució d'aleatorització (coneguda), condicionant implícitament als resultats potencials $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$ [IR15, Sec. 6.3].

Teorema 4.7 (Absència de biaix de l'estimador de Neyman [IR15, Sec. 6.3, Th. 6.1]). Suposem que l'assignació és completament aleatòria amb n_t unitats tractades, és a dir, tots els vectors $\mathbf{w} \in \mathcal{A}_n$ tals que $\sum_{i=1}^n w_i = n_t$ tenen probabilitat $\binom{n}{n_t}^{-1}$. Aleshores:

$$\mathbb{E}_{\mathbf{W}}[\hat{\tau}_{\text{dif}} \mid \mathbf{Y}(0), \mathbf{Y}(1)] = \text{SATE}.$$

En particular, $\hat{\tau}_{\text{dif}}$ és un estimador sense biaix del SATE respecte de la distribució d'aleatorització.

Notació 4.8 (Variàncies de mostra finita). Escrivim $\bar{Y}(w) = n^{-1} \sum_{i=1}^n Y_i(w)$ per a $w \in \{0, 1\}$. Denotem les variàncies de mostra finita dels resultats potencials i dels efectes individuals per:

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i(1) - \bar{Y}(1))^2, \quad S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i(0) - \bar{Y}(0))^2, \quad S_{\tau}^2 = \frac{1}{n-1} \sum_{i=1}^n (\tau_i - \text{SATE})^2.$$

Teorema 4.9 (Variància de Neyman [IR15, Sec. 6.4, Th. 6.2]). *Sota les hipòtesis de 4.7, la variància de l'estimador (4.1) respecte de l'aleatorització és:*

$$\text{Var}_{\mathbf{W}}(\hat{\tau}_{\text{dif}}) = \frac{S_1^2}{n_t} + \frac{S_0^2}{n_c} - \frac{S_\tau^2}{n}.$$

Observació 4.10. El terme S_τ^2 mesura l'heterogeneïtat dels efectes individuals. No el podem estimar directament: caldria observar $Y_i(1)$ i $Y_i(0)$ per a cada unitat, la qual cosa ens retorna al problema fonamental de la inferència causal (cf. 3.15; [Hol86]; [IR15, Sec. 6.4]).

Corol·lari 4.11 (Estimador conservador de la variància). *Suposem, a més, que $n_t, n_c > 1$. Definim les mitjanes observades $\bar{Y}_1^{\text{obs}} = n_t^{-1} \sum_{j:W_j=1} Y_j^{\text{obs}}$ i $\bar{Y}_0^{\text{obs}} = n_c^{-1} \sum_{j:W_j=0} Y_j^{\text{obs}}$, i les variàncies mostrals observades dins de cada grup:*

$$s_1^2 = \frac{1}{n_t - 1} \sum_{i:W_i=1} (Y_i^{\text{obs}} - \bar{Y}_1^{\text{obs}})^2, \quad s_0^2 = \frac{1}{n_c - 1} \sum_{i:W_i=0} (Y_i^{\text{obs}} - \bar{Y}_0^{\text{obs}})^2.$$

L'estimador de Neyman de la variància és:

$$\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}_{\text{dif}}) = \frac{s_1^2}{n_t} + \frac{s_0^2}{n_c}.$$

Aleshores [IR15, Sec. 6.5]:

$$\mathbb{E}_{\mathbf{W}} \left[\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}_{\text{dif}}) \mid \mathbf{Y}(0), \mathbf{Y}(1) \right] - \text{Var}_{\mathbf{W}}(\hat{\tau}_{\text{dif}}) = \frac{S_\tau^2}{n} \geq 0.$$

Observació 4.12. Si l'efecte causal és constant, $S_\tau^2 = 0$ i l'estimador és sense biaix. En general, és conservador en esperança perquè no pot restar el terme no observable S_τ^2/n .

4.2

ESTUDIS OBSERVACIONALS

Aquí abordem el cas més habitual i més delicat: les dades on l'assignació del tractament no ha estat controlada per l'investigador. La secció explica quines hipòtesis fan possible llegir un estudi observacional com si fos un experiment condicionat a covariants.

En un estudi observacional, l'investigador *observa* qui ha rebut el tractament i qui no, però no ha pogut decidir aquesta assignació aleatòriament. Per això, la comparació directa entre tractats i controls pot barrejar l'efecte del tractament amb diferències prèvies entre grups; ja d'entrada, el grup tractat pot tenir un pronòstic diferent del grup de control fins i tot abans de rebre el tractament [HR20, Cap. 3 i 7].

En un experiment ideal, la no-confusió prové del disseny. En un estudi observacional, és una hipòtesi: cal defensar que, un cop condicionem per covariants mesurades, el tractament rebut no aporta informació addicional sobre els resultats potencials [HR20, Cap. 3]. En la notació del capítol anterior, aquesta és la ignorabilitat condicional de 3.45, juntament amb la consistència de 3.9 i la positivitat de 3.46. Sota

aquestes condicions, la proposició d'ajustament 3.53 permet escriure l'efecte causal en termes de la distribució observacional.

La idea, doncs, és que un estudi observacional ben justificat intenta imitar un experiment aleatoritzat condicionat a covariants. Si hi ha confusors no mesurats, o si existeix un subconjunt mesurable de l'espai de covariants amb probabilitat positiva on $P(W = 1 | \mathbf{X})$ pren el valor 0 o 1, aquesta lectura falla [HR20, Sec. 3.1–3.3].

ASSIGNACIÓ NO-CONFOSA EN ESTUDIS OBSERVACIONALS

Es precisa quina hipòtesi substitueix l'aleatorització quan l'investigador no controla el tractament. L'objectiu és formular els estudis observacionals com experiments condicionals.

En [IR15, Secs. 3.3–3.7 i 12.1–12.2], un estudi observacional és un estudi en què la forma funcional del mecanisme d'assignació és desconeguda. En la formulació de superpoblació adoptada al capítol anterior, el mecanisme és automàticament individualístic, un cop fixats els paràmetres globals, gràcies a l'assumpció de mostreig i.i.d. de la distribució conjunta; per obtenir comparacions causals exigim, a més, que sigui probabilístic i no-confós. La part de no-confusió diu que, fixades les covariants pretractament, la probabilitat de tractament no depèn dels resultats potencials:

$$P(W_i = 1 | Y_i(0), Y_i(1), \mathbf{X}_i) = P(W_i = 1 | \mathbf{X}_i), \quad (4.2)$$

amb $0 < P(W = 1 | \mathbf{X} = \mathbf{x}) < 1$ per als valors $\mathbf{x}$ tals que $p(\mathbf{x}) > 0$. Dins de cada valor de $\mathbf{X}_i$, el tractament es comporta com si s'hagués assignat aleatòriament, encara que l'investigador no conegui les probabilitats exactes d'assignació [IR15, Sec. 12.2].

La condició (4.2) no es pot verificar només amb $(Y_i^{\text{obs}}, W_i, \mathbf{X}_i)$, perquè parla de resultats potencials no observats [IR15, Sec. 12.2]. Per això les covariants s'han de triar sense fer servir els resultats i s'han de justificar amb coneixement del problema [IR15, Secs. 12.5 i 13.3]; [HR20, Cap. 3], ja que el balanç de covariants no és una prova de la ignorabilitat.

ESTRATIFICACIÓ PEL VALOR DE PROPENSIO

El valor de propensió redueix el problema d'ajustar per moltes covariants a l'ajust per una sola funció. S'hi introdueix la propietat d'equilibratge i es mostra com motiva la subclassificació.

Ajustar directament per totes les covariants pot ser inviable si $\mathbf{X}_i$ és d'alta dimensió o gairebé contínua (els estrats exactes poden quedar buits o contenir només tractats o només controls). [RR83] proposen substituir $\mathbf{X}_i$ per una funció escalar que manté la comparabilitat necessària. Aquí farem servir la versió poblacional del valor de propensió:

$$e(\mathbf{x}) = P(W = 1 | \mathbf{X} = \mathbf{x}) = \mathbb{E}[W | \mathbf{X} = \mathbf{x}],$$

de manera que $e(\mathbf{X}_i)$ és l'avaluació d'aquesta funció a les covariants observades de la unitat i . Com que hem assumit $W \in \{0, 1\}$, aquesta probabilitat condicional coincideix amb l'esperança condicional.

Definició 4.13 (Valor d'equilibratge [RR83]; [HR20, Technical Point 15.1]). Una funció mesurable b de les covariants és un *valor d'equilibratge* si, posant $B_i := b(\mathbf{X}_i)$, es compleix $W_i \perp_P \mathbf{X}_i \mid B_i$. Equivalentment, dins de cada nivell de $b(\mathbf{X}_i)$, la distribució de les covariants és la mateixa en el grup tractat i en el grup de control. Les igualtats condicionals s'entenen per a $P(B_i)$ -gairebé tot valor de B_i ; en el cas discret, això vol dir per als valors t tals que $P(B_i = t) > 0$.

Tres propietats fan útil el valor de propensió: conserva la ignorabilitat i la positivitat que ja valien condicionant per totes les covariants, és efectivament un valor d'equilibratge i, finalment, és el resum menys fi dins d'aquesta família, en el sentit que es pot recuperar a partir de qualsevol altre valor d'equilibratge.

Proposició 4.14 (Equilibratge, ignorabilitat i positivitat [IR15, Sec. 12.3, Lema 12.2]). *Suposem que $b(\mathbf{X}_i)$ és un valor d'equilibratge. Si es compleix la ignorabilitat condicional donada $\mathbf{X}_i$ aleshores també es compleix la ignorabilitat condicional donada $b(\mathbf{X}_i)$, és a dir:*

$$(Y_i(0), Y_i(1)) \perp_P W_i \mid \mathbf{X}_i \implies (Y_i(0), Y_i(1)) \perp_P W_i \mid b(\mathbf{X}_i).$$

Si, a més, $0 < e(\mathbf{X}_i) < 1$ quasi segurament, aleshores $0 < P(W_i = 1 \mid b(\mathbf{X}_i)) < 1$ quasi segurament (cf. [RR83, Th. 3]; [HR20, Technical Point 15.1]).

Proposició 4.15 (El valor de propensió és d'equilibratge [RR83, Th. 1]; [IR15, Lema 12.1]). *Sigui $e(\mathbf{x}) = P(W = 1 \mid \mathbf{X} = \mathbf{x})$. Aleshores $e(\mathbf{X}_i)$ és un valor d'equilibratge, és a dir, $W_i \perp_P \mathbf{X}_i \mid e(\mathbf{X}_i)$. En particular, si es compleix la ignorabilitat condicional donada $\mathbf{X}_i$, llavors també es compleix:*

$$(Y_i(0), Y_i(1)) \perp_P W_i \mid e(\mathbf{X}_i).$$

Proposició 4.16 ([RR83, Th. 2]; [IR15, Sec. 12.3, Lema 12.3]). *Sigui $b(\mathbf{X}_i)$ un valor d'equilibratge. Aleshores existeix una funció f tal que $e(\mathbf{X}_i) = f(b(\mathbf{X}_i))$ quasi segurament. Qualsevol valor d'equilibratge conté, com a mínim, la informació necessària per recuperar $e(\mathbf{X}_i)$.*

Els tres resultats parlen del mecanisme d'assignació. La consistència entra en el pas d'identificació, quan substituïm $Y_i(w)$ per Y_i^{obs} en les unitats amb $W_i = w$. A partir d'aquí assumirem consistència, ignorabilitat condicional i positivitat, llevat que s'indiqui el contrari [RR83]; [IR15, Secs. 12.2–12.4]; [HR20, Cap. 15].

El valor de propensió no elimina la confusió no mesurada. Només diu que, si les covariants $\mathbf{X}_i$ ja justificaven la ignorabilitat, $e(\mathbf{X}_i)$ també ho fa. La legitimitat causal de la llista de covariants i l'ús del valor de propensió per equilibrar-les són dues qüestions diferents [IR15, Secs. 12.2, 12.5, 13.3 i 14.4].

Sota ignorabilitat forta (cf. 3.47), [RR83, Cor. 4.2] garanteix que l'ajustament per estrats exactes d'un valor d'equilibratge és no esbiaixat quan les diferències dins dels estrats es ponderen per les seves fraccions poblacionals. A la pràctica, però, dividim les unitats en intervals de valors semblants de $e(\mathbf{X}_i)$ o, habitualment, de $\hat{e}(\mathbf{X}_i)$.

Si $I_1, \dots, I_G$ és una partició de la mostra ordenada pel valor de propensió, agreguem les diferències dins d'estrat amb pesos n_g/n [IR15, Sec. 17.3.2]; [HR20, Sec. 15.3].

Definició 4.17 (Estimador per subclassificació). Siguin $I_1, \dots, I_G$ estrats disjunts de la mostra, amb $n_g = |I_g|$, $n_{tg} = \sum_{i \in I_g} W_i$ i $n_{cg} = \sum_{i \in I_g} (1 - W_i)$. Si $n_{tg}, n_{cg} > 0$ per a tot g , definim:

$$\hat{\tau}_{\text{sub}} = \sum_{g=1}^G \frac{n_g}{n} \left[\frac{1}{n_{tg}} \sum_{i \in I_g} W_i Y_i^{\text{obs}} - \frac{1}{n_{cg}} \sum_{i \in I_g} (1 - W_i) Y_i^{\text{obs}} \right].$$

La definició anterior recull l'estimador pràctic. Si els estrats fossin exactes i les hipòtesis d'identificació fossin certes, cada diferència compararia tractats i controls amb covariants equilibrades [RR83, Th. 4 i Cor. 4.2]. Com que acostumem a considerar continu el valor de propensió, la subclassificació només aproxima l'ajust exacte [HR20, Sec. 15.3]. Per això cal revisar solapament, presència de tractats i controls dins de cada estrat, i balanç després de subclassificar [IR15, Sec. 17.3]; [HR20, Sec. 15.3].

PONDERACIÓ INVERSA

La ponderació inversa escriu el mateix estimand amb pesos.

La proposició d'identificació per ajustament expressa el PATE amb mitjanes condicionals. La ponderació inversa l'escriu amb pesos: les unitats (donades les seves covariants) poc probables dins del grup on han aparegut compten més. D'aquesta manera, la població ponderada compensa el desequilibri de covariants entre tractats i controls [HR20, Sec. 2.4]. La idea adapta l'estimador de Horvitz–Thompson al valor de propensió [HT52]; [IR15, Sec. 17.8].

Proposició 4.18 (Ponderació inversa i PATE [IR15, Sec. 17.8]). Supposem que les unitats formen una mostra aleatòria de la població objectiu, que $Y_i(0)$ i $Y_i(1)$ són integrables, i que es compleixen consistència, ignorabilitat condicional i positivitats (cf. 3.9, 3.45 i 3.46). Supposem també que el valor de propensió poblacional $e(\mathbf{x}) = P(W = 1 \mid \mathbf{X} = \mathbf{x})$ és conegut. Definim la versió no normalitzada de l'estimador de ponderació inversa com:

$$\hat{\tau}_{\text{IPW}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{W_i Y_i^{\text{obs}}}{e(\mathbf{X}_i)} - \frac{(1 - W_i) Y_i^{\text{obs}}}{1 - e(\mathbf{X}_i)} \right) = \frac{1}{n} \left(\sum_{i: W_i=1} \lambda_i Y_i^{\text{obs}} - \sum_{i: W_i=0} \lambda_i Y_i^{\text{obs}} \right),$$

on:

$$\lambda_i = \frac{1}{e(\mathbf{X}_i)^{W_i} (1 - e(\mathbf{X}_i))^{1-W_i}} = \begin{cases} \frac{1}{1 - e(\mathbf{X}_i)}, & \text{si } W_i = 0, \\ \frac{1}{e(\mathbf{X}_i)}, & \text{si } W_i = 1. \end{cases}$$

Aleshores $E[\hat{\tau}_{\text{IPW}}] = \text{PATE}$ (cf. 3.25), on l'esperança es pren respecte del mostreig poblacional i del mecanisme d'assignació, com hem fixat a 4.1.

Nota 4.19. A la pràctica, el valor de propensió se substitueix per una estimació $\hat{e}(\mathbf{X}_i)$, ja que rarament és conegut. La igualtat anterior correspon al cas ideal amb e conegut; amb $\hat{e}$ apareixen problemes de modelització, balanç i variància que queden fora de l'abast d'aquest treball.

Cas pràctic

Qui és capaç de calcular la collita que la natura dona als camps no treballats per l'home?

Henry David THOREAU, *Journal* (1845-1846)

Notació 5.1 (Convenció notacional del capítol). Mantenim la convenció dels capítols 1 i 2, en particular la convenció d'assignacions de 1.2: $\mathbf{v} = (v_i)_{i \in I} \in \mathcal{X}$ i $\mathbf{s} := \mathbf{v}_{\mathbf{S}} = (v_i)_{V_i \in \mathbf{S}} \in \mathcal{X}_{\mathbf{S}}$ és la seva restricció a $\mathbf{S} \subseteq \mathcal{V}$. Si $\mathbf{a} := \mathbf{v}_{\mathbf{A}}$ i $\mathbf{b} := \mathbf{v}_{\mathbf{B}}$ són restriccions de la mateixa assignació, escriurem $\mathbf{a} \cap \mathbf{b} := \mathbf{v}_{\mathbf{A} \cap \mathbf{B}}$, $\mathbf{a} \cup \mathbf{b} := \mathbf{v}_{\mathbf{A} \cup \mathbf{B}}$ i $\mathbf{a} \setminus \mathbf{b} := \mathbf{v}_{\mathbf{A} \setminus \mathbf{B}}$. Durant tot el capítol, suposem l'SCM $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}))$, on $\mathcal{U}$ és el conjunt de variables exògenes i $\mathcal{V} = \{V_i\}_{i \in I}$ és el conjunt de variables endògenes observades, amb $I = \{1, \dots, n\}$. Seguint [SP06], en aquest capítol suposem que $\mathcal{V}$ té domini finit. Aquesta convenció local no restringeix el marc general dels capítols anteriors: l'extensió contínua d'ID i IDC requeriria una formalització addicional i queda fora de l'abast del treball. Per alleugerir la notació, si $\mathbf{S} \subseteq \mathcal{V}$ i $\mathbf{s} := \mathbf{v}_{\mathbf{S}}$, escriurem també $\mathbf{v} \setminus \mathbf{s} := \mathbf{v}_{\mathcal{V} \setminus \mathbf{S}}$; així, $\text{do}(\mathbf{v} \setminus \mathbf{s})$ imposa a les variables de $\mathcal{V} \setminus \mathbf{S}$ els valors corresponents de $\mathbf{v}$. Els sumatoris com $\sum_{\mathbf{v} \setminus \mathbf{Y}}$ indiquen marginalització discreta respecte de $\mathcal{V} \setminus \mathbf{Y}$; anàlogament, $\sum_{\mathbf{v} \setminus (\mathbf{Y} \cup \mathbf{X})}$ marginalitza sobre $\mathcal{V} \setminus (\mathbf{Y} \cup \mathbf{X})$ i $\sum_{\mathbf{s} \setminus \mathbf{Y}}$ marginalitza sobre $\mathbf{S} \setminus \mathbf{Y}$. Sigui:

$$\mathcal{P}_*^{\mathcal{M}} = \{P(\mathcal{V} \mid \text{do}(\mathbf{x})) \mid \mathbf{X} \subseteq \mathcal{V}, \mathbf{x} \in \mathcal{X}_{\mathbf{X}}\},$$

la família de distribucions intervingudes del model $\mathcal{M}$. La projecció latent de $\mathcal{M}$ sobre les variables observades indueix una CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*^{\mathcal{M}})$, amb diagrama causal $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, on les arestes bidirigides representen confusió latent induïda per variables de $\mathcal{U}$. Fixem un ordre topològic $<$ sobre $\mathcal{V}$ compatible amb $\mathcal{G}$ i ordenem les variables seguint aquest ordre; és a dir, per a tota aresta dirigida $V_i \rightarrow V_j \in \mathcal{E}$ es compleix que $V_i < V_j$, la qual cosa implica que $i < j$. Reprenem també la notació del capítol 2: $\mathcal{V}^{(i)} = \{V_1, \dots, V_i\}$, $\mathbf{v}^{(i)} = \mathbf{v}_{\mathcal{V}^{(i)}}$, $\mathcal{G}[\mathbf{S}]$ és el subgraf induït per $\mathbf{S}$ i $C(\mathcal{G})$ és la partició de $\mathcal{G}$ en C-components maximals.

5.1

C-BOSCOS I HEDGES

Refinem les C-components en C-boscós i hedges, que són les estructures gràfiques que certifiquen la no-identificabilitat d'alguns efectes causals. Aquest llenguatge serà el pont entre la teoria de la identificació i el cas de Berkeley, on apareix un bow que força una fallada de l'algorisme ID.

Una de les fortaleces de les CBNs Markovianes és que els pares d'una variable la «protegeixen» de la influència dels seus no-descendents. En les CBNs semi-Markovianes, que són les que ens interessen aquí, aquesta garantia s'ha de reformular mitjançant les C-components i els pares estesos introduïts al capítol

2. Recordem-ne només el punt que necessitem: una C-component agrupa variables connectades per camins formats per arestes bidirigides, i les C-components maximals formen la partició $C(\mathcal{G})$. En aquest capítol no redefinirem aquestes components, sinó que les refinarem en arbres, boscos i hedges.

Definició 5.2 (Bosc i arbre causal, [SP06; SP08]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ el diagrama causal d'una CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*^M)$ induïda per la projecció latent d'un SCM M .

- $\mathcal{G}$ és un **bosc causal** (*forest*) si cada vèrtex té, com a màxim, un fill directe.
- $\mathcal{G}$ és un **arbre causal** (*tree*) si és un bosc causal on, a més a més, existeix un únic vèrtex sense fills directes (l'arrel causal de l'arbre).

En aquesta definició, els fills directes s'entenen respecte de les arestes dirigides.

Nota 5.3 ([SP08]). Notem, doncs, que en el context dels models causals, aquestes dues definicions donen la volta a la direcció habitual de les arestes que normalment s'agafa en la teoria de grafs més clàssica [BM08]. Allà, les arestes surten de l'arrel i es ramifiquen cap a les fulles (així doncs, un node típicament té un sol pare, però pot tenir múltiples fills i l'arrel és l'únic node sense pares). En la terminologia de [SP08], en canvi, l'arrel és el nostre conjunt terminal o *sink*: el vèrtex sense descendents on van a parar els camins dirigits.

Amb les C-components i la partició $C(\mathcal{G})$ ja fixades, aquestes estructures permeten definir C-arbres i C-boscos.

Definició 5.4 (C-arbre, C-tree [SP06, Def. 4]; [SP08, Def. 11]). Sigui $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ una C-component que sigui, a més, un arbre (cf. 5.2). Si existeix $Y \in \mathcal{V}_{\mathcal{H}}$ de manera que $\text{An}_Y^{\mathcal{H}} = \mathcal{V}_{\mathcal{H}}$, direm que $\mathcal{H}$ és un **C-arbre arrelat en Y** (*Y-rooted C-tree*).

Definició 5.5 (C-bosc, C-forest [SP06, Def. 5]; [SP08, Def. 15]). Sigui $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ una C-component que sigui, a més, un bosc (cf. 5.2), i sigui $\mathbf{R} = \text{Sink}(\mathcal{H})$ el seu conjunt d'arrels, calculat respecte de les arestes dirigides. Direm que $\mathcal{H}$ és un **C-bosc arrelat en $\mathbf{R}$** (*R-rooted C-forest*).

L'anàlisi de la identificabilitat ens porta a estudiar certes configuracions dins del graf. En particular, quan $\mathcal{G}$ conté un parell de C-boscos diferents que compleixen certes propietats, s'origina una estructura anomenada *hedge*. Aquestes subestructures apareixen de manera natural en l'estudi d'aquest problema (cf. [SP06; SP08]) i ens ajuden a tancar el cercle de la identificabilitat per caracteritzar completament la no-identificabilitat d'efectes causals.

Definició 5.6 (Hedge [SP06, Def. 6]; [SP08, Def. 16]). Siguin $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{V}$ dos conjunts disjunts de variables, i siguin $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ i $\mathcal{H}' = (\mathcal{V}_{\mathcal{H}'}, \mathcal{E}_{\mathcal{H}'})$ dos C-boscos de $\mathcal{G}$ arrelats en el mateix conjunt $\mathbf{R}$, amb $\mathcal{H}'$ subgraf de $\mathcal{H}$. Direm que $\mathcal{H}$ i $\mathcal{H}'$ formen un *hedge* per a $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ en $\mathcal{G}$ si es compleixen les condicions següents:

$$\mathcal{V}_{\mathcal{H}} \cap \mathbf{X} \neq \emptyset, \quad \mathcal{V}_{\mathcal{H}'} \cap \mathbf{X} = \emptyset, \quad \mathbf{R} \subseteq \text{An}_{\mathbf{Y}}^{\mathcal{G}_{\bar{\mathbf{x}}}}.$$

Observació 5.7 ([SP06; SP08; Pea09]). Aquesta definició generalitza el cas dels C-arbres. En particular, qualsevol *Y-rooted C-tree* i el subgraf trivial format únicament pel node arrel Y determinen un cas particular de *hedge*. Així, els *hedges* proporcionen l'estructura adequada per descriure quan un efecte causal no es pot recuperar a partir de la distribució observacional.

Exemple 5.8. Considerem una CBN semi-Markoviana amb diagrama causal $\mathcal{G}$, representat a la figura 5.1. Les seves C-components maximals són $\mathbf{S}_1 = \{A, B, C, D, M\}$, $\mathbf{S}_2 = \{Y, Z\}$ i $\mathbf{S}_3 = \{W\}$ (podem veure $\mathcal{G}[\mathbf{S}_1]$ i $\mathcal{G}[\mathbf{S}_2]$ representats a les subfigures 5.1b i 5.1c, mentre que $\mathcal{G}[\mathbf{S}_3] = \{W\}$ és la component trivial).

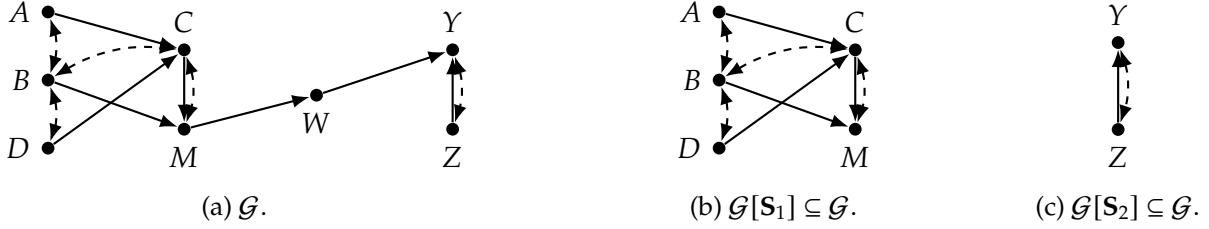


Figura 5.1: Una CBN semi-Markoviana amb tres C-components maximals. La parella $(\mathcal{G}[\mathbf{S}_1], \mathcal{H})$, on $\mathcal{H}$ és el subgraf reduït al node M , determina un *hedge* per a $p(y \mid \text{do}(a))$.

En particular, $\mathcal{G}$ no és una C-component, ja que no tots els seus vèrtexs es poden connectar mitjançant camins formats únicament per arestes bidirigides. A més, aquestes tres C-components són també C-trees. D'una banda, $\mathcal{G}[\mathbf{S}_1]$ és un M -rooted C-tree, perquè $\text{An}_M^{\mathcal{G}[\mathbf{S}_1]} = \{A, B, C, D, M\}$. D'altra banda, $\mathcal{G}[\mathbf{S}_2]$ és un Y -rooted C-tree, ja que $\text{An}_Y^{\mathcal{G}[\mathbf{S}_2]} = \{Y, Z\}$. Finalment, $\mathcal{G}[\mathbf{S}_3] = \{W\}$ és trivialment un W -rooted C-tree.

Vegem ara que $\mathcal{G}$ conté un *hedge* per a $p(y \mid \text{do}(a))$. Considerem el subgraf $\mathcal{G}[\mathbf{S}_1]$. Aquest és un $\{M\}$ -rooted C-forest i, a més, es compleix que $\{A\} \cap \mathbf{S}_1 \neq \emptyset$. Considerem ara el subgraf $\mathcal{H}$ format únicament pel node M i sense cap aresta.

Aleshores, $\mathcal{H}$ és també un $\{M\}$ -rooted C-forest²⁰, i tenim $\{A\} \cap \mathcal{V}_{\mathcal{H}} = \{A\} \cap \{M\} = \emptyset$ i $\mathcal{H} \subseteq \mathcal{G}[\mathbf{S}_1]$. D'altra banda, tenim que $\{M\} \subseteq \text{An}_Y^{\mathcal{G}[\mathbf{S}_2]}$, ja que el camí dirigit $M \rightarrow W \rightarrow Y$ queda intacte. Per tant, $\{M\} \subseteq \text{An}_Y^{\mathcal{G}[\mathbf{S}_2]}$. Per tot plegat, tenim que la parella $(\mathcal{G}[\mathbf{S}_1], \mathcal{H})$ forma un *hedge* per a $p(y \mid \text{do}(a))$ en $\mathcal{G}$. Pel criteri d'identificació mitjançant *hedges*, concloem que l'efecte causal $p(y \mid \text{do}(a))$ no és identificable.

Teorema 5.9 (Criteri d'identificació de *hedges* [SP06, Cor. 3]; [SP08, Th. 19]). *Siguin $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{V}$ dos conjunts disjunts de variables. Aleshores, $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ és identificable a partir de $p(\mathbf{v})$ en una CBN semi-Markoviana $(\mathcal{G}, \mathcal{P}_*)$ si, i només si, $\mathcal{G}$ no conté cap *hedge* per a $p(\mathbf{y}' \mid \text{do}(\mathbf{x}'))$, per a qualssevol $\mathbf{X}' \subseteq \mathbf{X}$ i $\mathbf{Y}' \subseteq \mathbf{Y}$.*

5.2

ALGORISMES ID i IDC

*Aquesta secció converteix la teoria de la identificació en un procediment efectiu. L'objectiu és entendre com l'algorisme ID descompon una consulta $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ fins a obtenir una expressió observacional, o bé detecta una obstrucció de tipus *hedge*; després veurem com IDC redueix el cas condicionat $p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x}))$ al cas anterior.*

Els criteris *back-door* i *front-door* donen dues portes d'entrada (mai millor dit) al problema d'identificació, però no exhaurixen tots els casos possibles. En grafs grans, buscar una derivació manual de l'estimand pot convertir-se ràpidament en una tasca complicada. Els algorismes ID i IDC resolen justament això:

²⁰ Com que el graf trivial amb un únic vèrtex és connex, $\mathcal{H}$ és una C-component trivial i, a més, un $\{M\}$ -rooted C-forest.

substitueixen la cerca manual per una descomposició guiada per C-components i, quan la consulta no és identificable, la fallada queda testimoniada per una obstrucció de tipus *hedge* [SP06; SP08; SP12].

Lema 5.10 (Factorització per C-components, [TP02, Lem. 1]; [Bar+20, Ap. D, Th. 11]). *Sigui $(\mathcal{G}, \mathcal{P}_*)$ una CBN semi-Markoviana, amb diagrama causal $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, i sigui $C(\mathcal{G}) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$ la partició de $\mathcal{G}$ en C-components maximals. Aleshores la PMF observacional $p(\mathbf{v})$ factoritza com:*

$$p(\mathbf{v}) = \prod_{j=1}^k p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)).$$

A més, cada factor és identificable a partir de $p(\mathbf{v})$ i es pot escriure, per a qualsevol ordre topològic fixat, com

$$p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)) = \prod_{\{i \mid V_i \in \mathbf{S}_j\}} p(v_i \mid \mathbf{v}^{(i-1)}).$$

Equivalentment, si $C^{(i)}$ és la C-component de V_i en $\mathcal{G}[\mathcal{V}^{(i)}]$, llavors $p(v_i \mid \mathbf{v}^{(i-1)}) = p(v_i \mid \mathbf{v}_{\text{Pa}_i^+}^+)$, on Pa_i^+ és el conjunt de parees estesos de la definició 2.23.

L'ALGORISME ID

L'algorisme ID 1 rep una consulta $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$, $p(\mathbf{v})$ i el diagrama causal semi-Markovià $\mathcal{G}$. La seva sortida és una expressió en termes de $p(\mathbf{v})$ o bé una fallada de la forma $\text{FAIL}(\mathcal{G}', \mathcal{G}'[\mathbf{S}])$ en alguna crida recursiva; d'aquesta parella de grafs se'n poden extreure dos C-boscós que formen un *hedge*. En el pseudocodi següent, si $\mathbf{A} = \text{An}_{\mathbf{Y}}^{\mathcal{G}}$, llavors $p(\mathbf{a})$ denota la marginal de p sobre els avantpassats de $\mathbf{Y}$, i $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$ denota la partició en C-components maximals del subgraf induït.

En el mateix article on es va presentar l'algorisme ID [SP06], els autors van demostrar que és correcte i complet, és a dir, que retorna una expressió correcta quan la consulta és identificable i, quan no ho és, garanteix la no-identificabilitat.

Teorema 5.11 ([SP06, Lem. 3, Th. 5, Cor. 2]). **ID** sempre acaba, quan retorna una expressió per a $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ aquesta és correcta. És un algorisme complet.

Línies 2–4. És el cas base més senzill. Si $\mathbf{X} = \emptyset$, no queda cap intervenció per eliminar: la consulta $p(\mathbf{y} \mid \text{do}(\emptyset))$ és simplement la marginal $p(\mathbf{y})$. Per tant, la sortida és $\sum_{\mathbf{v} \setminus \mathbf{y}} p(\mathbf{v})$.

Línies 5–8. Posem $\mathbf{A} := \text{An}_{\mathbf{Y}}^{\mathcal{G}}$. Per definició, $\mathbf{A}$ és el conjunt de variables des de les quals existeix un camí dirigit cap a algun node de $\mathbf{Y}$. En particular, si $V_i \in \mathbf{A}$ i $V_j \rightarrow V_i$, aleshores $V_j \in \mathbf{A}$. Equivalentment, no hi ha cap aresta dirigida de $\mathcal{V} \setminus \mathbf{A}$ cap a $\mathbf{A}$. Això permet reduir el problema d'identificació a $\mathcal{G}[\mathbf{A}]$. Les variables de $\mathcal{V} \setminus \mathbf{A}$ no poden aparèixer com a predecessors dirigits de cap variable de $\mathbf{A}$ i, per tant, les intervencions sobre variables fora de $\mathbf{A}$ no poden modificar la llei intervinguda de $\mathbf{Y}$. Així, $P(\mathbf{y} \mid \text{do}(\mathbf{x})) = P(\mathbf{y} \mid \text{do}(\mathbf{x} \cap \mathbf{a}))$; el subproblema en $\mathcal{G}[\mathbf{A}]$ pren com a distribució observacional la marginal $p(\mathbf{a}) = \sum_{\mathbf{v} \setminus \mathbf{a}} p(\mathbf{v})$ (en el cas discret), amb $\mathbf{a}$ un valor d' $\mathbf{A}$, i només conserva les intervencions dins d' $\mathbf{A}$. El lema següent formalitza que aquesta reducció preserva l'efecte causal que es vol identificar.

Algorisme 1 ID: identificació de $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ **Entrada:** Conjunts disjunts $\mathbf{Y}, \mathbf{X} \subseteq \mathcal{V}$, $p(\mathbf{v})$ i diagrama causal semi-Markovià $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.**Sortida:** Expressió de $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ en termes de $p(\mathbf{v})$, o bé una fallada de la forma $\text{FAIL}(\mathcal{G}', \mathcal{G}'[\mathbf{S}])$.

```

1: function ID( $\mathbf{Y}, \mathbf{X}, p, \mathcal{G}$ )
2:   if  $\mathbf{X} = \emptyset$  then
3:     return  $\sum_{\mathbf{v} \setminus \mathbf{Y}} p(\mathbf{v})$ 
4:   end if
5:   if  $\mathcal{V} \neq \text{An}_{\mathbf{Y}}^{\mathcal{G}}$  then
6:      $\mathbf{A} \leftarrow \text{An}_{\mathbf{Y}}^{\mathcal{G}}$ 
7:     return ID( $\mathbf{Y}, \mathbf{X} \cap \mathbf{A}, p(\mathbf{a}), \mathcal{G}[\mathbf{A}]$ )
8:   end if
9:    $\mathbf{W} \leftarrow (\mathcal{V} \setminus \mathbf{X}) \setminus \text{An}_{\mathbf{Y}}^{\mathcal{G}_{\bar{\mathbf{X}}}}$ 
10:  if  $\mathbf{W} \neq \emptyset$  then
11:    return ID( $\mathbf{Y}, \mathbf{X} \cup \mathbf{W}, p, \mathcal{G}$ )
12:  end if
13:  if  $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$  amb  $k > 1$  then
14:    return  $\sum_{\mathbf{v} \setminus (\mathbf{Y} \cup \mathbf{X})} \prod_{j=1}^k \text{ID}(\mathbf{S}_j, \mathcal{V} \setminus \mathbf{S}_j, p, \mathcal{G})$ 
15:  end if
16:  Sigui  $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}]\}$ 
17:  if  $C(\mathcal{G}) = \{\mathcal{G}\}$  then
18:    return FAIL( $\mathcal{G}, \mathcal{G}[\mathbf{S}]$ )
19:  end if
20:  if  $\mathcal{G}[\mathbf{S}] \in C(\mathcal{G})$  then
21:    return  $\sum_{\mathbf{s} \setminus \mathbf{Y}} \prod_{i \mid V_i \in \mathbf{S}} p(v_i \mid \mathbf{v}^{(i-1)})$ 
22:  end if
23:  if existeix  $\mathbf{S}'$  tal que  $\mathbf{S} \subset \mathbf{S}'$  i  $\mathcal{G}[\mathbf{S}'] \in C(\mathcal{G})$  then
24:     $q \leftarrow \prod_{i \mid V_i \in \mathbf{S}'} p(v_i \mid \mathbf{s}'^{(i-1)}, \mathbf{v}^{(i-1)} \setminus \mathbf{s}')$ 
25:    return ID( $\mathbf{Y}, \mathbf{X} \cap \mathbf{S}', q, \mathcal{G}[\mathbf{S}']$ )
26:  end if
27: end function

```

Lema 5.12 ([SP06, Lem. 5]). Sigui $\mathbf{A} = \text{An}_{\mathbf{Y}}^{\mathcal{G}}$. Aleshores $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$, calculat a partir de $p(\mathbf{v})$ en $\mathcal{G}$, coincideix amb $p(\mathbf{y} \mid \text{do}(\mathbf{x} \cap \mathbf{a}))$, calculat a partir de $p(\mathbf{a})$ en el subgraf induït $\mathcal{G}[\mathbf{A}]$.

Línies 9–12. Es fa servir el graf mutilat $\mathcal{G}_{\bar{\mathbf{X}}}$. El conjunt $\mathbf{W} = (\mathcal{V} \setminus \mathbf{X}) \setminus \mathbf{A}$, on en aquest cas $\mathbf{A} = \text{An}_{\mathbf{Y}}^{\mathcal{G}_{\bar{\mathbf{X}}}}$, conté variables que, un cop feta la intervenció sobre $\mathbf{X}$, tampoc poden afectar $\mathbf{Y}$. Per la regla 3 del *do-calculus*, intervenir sobre aquestes variables no canvia $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$.

Lema 5.13 ([SP06, Lem. 6]). Sigui $\mathbf{W} = (\mathcal{V} \setminus \mathbf{X}) \setminus \mathbf{A}$, on $\mathbf{A} = \text{An}_{\mathbf{Y}}^{\mathcal{G}_{\bar{\mathbf{X}}}}$. Aleshores, per a qualsevol valor $\mathbf{w}$ de $\mathbf{W}$, $p(\mathbf{y} \mid \text{do}(\mathbf{x})) = p(\mathbf{y} \mid \text{do}(\mathbf{x}), \text{do}(\mathbf{w}))$.

Notem que 5.13 no tria cap valor concret per a la intervenció $\text{do}(\mathbf{w})$: afirma que, per a qualsevol valor admès $\mathbf{w}$, la distribució $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ no en depèn.

Línies 13–15. Si després d'eliminar $\mathbf{X}$ el subgraf $\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]$ es trenca en diverses C-components, el problema es pot separar en factors independents en el sentit del lema 5.10. És el cor recursiu de l'algorisme:

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{y} \cup \mathbf{x})} \prod_{j=1}^k p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)).$$

Cada factor és una nova consulta causal del mateix tipus, però sobre una component més petita. Aquesta és la raó per la qual ID és recursiu: la identificació global es transforma en una col·lecció de problemes locals.

Lema 5.14 ([SP06, Lem. 4]). *Sigui $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$. Aleshores*

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{y} \cup \mathbf{x})} \prod_{j=1}^k p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)).$$

Si la condició anterior no es compleix, el graf $\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]$ (el subgraf amb variables que *no* estem intervenint) té una única C-component $\mathcal{G}[\mathbf{S}]$. A partir d'aquí apareixen tres possibilitats.

Línies 16–19. $\mathcal{G}$ és una única C-component, pel que no hi ha manera de descompondre el problema. Aquestes línies impliquen l'existència de dos C-boscós que formen un *hedge* i, pel criteri de 5.9, la consulta no es pot expressar només amb $p(\mathbf{v})$.

Teorema 5.15 ([SP06, Th. 6]; [SP08, Th. 39]). *Suposem que una crida de l'algorisme ID falla i retorna $\text{FAIL}(\mathcal{G}, \mathcal{G}[\mathbf{S}])$. Aleshores existeixen $\mathbf{X}' \subseteq \mathbf{X}$ i $\mathbf{Y}' \subseteq \mathbf{Y}$ tals que $\mathcal{G}$ i $\mathcal{G}[\mathbf{S}]$ contenen dos C-boscós $\mathcal{F}$ i $\mathcal{F}'$ que formen un hedge per a $p(\mathbf{y}' \mid \text{do}(\mathbf{x}'))$.*

Línies 20–22. Com $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}]\}$ per la línia 16, tenim $\mathcal{V} \setminus \mathbf{X} = \mathbf{S} \iff \mathbf{X} = \mathcal{V} \setminus \mathbf{S}$. En aquest cas tenim que $\mathcal{G}[\mathbf{S}] \in C(\mathcal{G})$, pel que les variables de $\mathbf{S}$ estan connectades per camins bidireccionals (però no tenen cap aresta bidireccional amb variables de $\mathbf{X}$). Com que la consulta només demana $\mathbf{Y} \subseteq \mathbf{S}$, marginalitzem les variables de $\mathbf{S} \setminus \mathbf{Y}$ i apliquem la segona part de 5.10:

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = p(\mathbf{y} \mid \text{do}(\mathbf{v} \setminus \mathbf{s})) = \sum_{\mathbf{s} \setminus \mathbf{y}} p(\mathbf{s} \mid \text{do}(\mathbf{v} \setminus \mathbf{s})) = \sum_{\mathbf{s} \setminus \mathbf{y}} \prod_{\{i \mid V_i \in \mathbf{S}\}} p(v_i \mid \mathbf{v}^{(i-1)}).$$

Línies 23–25. Cobreixen el cas intermedi. La component $\mathcal{G}[\mathbf{S}]$ no és maximal dins de $\mathcal{G}$, però està continguda en una C-component maximal més gran $\mathcal{G}[\mathbf{S}']$. Intuïtivament, aquesta línia diu que abans de resoldre $\mathbf{S}$ cal recuperar el bloc de confusió més gran; les intervencions sobre $\mathcal{V} \setminus \mathbf{S}'$ queden incorporades en q . El lema següent és la traducció de [SP06, Lem. 8]: l'última crida recursiva no canvia la distribució intervinguda que es vol calcular, sinó que la reexpressa dins de la C-component maximal $\mathbf{S}'$.

Lema 5.16 ([SP06, Lem. 8]). *En una crida de l'algorisme ID amb $\mathbf{X} \cap \mathbf{Y} = \emptyset$, suposem que $C(\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]) = \{\mathcal{G}[\mathbf{S}]\}$, que $\mathcal{G}[\mathbf{S}] \notin C(\mathcal{G})$, i que existeix $\mathbf{S}'$ amb $\mathbf{S} \subset \mathbf{S}'$ i $\mathcal{G}[\mathbf{S}'] \in C(\mathcal{G})$. Posem $\mathbf{X}_{\text{ext}} = \mathbf{X} \setminus \mathbf{S}'$ i $\mathbf{X}_{\text{int}} = \mathbf{X} \cap \mathbf{S}'$, i*

escriuim $\mathbf{x}_{\text{ext}} = \mathbf{x}_{\mathbf{x}_{\text{ext}}}$, $\mathbf{x}_{\text{int}} = \mathbf{x}_{\mathbf{x}_{\text{int}}}$ i $\mathbf{s}' = \mathbf{v}_{\mathbf{s}'}$. Definim:

$$\tilde{q}(\mathbf{s}') = \prod_{\{i | V_i \in \mathbf{S}'\}} p(v_i | \mathbf{s}'^{(i-1)}, \mathbf{v}^{(i-1)} \setminus \mathbf{s}').$$

Denotem per $\tilde{p}$ les probabilitats calculades en el problema reduït amb graf $\mathcal{G}[\mathbf{S}']$ i expressió d'entrada $\tilde{q}$. Aleshores

$$p(\mathbf{y} | \text{do}(\mathbf{x})) = \sum_{\mathbf{s} \setminus \mathbf{y}} \tilde{p}(\mathbf{s} | \text{do}(\mathbf{x}_{\text{int}})).$$

És a dir, la consulta $p(\mathbf{y} | \text{do}(\mathbf{x}))$ en el problema original coincideix amb la consulta reduïda que calcula $\text{ID}(\mathbf{Y}, \mathbf{X}_{\text{int}}, \tilde{q}, \mathcal{G}[\mathbf{S}'])$.

L'ALGORISME IDC

L'algorisme ID resol consultes de l'estil $p(\mathbf{y} | \text{do}(\mathbf{x}))$; ara bé, sovint la pregunta causal que ens ocupa té la forma $p(\mathbf{y} | \mathbf{z}, \text{do}(\mathbf{x}))$, és a dir, volem l'efecte d'una acció en un context on s'observen certes variables. Aquest condicionament pot fer identificable una consulta que sense condicionar no ho era, però també pot obrir dependències que dificultin la identificació [SP12]. L'algorisme IDC estén l>ID per resoldre aquests casos en dues fases [SP12; SP08]. Primer, examina les variables de $\mathbf{Z}$ una a una i aplica la regla 2 del *do-calculus* sempre que sigui possible, convertint observacions en intervencions quan el graf ho justifica. Quan ja no es pot fer cap conversió més, aplica la definició de probabilitat condicionada:

$$p(\mathbf{y} | \text{do}(\mathbf{x}), \mathbf{z}) = \frac{p(\mathbf{y}, \mathbf{z} | \text{do}(\mathbf{x}))}{p(\mathbf{z} | \text{do}(\mathbf{x}))}.$$

Aquesta igualtat s'entén per als valors $\mathbf{z}$ tals que $p(\mathbf{z} | \text{do}(\mathbf{x})) > 0$. D'aquesta manera, IDC redueix el problema condicionat a una crida d>ID sobre la distribució intervinguda conjunta i normalitza el resultat dividint per la marginal del denominador. La clau és que la part difícil torna a ser una consulta no condicionada; si aquesta conjunta no és identificable, tornem a obtenir un *hedge*.

El punt subtil és que aquesta primera fase no depèn de l'ordre arbitrari amb què mirem les variables de $\mathbf{Z}$. Les aplicacions successives de la regla 2 permeten traslladar a l'operador $\text{do}(\cdot)$ totes les variables de $\mathbf{Z}$ que satisfan la condició gràfica corresponent. Quan aquest conjunt maximal ja s'ha esgotat, les variables condicionants restants no poden eliminar-se per aquesta via, i IDC continua amb una crida a ID.

Lema 5.17 (Reducció del cas condicionat [SP12, Cor. 1 i Th. 6]). *Sigui $\mathbf{Z}_0 \subseteq \mathbf{Z}$ el conjunt maximal de variables condicionants que poden ser convertides en intervencions mitjançant aplicacions successives de la regla 2 del *do-calculus*, i sigui $\mathbf{z}_0 := \mathbf{z}_{\mathbf{Z}_0}$. Aleshores*

$$p(\mathbf{y} | \mathbf{z}, \text{do}(\mathbf{x})) = p(\mathbf{y} | \mathbf{z} \setminus \mathbf{z}_0, \text{do}(\mathbf{x} \cup \mathbf{z}_0)),$$

i la consulta condicionada original és identificable si, i només si, ho és la conjunta intervinguda

$$p(\mathbf{y}, \mathbf{z} \setminus \mathbf{z}_0 | \text{do}(\mathbf{x} \cup \mathbf{z}_0)).$$

Algorisme 2 IDC: identificació de $p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x}))$

Entrada: Conjunts disjunts $\mathbf{Y}, \mathbf{X}, \mathbf{Z} \subseteq \mathcal{V}$, $p(\mathbf{v})$ i diagrama causal semi-Markovià $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Sortida: Expressió de $p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x}))$ en termes de $p(\mathbf{v})$, o bé una fallada propagada per ID.

```

1: function IDC( $\mathbf{Y}, \mathbf{X}, \mathbf{Z}, p, \mathcal{G}$ )
2:   if existeix  $Z \in \mathbf{Z}$  tal que  $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{X}}, \mathbf{Z}}} Z \mid \mathbf{X}, \mathbf{Z} \setminus \{Z\}$  then
3:     return IDC( $\mathbf{Y}, \mathbf{X} \cup \{Z\}, \mathbf{Z} \setminus \{Z\}, p, \mathcal{G}$ )
4:   else
5:      $p' \leftarrow \text{ID}(\mathbf{Y} \cup \mathbf{Z}, \mathbf{X}, p, \mathcal{G})$ 
6:     return  $\frac{p'}{\sum_{\mathbf{y}'} p'}$ 
7:   end if
8: end function

```

Línies 2–3. La primera branca d'IDC aplica la regla 2 del *do-calculus*. Si existeix $Z \in \mathbf{Z}$ tal que $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{X}}, \mathbf{Z}}} Z \mid \mathbf{X}, \mathbf{Z} \setminus \{Z\}$, aleshores observar $Z = z$ és equivalent, per a aquesta consulta, a intervenir sobre Z amb valor z :

$$p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x})) = p(\mathbf{y} \mid \mathbf{z} \setminus \{z\}, \text{do}(\mathbf{x}), \text{do}(z)).$$

La condició gràfica es comprova en el graf mutilat $\mathcal{G}_{\bar{\mathbf{X}}, \mathbf{Z}}$. Les arestes bidirigides es conserven i continuen representant confusió latent. Per tant, la condició no afirma que Z sigui irrellevant en el graf original, sinó que, després de les intervencions ja imposades i condicionant per la resta de variables observades, substituir l'observació de Z per la intervenció $\text{do}(Z)$ no altera la distribució de $\mathbf{Y}$. En aquest cas, IDC trasllada Z de $\mathbf{Z}$ a $\mathbf{X}$ i repeteix el test fins que cap variable condicionant restant satisfà la condició.

Línies 4–6. Quan ja no queda cap variable condicionant que es pugui convertir en intervenció, entra en joc el lema 5.17. IDC demana a ID una expressió per a la conjunta intervinguda $p(\mathbf{y}, \mathbf{z} \mid \text{do}(\mathbf{x}))$. Si aquesta crida falla, IDC propaga la fallada; si té èxit, normalitza:

$$p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x})) = \frac{p(\mathbf{y}, \mathbf{z} \mid \text{do}(\mathbf{x}))}{\sum_{\mathbf{y}'} p(\mathbf{y}', \mathbf{z} \mid \text{do}(\mathbf{x}))}.$$

on $\sum_{\mathbf{y}'} p(\mathbf{y}', \mathbf{z} \mid \text{do}(\mathbf{x})) = p(\mathbf{z} \mid \text{do}(\mathbf{x})) > 0$. És important que la suma del denominador sigui sobre els valors de $\mathbf{Y}$, no sobre $\mathbf{Z}$: $\mathbf{z}$ és el context que mantenim constant. Així, la sortida d'ID per a la conjunta $p(\mathbf{y}, \mathbf{z} \mid \text{do}(\mathbf{x}))$ es transforma en una distribució condicionada normalitzant sobre el suport rellevant de $\mathbf{Y}$.

De manera similar a l'ID, els autors de l'IDC van demostrar que és correcte i complet a [SP12].

Teorema 5.18 (Correctesa i completesa d'IDC [SP12, Th. 7 i Th. 8]). **IDC** sempre acaba, quan retorna una expressió per a $p(\mathbf{y} \mid \mathbf{z}, \text{do}(\mathbf{x}))$ aquesta és correcta. És un algorisme complet.

PARADOXA DE SIMPSON

Useu la paradoxa de Simpson per separar tres qüestions: la reversió estadística que apareix en agregar estrats, les condicions algebraïques que la fan possible i la decisió causal de si cal ajustar o no per l'estrat.

La paradoxa de Simpson és una reversió d'associació produïda per agregació: una relació que té una direcció dins de cada estrat pot tenir la direcció contrària quan col·lapsem sobre els estrats (és a dir, quan fem una lectura agregada) [Sam93, p. 81]; [SW21, Sec. 2.1]. Simpson ho va il·lustrar amb taules $2 \times 2 \times 2$ i va remarcar que amalgamar taules pot canviar-ne la interpretació [Sim51, pp. 240–241]. El punt important, sobretot en un context causal, és que no hi ha una regla purament estadística que digui si la taula agregada és sempre preferible a les taules estratificades, o a l'inrevés [SW21, Secs. 3.3–3.4]; [Pea09, Sec. 6.1].

Definició 5.19 (Reversió d'associació [Sam93, p. 81]). En una taula $2 \times 2 \times k$, hi ha *reversió d'associació* quan la direcció de l'associació entre dues variables canvia en col·lapsar sobre una covariable Z .

Exemple 5.20. Considerem 52 pacients: 40 reben tractament i 12 formen el grup control. La taula mostra els resultats separats per gènere. És la mateixa mena de reversió que Pearl presenta a [Pea09, Ch. 6], però amb el signe invertit: aquí el tractament sembla pitjor en l'agregat i millor dins de cada estrat, mentre que la versió típica fa l'elecció oposada. Les dues formes són matemàticament equivalents, perquè intercanviar tractament/control o èxit/fracàs canvia el signe de l'associació però no el fenomen d'agregació.

	Població completa, $N = 52$			Homes (M), $N = 20$			Dones ($\neg M$), $N = 32$		
	Èxit (S)	Fracàs ($\neg S$)	Taxa	S	$\neg S$	Taxa	S	$\neg S$	Taxa
Tractament (T)	19	21	0.475 (47.5%)	8	5	0.615 (61.5%)	11	16	0.407 (40.7%)
Control ($\neg T$)	6	6	0.500 (50.0%)	4	3	0.571 (57.1%)	2	3	0.400 (40.0%)

Taula 5.1: Reversió d'associació en una taula $2 \times 2 \times 2$: el tractament té una taxa d'èxit superior dins de cada estrat, però inferior a la taula agregada. Variant amb signe invertit del cas artificial de Simpson [Sim51, pp. 240–241].

Si mirem només la població completa, el medicament sembla perjudicial: $\frac{19}{40} = 0.475$ (47.5%) dels pacients tractats es recuperen, davant de $\frac{6}{12} = 0.500$ (50.0%) dels pacients no tractats. En canvi, dins de cada estrat el signe s'inverteix: $\frac{8}{13} > \frac{4}{7}$ i $\frac{11}{27} > \frac{2}{5}$. Per tant, la lectura agregada i l'estratificada donen signes oposats per a la diferència de taxes d'èxit. Així, sembla que estratificant el tractament és més beneficiós que si no ho fem.

Per separar el fenomen estadístic de la seva interpretació causal, primer reescrivim l'exemple com una operació sobre taules de contingència, el llenguatge estadístic natural per a variables categòriques [Vil24a, Sec. 4.4]; cf. B.8. Sigui $D_i = (a_i, b_i, c_i, d_i)$ la taula 2×2 del i -èsim estrat, amb primera fila T i segona fila $\neg T$:

	Població $D = D_1 + D_2$		Subpoblació D_1		Subpoblació D_2	
	S	$\neg S$	S	$\neg S$	S	$\neg S$
Tractament (T)	$a_1 + a_2$	$b_1 + b_2$	a_1	b_1	a_2	b_2
Control ($\neg T$)	$c_1 + c_2$	$d_1 + d_2$	c_1	d_1	c_2	d_2

Taula 5.2: Representació abstracta d'una taula de contingència 2×2 amb dues subpoblacions. En l'exemple: $(a_1, b_1, c_1, d_1) = (8, 5, 4, 3)$ i $(a_2, b_2, c_2, d_2) = (11, 16, 2, 3)$.

Nota 5.21 ([GM87]). Sigui $D_i = (a_i, b_i, c_i, d_i)$, $i = 1, \dots, n$, la taula 2×2 de contingència corresponent a la i -èsima de n subpoblacions mútuament exclusives i conjuntament exhaustives (és a dir, que formen una *partició* de la població), tals que $a_i b_i c_i d_i \neq 0$ per a tot i . Sigui $N_i = a_i + b_i + c_i + d_i$ la mida de la mostra de la i -èsima subpoblació i $N = \sum_{i=1}^n N_i$ la mida de la mostra de tota la població.

L'amalgamació és un procés pel qual obtenim una taula agregada

$$\mathbf{A} = (A, B, C, D) = \left(\sum_i a_i, \sum_i b_i, \sum_i c_i, \sum_i d_i \right),$$

tal que $A + B + C + D = N$.

Ara bé, la reversió que hem vist en el nostre exemple és un cas particular d'un fenomen més ampli. [GM87, pp. 694–695] en diuen *paradoxa d'amalgamació*: una mesura d'associació de D pot quedar fora del rang de les mesures dels D_i .

Definició 5.22 (Mesura d'associació [GM87]). Una *mesura d'associació* és una funció α definida sobre taules 2×2 que pren el valor 0 quan no hi ha associació entre les dues variables de la taula. En el cas d'una taula $D_i = (a_i, b_i, c_i, d_i)$, aquesta situació correspon a

$$a_i d_i = b_i c_i.$$

Seguirem la convenció que $\alpha(D_i) > 0$ indica associació positiva i $\alpha(D_i) < 0$ associació negativa. Algunes de les més usades són [SW21]:

$$\begin{aligned} \pi_D &= \frac{a}{a+b} - \frac{c}{c+d} & \pi_Y &= \frac{ad-bc}{N^2} & \pi_R &= \log \left(\frac{a}{a+b} \cdot \frac{c+d}{c} \right) \\ \pi_W &= \log \left(\frac{a}{a+c} \cdot \frac{b+d}{b} \right) & \pi_O &= \log \frac{ad}{bc} & \pi_C &= \log \left(\frac{d}{c+d} \cdot \frac{a+b}{a} \right) \end{aligned}$$

[GM87, pp. 698–701] estudien, entre d'altres, diferències de proporcions, mesures de Yule, raons d'avantatges i pesos d'evidència; l'amalgamació no es comporta igual per a totes elles.

Definició 5.23 (Paradoxa d'amalgamació). Sigui α una mesura d'associació per a taules 2×2 i sigui $D = \sum_i D_i$ la taula agregada. Diem que hi ha *paradoxa d'amalgamació* si

$$\alpha(D) > \max_i \alpha(D_i) \quad \text{o bé} \quad \alpha(D) < \min_i \alpha(D_i).$$

La definició depèn de la mesura α . Per a una taula $D_i = (a_i, b_i, c_i, d_i)$, la comparació associativa de taxes d'èxit entre tractament i control queda determinada pel signe de

$$\frac{a_i}{a_i + b_i} - \frac{c_i}{c_i + d_i}.$$

Com que tots els denominadors són positius,

$$\frac{a_i}{a_i + b_i} > \frac{c_i}{c_i + d_i} \iff a_i(c_i + d_i) > c_i(a_i + b_i) \iff a_i d_i > b_i c_i.$$

Per tant, en aquest cas discret, la direcció de l'associació també es pot llegir amb el signe de $a_i d_i - b_i c_i$. En el nostre exemple, $8 \cdot 3 > 5 \cdot 4$, $11 \cdot 3 > 16 \cdot 2$, però a la taula agregada $19 \cdot 6 < 21 \cdot 6$.

Corol·lari 5.24 (Condicció d'enllaç doble [Sam93, pp. 81–83]). *En una taula candidata a reversió d'associació en el sentit de 5.19, no pot aparèixer reversió si la covariable Z no està doblement enllaçada amb (T, S) . És a dir, la reversió queda descartada si es compleix almenys una de les condicions següents:*

$$Z \perp\!\!\!\perp_P T, \quad Z \perp\!\!\!\perp_P S, \quad Z \perp\!\!\!\perp_P S \mid T, \quad Z \perp\!\!\!\perp_P T \mid S.$$

El doble enllaç és una condició estadística necessària per a aquesta mena de reversió, però no decideix cap pregunta causal. En l'exemple, el gènere està associat amb el tractament i amb el resultat: $P(T \mid M) = \frac{13}{20}$, mentre que $P(T \mid \neg M) = \frac{27}{32}$, i les taxes de recuperació també varien entre estrats. Això explica com pot aparèixer la reversió. No diu, encara, si cal agregar o estratificar. De manera similar, saber que hi ha paradoxa d'amalgamació en el sentit de 5.23 és un **diagnòstic estadístic** sobre la mesura α , **no una resposta causal**.

Exemple 5.25 (5.20, continuació). Seguint les condicions de reversió d'associació de Samuels per a taules $2 \times 2 \times k$ [Sam93, pp. 82–83], amb $D_1 = \text{homes}$ i $D_2 = \text{dones}$, posem

$$\begin{aligned} q_i &= P(T = 1 \mid Z = i), \\ p_i &= P(S = 1 \mid T = 0, Z = i), \\ \Delta_i &= P(S = 1 \mid T = 1, Z = i) - p_i. \end{aligned}$$

Aleshores

$$q_1 = \frac{13}{20}, \quad q_2 = \frac{27}{32}, \quad p_1 = \frac{4}{7}, \quad p_2 = \frac{2}{5} \implies \Delta_1 = \frac{8}{13} - \frac{4}{7} \approx 0.0440, \quad \Delta_2 = \frac{11}{27} - \frac{2}{5} \approx 0.0074.$$

És a dir, dins de tots dos estrats el tractament té associació positiva amb la recuperació. En canvi, a l'agregat: $\frac{19}{40} - \frac{6}{12} = -0.025$. Samuels [Sam93, pp. 82–83] dona condicions necessàries i suficients perquè es produeixi aquesta reversió en taules $2 \times 2 \times k$. En aquest exemple, la distribució del tractament entre estrats és prou desigual perquè l'associació marginal acabi tenint el signe contrari.

Nota 5.26 (Disseny de mostreig). Samuels distingeix entre reversió poblacional i reversió produïda pel disseny de mostreig. Amb mostreig aleatori simple, o amb alguns dissenys producte-binomial, les

frequències esperades no presenten reversió si la població original tampoc en presenta²¹; amb mostreig aparellat per estrats (*matching*), la reversió queda descartada tant a la mostra com a la taula [Sam93, p. 84, Corol·laris 2.7–2.8]. En una línia semblant, Good i Mittal [GM87, pp. 696–703] estudien quan l'amalgamació preserva el signe de l'associació: segons la mesura utilitzada, poden ser suficients condicions d'uniformitat per files, per columnes o per totes dues. Tot plegat mostra que la possibilitat de col·lapsar una taula depèn de tres elements: la mesura d'associació, el disseny de mostreig i la pregunta que es vol respondre.

La pregunta causal no és quina associació cal mirar, sinó quin estimand volem. La quantitat $P(S = 1 \mid T = t)$ és associativa: descriu la taxa d'èxit entre els individus observats amb $T = t$. En canvi, $P(S = 1 \mid \text{do}(T = t))$ és intervencional: descriu què passaria si forcéssim el tractament a valer t en la població. Pearl formula la selecció de covariants com un problema causal, no només estadístic [Pea09, Sec. 3.3]. La mateixa distinció es pot expressar dient que Z pot tenir rols causals diferents: causa comuna, mediador, descendent o col·lisionador [SW21, Secs. 3.2–3.4].

Proposició 5.27 (*Sure thing principle* [Pea09, Sec. 6.1.4]). *Suposem que Z pren un nombre finit de valors. Si, per a tot z amb probabilitat positiva,*

$$P(S = 1 \mid \text{do}(T = 1), Z = z) > P(S = 1 \mid \text{do}(T = 0), Z = z),$$

i la intervenció sobre T no canvia la distribució dels estrats,

$$P(Z = z \mid \text{do}(T = 1)) = P(Z = z \mid \text{do}(T = 0)) = P(Z = z),$$

aleshores

$$P(S = 1 \mid \text{do}(T = 1)) > P(S = 1 \mid \text{do}(T = 0)).$$

Exemple 5.28 (5.20, continuació). Suposem que Z és una variable pretractament que afecta tant l'assignació efectiva al tractament com el resultat, com representem a la figura 5.2.

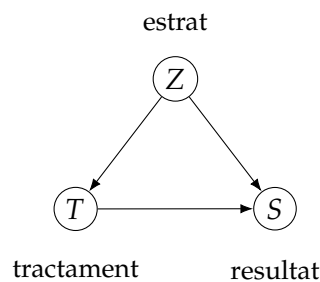


Figura 5.2: Exemple de Simpson quan l'estrat Z actua com a causa comuna de tractament i resultat.

El camí $T \leftarrow Z \rightarrow S$ és un camí *back-door* entre T i S . Com que Z no és descendent de T i bloqueja aquest camí, Z satisfà el criteri de la porta enrere respecte de (T, S) , en el sentit de 2.29. Pel teorema

²¹ És a dir, el mostreig aleatori per si sol no crea la paradoxa de Simpson en les freqüències esperades a menys que el fenomen ja existeixi en la població general que s'està estudiant

d'ajustament 2.31, l'efecte causal és identificable mitjançant

$$P(S = 1 \mid \text{do}(T = t)) = \sum_z P(S = 1 \mid T = t, Z = z)P(Z = z).$$

Per tant, l'efecte causal ajustat és:

$$P(S = 1 \mid \text{do}(T = 1)) - P(S = 1 \mid \text{do}(T = 0)) = \left(\frac{8}{13} \cdot \frac{20}{52} + \frac{11}{27} \cdot \frac{32}{52} \right) - \left(\frac{4}{7} \cdot \frac{20}{52} + \frac{2}{5} \cdot \frac{32}{52} \right) \approx 0.0215.$$

Fixem-nos que l'efecte és positiu, encara que l'associació marginal sigui negativa.

Observació 5.29. Aquesta conclusió depèn del model causal assumit. Si Z és un factor de confusió observat, podem ajustar per Z i això significa que podem tancar el camí de porta enrere. Si Z és un mediador, un descendent del tractament o un col·lisionador, el mateix ajustament podria eliminar part de l'efecte causal o introduir biaix. Pearl subratlla precisament aquest punt a [Pea09, Sec. 6.1.3]: una mateixa distribució observacional pot ser compatible amb estructures causals diferents, per exemple amb Z com a confusor, com a variable posttractament o com a variable que només actua sobre el tractament. La taula, doncs, no és capaç de resoldre aquesta qüestió per nosaltres perquè treballa a nivell associatiu; ens cal una hipòtesi causal [Pea09, Sec. 3.3]; [SW21, Secs. 3.3–3.4].

5.4

ADMISSIONS A BERKELEY (1973)

Apliquem la part teòrica a les admissions de Berkeley separant tres plans: l'associació estadística marginal i estratificada, la formulació de preguntes causals amb $\text{do}(\cdot)$ i la identificabilitat d'aquestes preguntes amb ID/IDC.

Per tancar el treball, apliquem la teoria desenvolupada al cas clàssic de les admissions de postgrau a Berkeley l'any 1973 [BHO75]. L'estudi original analitza 12 763 sol·licituds a 101 unitats de postgrau, segregades per gènere i departament, amb l'objectiu de determinar si hi ha hagut algun tipus de biaix de selecció que involucri el gènere.

No podem decidir amb una sola taula si les dones han patit discriminació respecte als homes en les admissions de postgrau a Berkeley. Hem de mostrar per què aquesta pregunta necessita un estimand i un model causal explícits. La versió completa amb els càlculs, el codi Python, les taules i les figures, així com l'anàlisi estadística i altres explicacions detallades, es troba al fitxer `berkeley.ipynb` del repositori del treball²².

Aquí treballarem amb `UCBAdmissions`, la versió reduïda més habitual, que conserva els sis departaments més grans i agrega 4526 sol·licituds segons admissió Y , gènere G i departament D . Per a la lectura estadística fem servir les eines de taules de contingència recollides a la secció B.8. En la taula marginal, la taxa d'admissió dels homes és superior a la de les dones: $\frac{1198}{2691} \approx 0.4452$ (44.52%) contra $\frac{557}{1835} \approx 0.3035$ (30.35%). El contrast χ -quadrat de Pearson d'independència entre G i Y rebutja clarament la hipòtesi

²² <https://github.com/mariovilar/TFG/blob/main/berkeley.ipynb>.

nul·la ($X^2 \approx 92.205$, $p \approx 7.81 \cdot 10^{-22}$). Usant (B.1), veiem que en aquesta taula 2×2 , el recompte esperat més petit és el de dones admeses, $\frac{1835 \cdot 1755}{4526} \approx 711.5$, molt per sobre del llindar pràctic de 5 indicat a B.98. El punt delicat de l'exemple no és el càlcul del p -valor marginal, sinó en com l'interpretem en presència d'estratificació per departament. L'OR marginal és

$$\widehat{OR}_{\text{marg}} = \frac{1198 \cdot 1278}{1493 \cdot 557} \approx 1.84.$$

Aquesta lectura que estem fent, però, és només a nivell associatiu: descriu com estan distribuïdes les admissions observades, però encara no estima cap intervenció.

Quan estratifiquem per departament, la interpretació canvia. Les taxes d'admissió varien molt entre departaments: els departaments A i B admeten al voltant de 64% i 63% de les sol·licituds, mentre que el departament F només n'admet 6.4%. A més, homes i dones no es distribueixen igual entre departaments: aproximadament 30.7% dels homes sol·liciten el departament A, però només 0.059 (5.9%) de les dones ho fan; en canvi, les dones es concentren més en departaments més selectius. La taxa marginal és la mitjana ponderada de les condicionades següent:

$$P(Y = 1 \mid G = g) = \sum_d P(Y = 1 \mid G = g, D = d) P(D = d \mid G = g),$$

que és una aplicació directa de la probabilitat total condicionada [San99, Sec. 1.5]. Per això, dues poblacions amb taxes estratificades semblants poden tenir taxes marginals diferents si els pesos $P(D = d \mid G = g)$ són diferents. Aquesta és justament l'explicació qualitativa que es destaca a [BHO75, pp. 399–400]: la proporció de dones sol·licitants tendeix a ser més alta en departaments difícils d'entrar.

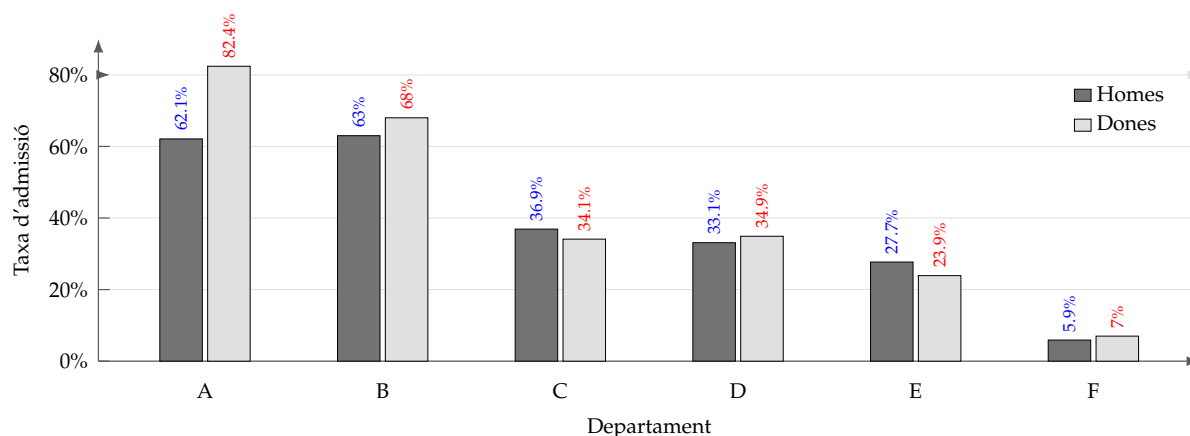


Figura 5.3: Taxa d'admissió per gènere i departament en UCBAmissions.

El cas no és una reversió de Simpson perfecta en tots els estrats, perquè els departaments C i E mantenen la direcció d'associació (és a dir, el signe de la mesura α emprada). Tot i així, mostra el fenomen característic de la paradoxa: l'associació marginal canvia significativament quan condicionem per una variable estructural. Amb la notació $T_i = (a_i, b_i, c_i, e_i)$ per a la taula 2×2 de cada departament, el signe de $a_i e_i - b_i c_i$ afavoreix els homes a la taula agregada, però afavoreix les dones als departaments A, B, D i F.

Per resumir l'evidència estratificada, usem l'estimador de Mantel–Haenszel de l'OR comuna i el contrast de Cochran–Mantel–Haenszel de les definicions B.100 i B.101. Si al departament d escrivim la taula 2×2 com

	$Y = 1$	$Y = 0$
$G = \text{home}$	a_d	b_d
$G = \text{dona}$	c_d	e_d

l'OR departamental és $\widehat{OR}_d = \frac{a_d e_d}{b_d c_d}$. En les dades, l'OR comuna de Mantel–Haenszel és $\widehat{OR}_{MH} \approx 0.905$, amb interval de confiança $[0.772, 1.060]$ i $p \approx 0.217$; així, després d'estratificar no queda evidència clara d'una associació comuna forta en favor dels homes. En canvi, el test de Breslow–Day–Tarone dona $p \approx 0.002$, de manera que la hipòtesi que l'odds ratio entre gènere i admissió és el mateix dins de tots els departaments és poc plausible [BD80]. Estratificar ajuda, però resumir els sis departaments amb una única OR també és una simplificació.

PREGUNTES CAUSALS

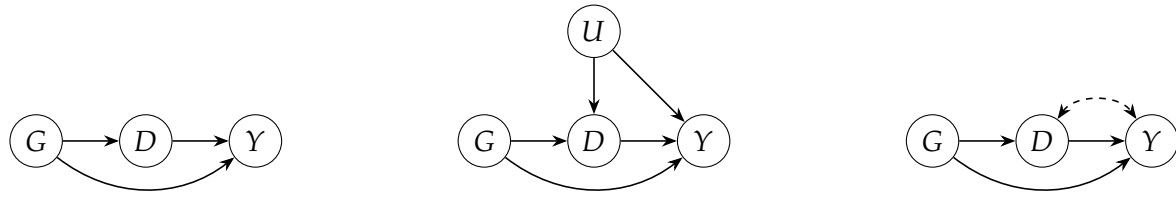
La pregunta «hi ha discriminació?» no té una única traducció matemàtica. En aquesta memòria considerem quatre consultes:

1. $P(Y \mid \text{do}(G = g))$: quina seria la taxa d'admissió si intervinguéssim per fer que tots els candidats fossin del gènere g ?
2. $P(Y \mid \text{do}(G = g), D = d)$: si agafem les persones que van sol·licitar entrar al departament d (ho observem: $D = d$), què passaria amb la seva admissió si només intervinguéssim en el seu gènere ($\text{do}(G = g)$)? Aquesta pregunta busca saber si, observada la tria del departament, el comitè d'aquell departament específic jutja de manera diferent depenent del gènere.
3. $P(Y \mid \text{do}(G = g), \text{do}(D = d))$: obliguem algú a ser del gènere g i a més l'obliguem a sol·licitar plaça al departament d . Serveix per aïllar el biaix institucional d'un departament, ignorant completament les tendències i preferències prèvies dels sol·licitants.
4. $P(Y \mid \text{do}(D = d))$: si simplement obliguéssim tota la població a demanar plaça al departament d , quina seria l'admissió global?

Aquestes quatre quantitats, com és evident, no tenen el mateix contingut causal. La figura 5.4 separa dues hipòtesis causals. A la subfigura (a) suposem un SCM Markovià $\mathcal{M}_0$ sense confusor latent entre departament i admissió; per tant, en la notació dels capítols 1 i 2 tenim $\mathcal{G}_{\mathcal{M}_0}^{\rightarrow} = \mathcal{G}_{\mathcal{M}_0}^{\mathcal{L}_1} = \mathcal{G}_{\mathcal{M}_0}$.

Les subfigures (b) i (c) no són una altra manera de dibuixar $\mathcal{M}_0$, sinó un segon SCM $\mathcal{M}_1$ amb les mateixes variables observades i un factor latent U entre D i Y . En (b) es mostra $\mathcal{G}_{\mathcal{M}_1}^+$, on el confusor apareix explícitament; en (c) es mostra el diagrama causal $\mathcal{G}_{\mathcal{M}_1}$ induït per la projecció latent sobre les variables observades.

En la CBN Markoviana de la figura 5.4a, totes aquestes consultes són identificables. En canvi, en el segon model, la projecció latent conté el *bow* $D \rightarrow Y$ i $D \leftrightarrow Y$. Aleshores les consultes que intervenen el departament deixen de ser identificables només amb la distribució observacional: el subgraf $\mathcal{H} = \mathcal{G}[\{D, Y\}]$ conté un *hedge* format per $\mathcal{F} = (\{D, Y\}, \{D \rightarrow Y, D \leftrightarrow Y\})$ i $\mathcal{F}' = (\{Y\}, \emptyset)$.



(a) Diagrama causal Markovià.

(b) Confusor explícit.

(c) Projectió latent.

Figura 5.4: Tres diagrames per a dues hipòtesis causals sobre el cas Berkeley. La subfigura (a) representa un SCM Markovià $\mathcal{M}_0$. Les subfigures (b) i (c) corresponen a un altre SCM $\mathcal{M}_1$ amb confusió latent entre D i Y : (b) en mostra el graf augmentat $\mathcal{G}_{\mathcal{M}_1}^+$ i (c) en mostra el diagrama causal $\mathcal{G}_{\mathcal{M}_1}$ obtingut per projecció latent sobre les variables observades.

Fem l'execució amb el paquet Python `causaleffect`, que implementa ID i, quan rep una consulta condicionada, IDC [PVP21]. El fragment conceptualment rellevant del notebook és:

```

1 DAG = createGraph(["G->D", "D->Y", "G->Y", "D<->Y"])
2
3 ID({"Y"}, {"G"}, DAG)
4 ID({"Y"}, {"G"}, DAG, cond={"D"})
5
6 ID({"Y"}, {"G", "D"}, DAG)
7 ID({"Y"}, {"D"}, DAG)

```

Les dues primeres consultes són identificables; les dues últimes haurien d'arribar al cas de no-identificabilitat. La taula 5.3 resumeix l'estimand que hauríem d'obtenir quan la consulta és identificable i quan no ho és.

Consulta	Lectura	CBN Markoviana	CBN semi-Markoviana amb $D \leftrightarrow Y$
$P(Y \mid \text{do}(G = g))$	Què canviaria en l'admissió si intervinguéssim el gènere a g .	$P(Y \mid G = g).$	$P(Y \mid G = g).$
$P(Y \mid \text{do}(G = g), D = d)$	Què canviaria, a més, dins dels sol·licitants que han triat d .	$P(Y \mid G = g, D = d).$	$P(Y \mid G = g, D = d).$
$P(Y \mid \text{do}(G = g), \text{do}(D = d))$	Què passaria intervenint alhora el gènere i el departament.	$P(Y \mid G = g, D = d).$	no identificable.
$P(Y \mid \text{do}(D = d))$	Quina admissió tindriem si tothom sol·licités d .	$\sum_g p(g)P(Y \mid D = d, G = g).$	no identificable.

Taula 5.3: Comparació entre les consultes causals i els estimands obtinguts al notebook.

En el nostre graf G no té pares; per això observar $G = g$ i intervenir formalment sobre $G = g$ són equivalents. La diferència substantiva apareix amb D : observar que una sol·licitud és al departament $D = d$ també pot informar sobre variables latents que n'han condicionat l'elecció, mentre que $\text{do}(D = d)$ talla aquest mecanisme.

En la CBN Markoviana, la darrera fila de la taula és una estandardització per gènere:

$$P(Y = 1 \mid \text{do}(D = d)) = \sum_g P(G = g)P(Y = 1 \mid D = d, G = g),$$

amb pesos de gènere estimats per $\hat{P}(G = \text{home}) = \frac{2691}{4526}$ i $\hat{P}(G = \text{dona}) = \frac{1835}{4526}$. Si substituïm també les probabilitats condicionals per les freqüències observades dins de cada departament, obtenim l'estimador:

$$\hat{\theta}_d = \frac{2691}{4526} \hat{P}(Y = 1 \mid D = d, G = \text{home}) + \frac{1835}{4526} \hat{P}(Y = 1 \mid D = d, G = \text{dona}).$$

Per als sis departaments de UCBAmissions, això dona:

d	A	B	C	D	E	F
$\hat{\theta}_d$	0.7031	0.6505	0.3576	0.3384	0.2620	0.0636

Aquests valors no s'han de traslladar a la CBN semi-Markoviana amb $D \leftrightarrow Y$, perquè allà la mateixa consulta no és identificable.

Observació 5.30 (El *bow* i l'error de causal effect). Com ja hem explicat, les consultes amb $\text{do}(D)$ en la CBN semi-Markoviana haurien de retornar FAIL. El subproblema essencial és

$$\mathcal{H} = \mathcal{G}[\{D, Y\}], \quad D \rightarrow Y, \quad D \leftrightarrow Y.$$

Si intervenim D , aleshores $C(\mathcal{H}[\{Y\}]) = \{\mathcal{H}[\{Y\}]\}$ i, alhora, $C(\mathcal{H}) = \{\mathcal{H}\}$ perquè D i Y formen una única C-component per l'aresta bidirigida. És exactament el cas de fallada de l'algorisme ID: la parella

$$\mathcal{F} = (\{D, Y\}, \{D \rightarrow Y, D \leftrightarrow Y\}), \quad \mathcal{F}' = (\{Y\}, \emptyset)$$

és un hedge arrelat en Y que testimonia la no-identificabilitat [SP06; SP08].

En executar-ho, però, *causal effect* no retorna aquest FAIL, sinó l'error d'ígraph «Edge counts should be non-negative, found -1». Si anem al codi o a la documentació [PVP21, Sec. 4.2] veiem que una aresta bidirigida es codifica internament mitjançant dues arestes dirigides auxiliars amb un atribut de confusió. En presència del *bow* $D \rightarrow Y$ i $D \leftrightarrow Y$, la mutilació associada a $\text{do}(D = d)$ elimina només una part de la bidirigida, i després la reconstrucció de les parts dirigida i bidirigida a partir de la matriu d'adjacència queda malformada. La conclusió teòrica continua sent la mateixa (és a dir, no hi ha estimand observacional per a aquestes consultes en aquest graf), ja que arreglant-ho s'arribaria al FAIL en ambdós casos.

En definitiva, les dades agregades no proven si hi ha hagut discriminació o no; tampoc proven que no n'hi hagi pogut haver. El que mostren és que la taula marginal, les taules estratificades i les consultes intervencionals responen preguntes diferents. La marginal resumeix una disparitat global entre sol·licitants homes i dones; l'estratificació per departament separa una part important de la composició de les sol·licituds; i les expressions amb $\text{do}(\cdot)$ obliguen a precisar quin mecanisme volem imaginar modificat.

Per això no n'hi ha prou amb triar entre «agregar» o «condicionar»: cal dir quin escenari causal fa legítima aquesta tria.

El cas de Berkeley també il·lustra que una resposta negativa de l'algorisme no és un fracàs de l'anàlisi, sinó informació sobre els límits del que les dades observacionals poden determinar sota unes hipòtesis concretes. En la CBN Markoviana obtenim estimands senzills, mentre que en la CBN semi-Markoviana amb $D \leftrightarrow Y$ les consultes que intervenen el departament queden bloquejades pel *hedge*. Aquesta diferència no es veu només mirant percentatges, sinó formulant la pregunta en el llenguatge causal adequat. Aquesta és la lliçó aplicada de la paradoxa de Simpson: canviar de nivell d'agregació pot canviar l'associació, però decidir quin nivell respon una pregunta causal exigeix hipòtesis causals explícites.

Conclusions

Començàvem aquesta memòria explicant que Hume entenia la causalitat fonamentalment a través de la regularitat. Això plantejava una tensió molt senzilla de formular i molt difícil de resoldre: les dades ens mostren regularitats, però les preguntes causals que nosaltres ens formulem parlen d'intervencions, explicacions i mons que no hem observat.

Hem vist de primera mà que la distància entre una associació estadística i una afirmació causal no es pot salvar només amb més dades. Una distribució observacional pot mostrar correlacions, independències i patrons d'associació molt rics, però no pot, per si mateixa, fer afirmacions causals. Per moure'ns d'una capa a l'altra necessitem hipòtesis, coneixement extern i estructura causal. Per això l'objectiu principal de la memòria ha estat construir un llenguatge matemàtic capaç de fer explícites aquestes hipòtesis i de determinar quan una pregunta causal pot traduir-se en una quantitat estimable.

Hem separat els tres nivells de la jerarquia causal de Pearl: associació, intervenció i contrafactuals. Aquesta distinció ha servit de brúixola per a tot el treball, perquè mostra per què una distribució observacional pot respondre preguntes del primer nivell, però no determina per si sola què passaria si modifiquéssim el sistema. Hem formalitzat les eines gràfiques bàsiques —xarxes bayesianes, d -separació, propietats de Markov, diagrames causals i models causals estructurals— fins arribar a la idea central que un diagrama causal representa no només dependències, sinó també els mecanismes: les funcions estructurals.

Els criteris *back-door* i *front-door*, el *do-calculus* i, finalment, els algorismes ID i IDC, han permès precisar quan una intervenció pot reescriure's en termes observacionals. Aquesta és una de les conclusions conceptuals més importants del treball: identificar un efecte no vol dir estimar-lo encara, sinó saber que la pregunta causal està determinada, al seu torn, per altres preguntes. La diferència és que aquestes noves preguntes sí que poden respondre's amb les dades disponibles, perquè viuen dins la capa associativa $\mathcal{L}_1$.

El model de resultats potencials de Neyman–Rubin és l'altra cara de la moneda de la causalitat i ens ha servit com el contrapunt matemàtic perfecte per a aquesta memòria. A la meua manera d'entendre, la bellesa d'aquest treball rau justament en el fet que acabem el tercer capítol posant en comú dos llenguatges que venen d'escoles completament diferents: la matemàtica-estadística i la informàtica.

Allà on Pearl escriu intervencions sobre variables, Neyman–Rubin compara resultats potencials d'una mateixa unitat sota tractaments diferents. El problema fonamental de la inferència causal, els estimands com SATE, PATE o PATT, la consistència, la ignorabilitat i la positivitat mostren que aquesta formulació és especialment natural per pensar en dissenys experimentals i observacionals. I, com dèiem més amunt, la comparació posterior amb els SCMs i la derivació contrafactual de la fórmula *front-door* han posat la cirereta al pastís, tot tancant el cercle: els dos llenguatges no competeixen, sinó que il·luminen parts diferents d'una mateixa estructura.

També hem vist que en un experiment aleatoritzat, el mecanisme d'assignació dona una justificació per poder comparar grups sense risc de biaix de confusió per variables no observades. En un estudi observacional, però, no tenim aquesta garantia, de manera que ens veiem obligats a dependre d'hipòtesis fermes sobre com s'ha produït l'assignació i quines covariables hi influeixen. El valor de propensió, l'equilibratge i la ponderació inversa, per tant, només organitzen la comparació quan les hipòtesis que la fan legítima ja són defensables i no ens donen peu per elles mateixes a fer afirmacions causals.

Hem aplicat tot aquest coneixement a les admissions de Berkeley, que és un cas particular molt conegut de la paradoxa de Simpson: la taula marginal suggereix una cosa, mentre que l'estratificació per departament en suggereix una altra, i les consultes causals canvien segons el diagrama causal assumit. En particular, sota una CBN Markoviana algunes preguntes tenen estimands observacionals senzills, mentre que en una CBN semi-Markoviana amb confusió latent entre departament i admissió les consultes que intervenen el departament deixen de ser identificables. La conclusió, per tant, no és que les dades no diguin res, sinó que no diuen el mateix davant de totes les preguntes i tots els escenaris possibles. En causalitat, formular la pregunta correcta i construir el model adequat ja és una gran part de la resposta.

RISCOS I ÈTICA

Un dels riscos principals en l'anàlisi causal és confondre predicció amb decisió. En la terminologia de la jerarquia causal, passar de l'esglaó associatiu $\mathcal{L}_1$ a l'intervencional $\mathcal{L}_2$ sense un model causal és justament el salt que no és justificable en general [Bar+22]. Aquest punt és especialment rellevant en sistemes d'IA: un model predictiu pot aprendre molt bé patrons observats i, amb tot, fallar quan s'utilitza per recomanar intervencions o polítiques en un entorn diferent d'aquell on ha estat entrenat. La capacitat d'aprendre, processar i aprofitar la informació causal és crucial per desenvolupar una IA més general i abordar més reptes fora d'entorns controlats. Disciplines com el *causal reinforcement learning* estan començant a explorar aquesta intersecció [BZL24], però encara queda molt camí per recórrer. Altres línies recents demostren que la causalitat també és clau en el modelatge generatiu per entendre com es relacionen les formes de raonament causal i neuronal.

A més, la causalitat és un essencial per fer afirmacions científiques i hauria de constituir la base de la disciplina de la ciència de dades. Des de la disciplina de *causal data science* s'estan fent esforços en aquesta direcció [BP16]. Així doncs, cal separar de manera estricta la identificabilitat teòrica de la validesa del model. Que un efecte sigui identificable en un graf no vol dir que el graf representi correctament la realitat, ni que la mostra sigui un reflex fidel de la població objectiu. El biaix de confusió, el biaix de selecció, els canvis en la distribució de les dades (*dataset shifts*) i els problemes de validesa externa (o transportabilitat) apareixen precisament quan hem de fer afirmacions intervencionals i contrafactuals a partir de col·leccions de dades heterogènies i esbiaixades. En aquests casos, una fórmula causal pot ser formalment correcta dins del model matemàtic i, alhora, completament inadequada si s'aplica fora del seu domini de validesa.

En aplicacions amb persones els estimands tampoc són neutres, i ens trobem amb reptes ètics directament lligats als problemes metodològics anteriors. Decidir si interessa el PATE, el PATT, un CATE... pot canviar quins grups queden beneficiats o perjudicats per una decisió. Aquí prenem una rellevància especial el paper de la causalitat en les explicacions i en marcs com el *causal fairness analysis*, dissenyats per avaluar l'equitat. Per això, a més de comprovar la identificabilitat, cal explicitar la població objectiu, els mecanismes de selecció possibles, les covariants disponibles i la interpretació real del tractament. La formalització causal no pot substituir mai el judici ètic, però proporciona el llenguatge per no amagar hipòtesis sota l'aparença d'una predicció neutral.

M'agradaria acabar amb una idea tan senzilla com colpidora: **la clau no són les dades en si mateixes, sinó com s'estructuren**. Ja ho va dir Judea Pearl a [PM18]: «Data are dumb». Decidir què es mesura i, sobretot, què s'omet; triar els termes adequats; i filar un relat determinat pot conduir qui ho llegeix, de forma gairebé imperceptible, cap a una conclusió concreta. En un moment de la història on les dades estan per tot arreu, cal ser més crítics que mai amb tota la informació que rebem.

TREBALL FUTUR

Aquest treball s'ha centrat a construir el llenguatge necessari per entendre la identificació causal i la connexió entre Pearl i Neyman–Rubin. Aquest punt d'arribada deixa obertes diverses línies que queden fora de l'abast d'una introducció teòrica, però que en serien continuacions naturals.

La primera seria aprofundir en variants del problema d'identificació més enllà dels efectes intervencionals tractats amb ID i IDC. En aquest treball hem vist com els *hedges* caracteritzen la no-identificabilitat en el cas semi-Markovià recursiu [SP06; SP08], però hi ha consultes més riques —efectes específics de camí, plans condicionals o distribucions contrafactuals— que requereixen llenguatges i algorismes addicionals. En aquesta direcció, els SWIGs i el *po-calculus* ofereixen una via natural per continuar el pont entre DAGs i resultats potencials.

En particular, els algorismes ID* i IDC* de Shpitser i Pearl obren la continuació contrafactual d'aquest camí: parteixen de grafs de mons paral·lels i grafs contrafactuals per decidir quan una consulta del tercer nivell pot expressar-se en termes de distribucions intervencionals [SP07]. Això encaixa amb una de les intuïcions recurrents de la memòria: fins i tot quan la pregunta ja no parla només d'intervenir, sinó d'imaginar mons incompatibles amb el que s'ha observat, la resposta continua depenent de quina estructura causal estem disposats a assumir.

Una segona continuació seria desenvolupar amb més profunditat els mètodes d'estimació per a estudis observacionals. Aquí només hem introduït la subclassificació i la ponderació inversa com a portes d'entrada, però el mateix marc permet estudiar aparellament (*matching*), regressió, ajustament per covariants i combinacions diferents d'aquests procediments. Una continuació natural seria comparar aquests enfocaments i analitzar com canvien les conclusions quan es modifica el model de propensió o el model de resultat [IR15, Cap. 13–14].

També caldria ampliar el treball cap a situacions on la ignorabilitat no és defensable. Les variables instrumentals en són un exemple important: introdueixen una variable associada amb el tractament, però que només pot afectar el resultat a través d'aquest tractament, sota les hipòtesis del model [IR15, Cap. 23–25]. Aquesta direcció és rellevant perquè moltes aplicacions reals inclouen confusió no observada que no es pot resoldre simplement afegint covariants.

Finalment, una extensió natural seria considerar tractaments longitudinals i règims dinàmics. En moltes aplicacions pràctiques, l'objectiu no és avaluar una intervenció estàtica puntual, sinó una seqüència de decisions adaptades a la informació que s'actualitza al llarg del temps. Això porta al marc teòric de Robins, on eines com la g -fórmula permeten estimar efectes causals en presència de confusors dependents del temps que, alhora, han estat afectats per tractaments previs. Aquest enfocament, combinat amb l'emulació d'un assaig objectiu (*target trial emulation*) [HR20, Caps. 19 i 21, Sec. 22.3], connecta la inferència causal amb problemes seqüencials en epidemiologia i polítiques públiques. En una direcció complementària, també seria interessant estudiar procediments d'estimació causal independents de l'*estimand*, com els *Deep Causal Graphs*, capaços d'aprendre models estructurals flexibles i respondre múltiples consultes identificables sobre el mateix sistema sense redissenyar un estimador específic per a cada pregunta [PV22; PV20; Par23].

Demostracions seleccionades

Aquest capítol recull les demostracions que no hem inclòs en el cos principal del text. La idea és mantenir fluida l'exposició dels capítols anteriors sense perdre els detalls necessaris per verificar cada resultat.

Cada demostració indica el resultat corresponent, de manera que es pugui localitzar sense haver de rellegir tot el capítol.

A.1

MODELS CAUSALS I GRÀFICS DIRIGITS

Demostració de 1.9. Recordem que hem fixat, sense pèrdua de generalitat, un ordre topològic $<$ sobre les variables, compatible amb $\mathcal{G}$, i que les hem indexat seguint aquest ordre. Escrivem la prova amb PMFs; en el cas general, es pot fer una derivació anàloga amb integrals respecte de la mesura dominant producte de 1.1.

$\Rightarrow$ Utilitzem la regla de la cadena per a probabilitats:

$$p(v_1, \dots, v_n) = \prod_{i=1}^n p(v_i \mid v_1, \dots, v_{i-1}). \quad (\text{A.1})$$

Considerem el factor $p(v_i \mid v_1, \dots, v_{i-1})$ per a un cert $i \in I$. Com que l'ordre és topològic, tots els predecessors de V_i són no-descendents seus. Per la propietat local de Markov:

$$V_i \perp\!\!\!\perp_P \text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}} \mid \text{Pa}_i^{\mathcal{G}}.$$

En particular, per descomposició de la independència condicional: $V_i \perp\!\!\!\perp_P \text{Pred}_i \setminus \text{Pa}_i^{\mathcal{G}} \mid \text{Pa}_i^{\mathcal{G}}$. Per tant, per als valors on els condicionals estan ben definits, $p(v_i \mid v_1, \dots, v_{i-1}) = p(v_i \mid \text{pa}_i^{\mathcal{G}})$. Substituint aquesta igualtat en (A.1) per a cada i , obtenim la factorització sobre $\mathcal{G}$.

$\Leftarrow$ Suposem que P factoritza sobre $\mathcal{G}$, és a dir,

$$p(\mathbf{v}) = \prod_{j=1}^n p(v_j \mid \text{pa}_j^{\mathcal{G}}).$$

Fixem V_i i escrivim De_i per als seus descendents; la notació $\sum_{\text{de}_i}$ indica la suma sobre totes les assignacions d'aquests descendents. Volem veure que

$$p(v_i \mid \text{nd}_i) = p(v_i \mid \text{pa}_i^{\mathcal{G}}),$$

que és equivalent a (L). Marginalitzant els descendents de V_i en la factorització anterior,

$$p(v_i, \text{nd}_i) = \sum_{\text{de}_i} p(v_i | \text{pa}_i^{\mathcal{G}}) \prod_{V_j \in \text{ND}_i} p(v_j | \text{pa}_j^{\mathcal{G}}) \prod_{V_j \in \text{De}_i} p(v_j | \text{pa}_j^{\mathcal{G}}).$$

Si $V_j \in \text{ND}_i$, aleshores cap pare de V_j no pot ser V_i ni cap descendent de V_i . En efecte, si algun pare de V_j fos V_i , tindríem un camí dirigit $V_i \rightarrow V_j$, i V_j seria descendent de V_i . I si algun pare de V_j fos un descendent D de V_i , tindríem un camí dirigit $V_i \rightarrow \dots \rightarrow D \rightarrow V_j$, de manera que V_j també seria descendent de V_i . Per tant, els factors corresponents als no-descendents no depenen ni de v_i ni de les variables sobre les quals estem sumant. Per tant, els factors dels no-descendents i el terme $p(v_i | \text{pa}_i^{\mathcal{G}})$ surten del sumatori:

$$p(v_i, \text{nd}_i) = p(v_i | \text{pa}_i^{\mathcal{G}}) \prod_{V_j \in \text{ND}_i} p(v_j | \text{pa}_j^{\mathcal{G}}) \sum_{\text{de}_i} \prod_{V_j \in \text{De}_i} p(v_j | \text{pa}_j^{\mathcal{G}}).$$

L'últim sumatori val 1: és la marginalització successiva dels nuclis condicionals dels descendents, començant pels nodes terminals. Així,

$$p(v_i, \text{nd}_i) = p(v_i | \text{pa}_i^{\mathcal{G}}) \prod_{V_j \in \text{ND}_i} p(v_j | \text{pa}_j^{\mathcal{G}}).$$

Sumant ara sobre v_i obtenim

$$p(\text{nd}_i) = \prod_{V_j \in \text{ND}_i} p(v_j | \text{pa}_j^{\mathcal{G}}),$$

i, dividint les dues expressions (sempre que $p(\text{nd}_i) > 0$), queda $p(v_i | \text{nd}_i) = p(v_i | \text{pa}_i^{\mathcal{G}})$. Com que la propietat se satisfà per un $i \in I$ arbitrari, P satisfà la propietat local de Markov. ■

Demostració de 1.21.

⇒ En el graf $\mathcal{G}$, la propietat local de Markov afirma que $V_i \perp_{\mathcal{G}} \text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}} | \text{Pa}_i^{\mathcal{G}}$ per a tot i . Com que $\mathcal{G}$ és un I -mapa de P , tenim que $I(\mathcal{G}) \subseteq I(P)$ i, per tant:

$$V_i \perp_{\mathcal{G}} \text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}} | \text{Pa}_i^{\mathcal{G}} \implies V_i \perp_P \text{ND}_i \setminus \text{Pa}_i^{\mathcal{G}} | \text{Pa}_i^{\mathcal{G}}.$$

Per tant, P compleix la propietat local de Markov i, per 1.9, factoritza sobre $\mathcal{G}$.

⇐ Suposem que P factoritza sobre $\mathcal{G}$. Només cal veure que $I(\mathcal{G}) \subseteq I(P)$. Aquesta demostració, que usa eines fora de l'abast d'aquest treball, es pot consultar a [KF09, Sec. 4.5.1, Th. 4.9]. ■

Demostració de 1.45. Com que el graf endogen $\mathcal{G}_{\mathcal{M}}^{\rightarrow}$ és acíclic, admet un ordre topològic. Seguint aquest ordre, podem resoldre les equacions estructurals per substitució cap endavant: primer les variables sense pares endògens, després les variables que només depenen de les anteriors, i així successivament. Això defineix una aplicació mesurable global

$$\Psi_{\mathcal{M}} : \mathcal{U} \longrightarrow \mathcal{X}, \quad \mathbf{u} \longmapsto \mathbf{v}(\mathbf{u}),$$

on $\mathbf{v}(\mathbf{u})$ és l'única solució del sistema estructural per al valor exogen $\mathbf{u}$. Per tant, la distribució

observacional induïda queda determinada de manera única com la mesura imatge: $P(\mathcal{V}) = (\Psi_{\mathcal{M}})_{\#}P(\mathcal{U})$. O, formalment, $P(\mathcal{U})(\Psi_{\mathcal{M}}^{-1}(A))$, $A \subset \mathcal{X}$. Aquesta és exactament la idea de la distribució *entailed* d'un SCM de [PJS17, Prop. 6.3]: un cop fixades les funcions estructurals i la llei dels exògens, no hi ha cap altra distribució observacional compatible amb el model. ■

Demostració de 1.49. Seguim l'argument de [Bar+20, App. C, Th. 8]. Escriurem, com ells, la demostració en el cas discret. Cal tenir en compte que usarem conceptes que expliquem més endavant, com ara les C-components (cf. 2.20 i 1.46) i els pares estesos Pa_i^+ (cf. 2.23).

Fixem una assignació $\mathbf{v} = (v_1, \dots, v_n)$ de les variables endògenes i una assignació $\mathbf{u}$ dels factors exògens. Per la definició de distribució induïda,

$$p(\mathbf{v}) = \sum_{\mathbf{u}} \prod_{V_i \in \mathcal{V}} \mathbb{1}_{\{\psi_i(\mathbf{u})=v_i\}} p(\mathbf{u}).$$

Com que $\psi_i(\mathbf{u}) = f_i(\text{pa}_i, \mathbf{u}_i^c)$, cada indicador anterior es pot llegir com $\mathbb{1}_{\{\psi_i(\mathbf{u})=v_i\}} = p(v_i \mid \text{pa}_i^M, \mathbf{u}_i^c)$. Així,

$$p(\mathbf{v}) = \sum_{\mathbf{u}} \prod_{V_i \in \mathcal{V}} p(v_i \mid \text{pa}_i^M, \mathbf{u}_i^c) p(\mathbf{u}).$$

El pas clau de [Bar+20, App. C, Th. 8] consisteix a aplicar aquesta marginalització respectant l'ordre topològic que satisfà la definició de $\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}$. Dit d'una altra manera, si T_i és la C-component de V_i en el subgraf induït per $\mathcal{V}^{(i)}$, el pas inductiu de [TP02, Lem. 1] identifica el factor de la regla de la cadena amb el condicional donat per la component confosa ja revelada i els seus pares dirigits:

$$p(v_i \mid \mathbf{v}^{(i-1)}) = p(v_i \mid \mathbf{v}_{\text{Pa}(T_i) \setminus \{V_i\}}).$$

Amb la definició de pares estesos de 2.23, tenim $\text{Pa}(T_i) \setminus \{V_i\} = \text{Pa}_i^+$, i per tant

$$p(v_i \mid \mathbf{v}^{(i-1)}) = p(v_i \mid \text{pa}_i^+).$$

Així,

$$p(\mathbf{v}) = \prod_{V_i \in \mathcal{V}} p(v_i \mid \text{pa}_i^+).$$

Per la construcció de $\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}$, aquests pares estesos coincideixen amb els pares gràfics de V_i en $\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}$, és a dir, $\text{Pa}_i^{\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}} = \text{Pa}_i^+$. Per tant,

$$p(\mathbf{v}) = \prod_{V_i \in \mathcal{V}} p(v_i \mid \text{pa}_i^{\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}}),$$

i això és precisament la factorització d'una xarxa bayesiana sobre $\mathcal{G}_{\mathcal{M}}^{\mathcal{L}_1}$. ■

Demostració de 1.50. Ens basem en la demostració donada a [PJS17, Prop. 7.1]. En el cas real amb densitat respecte de Lebesgue podem prendre nuclis condicionals regulars de P , $P(V_i \in \cdot \mid V_1, \dots, V_{i-1})$. Prenem $\mathcal{U} = \{U_1, \dots, U_n\}$ amb $U_1, \dots, U_n$ independents i uniformes a $(0, 1)$. Com que P factoritza sobre $\mathcal{G}$, per a qualsevol ordre topològic $<$ sobre les variables compatible amb $\mathcal{G}$, i indexant les variables segons

aquest ordre, tenim

$$p(v_1, \dots, v_n) = \prod_{j=1}^n p(v_j \mid \text{pa}_j^{\mathcal{G}}).$$

Cada nucli condicional $P(V_j \in \cdot \mid \text{Pa}_j^{\mathcal{G}} = \text{pa}_j)$ es pot realitzar mitjançant una funció mesurable $f_j(\text{pa}_j, u_j)$ i una uniforme independent U_j . En el cas real, aquesta funció es pot escriure com la inversa generalitzada:

$$v_j = f_j(\text{pa}_j, u_j) := F_{V_j \mid \text{Pa}_j^{\mathcal{G}} = \text{pa}_j}^{-1}(u_j), \quad \forall j \in \{1, \dots, n\}$$

on $F_{V_j \mid \text{Pa}_j^{\mathcal{G}} = \text{pa}_j}^{-1}$ és la funció inversa generalitzada, definida per $F_Y^{-1}(a) := \inf\{y \in \mathbb{R} \mid F_Y(y) \geq a\}$. Per construcció, condicionat a $\text{Pa}_j^{\mathcal{G}} = \text{pa}_j$, la variable $f_j(\text{pa}_j, U_j)$ té distribució condicional $P(V_j \in \cdot \mid \text{Pa}_j^{\mathcal{G}} = \text{pa}_j)$. Com que els U_j són mútuament independents i cada equació només usa els pares de V_j en $\mathcal{G}$, el graf estructural endogen del model construït coincideix amb $\mathcal{G}$ i el model indueix una distribució que factoritza amb els mateixos nuclis condicionals que P . Per la factorització de Markov amb aquests mateixos nuclis, tal com en la prova de [PJS17, Ap. C.9], la distribució induïda coincideix amb P . ■

Demostració de 1.62.

1. Per la hipòtesi de factorització, en la configuració simple $X \rightarrow Y \leftarrow Z$ tenim $p(x, y, z) = p(x)p(z)p(y \mid x, z)$. Marginalitzant Y ,

$$\begin{aligned} p(x, z) &= \sum_y p(x, y, z) = \sum_y p(x) \cdot p(z) \cdot p(y \mid x, z) \\ &= p(x) \cdot p(z) \cdot \sum_y p(y \mid x, z) = p(x) \cdot p(z), \end{aligned}$$

on hem usat que $\sum_y p(y \mid x, z) = 1$. Per tant, $X \perp\!\!\!\perp_P Z$. El cas continu es prova de manera anàloga, substituint la suma per una integral.

2. Vegem ara quin és el cas excepcional en què condicionar pel col·lisionador no destrueix aquesta independència. Per Bayes,

$$p(x, z \mid y) = \frac{p(x, z, y)}{p(y)} = \frac{p(x) \cdot p(z) \cdot p(y \mid x, z)}{\sum_{x', z'} p(x') \cdot p(z') \cdot p(y \mid x', z')}. \quad (\text{A.2})$$

Per tant, perquè es compleixi $X \perp\!\!\!\perp_P Z \mid Y$, caldria que, per a cada valor fixat de y amb $p(y) > 0$, $p(x, z \mid y) = p(x \mid y) \cdot p(z \mid y)$.

Ara bé, per (A.2), això només passa en casos especials. En efecte, la independència condicional equival al fet que el terme que depèn conjuntament de x i z es pugui separar multiplicativament. És a dir, per a cada y , caldria que

$$p(y \mid x, z) = c(y) \cdot u(x, y) \cdot v(z, y), \quad (\text{A.3})$$

per a certes funcions u i v , on $u(x, y)$ depèn només de (x, y) , $v(z, y)$ depèn només de (z, y) , i $c(y)$ és un terme que només depèn de y .

En aquest cas excepcional, substituint (A.3) a (A.2), obtenim

$$\begin{aligned} p(x, z | y) &= \frac{c(y) \cdot p(x) \cdot p(z) \cdot u(x, y) \cdot v(z, y)}{c(y) \sum_{x', z'} p(x') \cdot p(z') \cdot u(x', y) \cdot v(z', y)} \\ &= \frac{p(x) \cdot u(x, y)}{\sum_{x'} p(x') \cdot u(x', y)} \cdot \frac{p(z) \cdot v(z, y)}{\sum_{z'} p(z') \cdot v(z', y)}. \end{aligned}$$

Per tant, en aquest cas $p(x, z | y)$ factoritza com a producte d'un terme que depèn només de x i y i d'un terme que depèn només de z i y . Recíprocament, si aquesta separació multiplicativa no és possible per a algun valor de y amb $p(y) > 0$, el quocient de Bayes en (A.2) no factoritza com $p(x | y)p(z | y)$. Aleshores $p(x | z, y) \neq p(x | y)$ per a algun valor de (z, y) , i per tant $X \not\perp_P Z | Y$. Aquest és el sentit precís de l'afirmació «fora dels casos especials» de 1.62. ■

Demostració de 1.53. Suposem, per reducció a l'absurd, que P factoritza sobre un subgraf propi $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ de $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, amb $\mathcal{E}' \subsetneq \mathcal{E}$. Aleshores hi ha dos nodes X, Y tals que $X \rightarrow Y \in \mathcal{E}$ i $X \rightarrow Y \notin \mathcal{E}'$ ²³.

Com que P factoritza sobre $\mathcal{G}'$, pel teorema d'ordered Markov tenim que Y és independent de qualsevol predecessor que no sigui pare seu en $\mathcal{G}'$, condicionadament als pares de Y en $\mathcal{G}'$: $Y \perp_P X | \text{Pa}_Y^{\mathcal{G}'}$.

Ara bé, $\text{Pa}_Y^{\mathcal{G}'}$ no conté ni X ni Y , i en $\mathcal{G}$ el camí directe $X \rightarrow Y$ no té nodes intermedis que puguin quedar bloquejats. Per tant, $X \not\perp_{\mathcal{G}} Y | \text{Pa}_Y^{\mathcal{G}'}$. Com que P és fidel a $\mathcal{G}$, això implica $Y \not\perp_P X | \text{Pa}_Y^{\mathcal{G}'}$, contradicció. Així, P no pot factoritzar sobre cap subgraf propi de $\mathcal{G}$. ■

Demostració de 1.54. La demostració es basa en provar el contrarrecíproc en ambdues direccions i l'hem detallat més que l'original. Recordem la hipòtesi local de l'enunciat: P admet una densitat respecte de la mesura dominant producte de 1.1. A més, com que P satisfà la propietat de Markov respecte de $\mathcal{G}$, podem utilitzar la factorització de P segons el graf $\mathcal{G}$ per l'equivalència entre propietat de Markov i factorització per a DAGs.

⇒ Vegem primer que si P satisfà la minimalitat causal respecte a $\mathcal{G}$, llavors es compleix la condició de dependència: $V_j \not\perp_P Y | \text{Pa}_j^{\mathcal{G}} \setminus \{Y\}$, per a tot V_j i per a tot $Y \in \text{Pa}_j^{\mathcal{G}}$. Suposem el contrari: existeix un node V_j i un pare $Y \in \text{Pa}_j^{\mathcal{G}}$ tals que $V_j \perp_P Y | \text{Pa}_j^{\mathcal{G}} \setminus \{Y\}$. Aquesta implica que la densitat condicional es pot simplificar: $p(v_j | \text{pa}_j^{\mathcal{G}}) = p(v_j | \text{pa}_j^{\mathcal{G}} \setminus \{y\})$ gairebé pertot. Si substituïm això en la factorització conjunta de P sobre $\mathcal{G}$, obtenim:

$$p(v_1, \dots, v_n) = p(v_j | \text{pa}_j^{\mathcal{G}} \setminus \{y\}) \prod_{k \neq j} p(v_k | \text{pa}_k^{\mathcal{G}}).$$

Aquesta nova factorització correspon a un graf $\mathcal{G}'$ on l'aresta $Y \rightarrow V_j$ ha estat eliminada ($\mathcal{G}' \subsetneq \mathcal{G}$). Com que P factoritza sobre aquest subgraf propi $\mathcal{G}'$, $\mathcal{G}$ no pot ser minimal.

⇐ Si P és compatible amb una densitat, la condició de Markov és equivalent a la factorització de Markov [Lau96, Th. 3.27]. Suposem ara que la condició de dependència es compleix per a tot pare

²³ Això equivaldria a dir que el graf $\mathcal{G}$ no és minimal, ja que podríem eliminar alguna aresta sense perdre la propietat de Markov.

del graf $\mathcal{G}$. Hem de veure que $\mathcal{G}$ és minimal. Raonem per reducció a l'absurd: suposem que $\mathcal{G}$ no fos minimal. Aleshores, P factoritzaria sobre algun subgraf propi $\mathcal{G}' \subsetneq \mathcal{G}$. Triem una aresta $Y \rightarrow V_i$ que pertany a $\mathcal{G}$ però no a $\mathcal{G}'$.

Sigui $\mathcal{G}''$ el graf obtingut de $\mathcal{G}$ eliminant només aquesta aresta $Y \rightarrow V_i$. Com que $\mathcal{G}' \subseteq \mathcal{G}''$, qualsevol distribució que factoritzi sobre $\mathcal{G}'$ també ho farà sobre $\mathcal{G}''$.

Atès que P factoritza sobre $\mathcal{G}''$, la regla de la cadena ens diu que $p(v_i \mid v_1, \dots, v_{i-1}) = p(v_i \mid \text{pa}_i^{\mathcal{G}''}) = p(v_i \mid \text{pa}_i^{\mathcal{G}} \setminus \{y\})$, pel que el factor de V_i no depèn de Y .

D'altra banda, per la hipòtesi inicial de l'enunciat, P també factoritza sobre el graf original $\mathcal{G}$, la qual cosa implica que $p(v_i \mid v_1, \dots, v_{i-1}) = p(v_i \mid \text{pa}_i^{\mathcal{G}})$. Igualant ambdues expressions obtenim:

$$p(v_i \mid \text{pa}_i^{\mathcal{G}}) = p(v_i \mid \text{pa}_i^{\mathcal{G}} \setminus \{y\}).$$

Això és exactament equivalent a afirmar que $V_i \perp\!\!\!\perp_P Y \mid \text{Pa}_i^{\mathcal{G}} \setminus \{Y\}$, fet que contradiu la hipòtesi de dependència per a tot pare. Així, $\mathcal{G}$ ha de ser minimal. ■

A.2

IDENTIFICABILITAT D'EFECTES CAUSALS

Demostració de 2.22. Sigui $\sim$ una relació sobre el conjunt de vèrtexs $\mathcal{V}$ definida per $X \sim Y \iff X = Y$ o bé existeix un camí format completament per arestes bidirigides entre X i Y . Com que els camins bidirigits es poden recórrer en ambdues direccions i concatenar-se, la relació $\sim$ és reflexiva, simètrica i transitiva. Per tant, és una relació d'equivalència i indueix una partició única de $\mathcal{V}$ en classes d'equivalència disjunts $S_1, \dots, S_k$ ²⁴.

Per definició, dos vèrtexs pertanyen a la mateixa classe si, i només si, estan connectats per un camí bidirigit. Així, els subgrafs induïts $\mathcal{G}[S_1], \dots, \mathcal{G}[S_k]$ són exactament les C-components maximals de $\mathcal{G}$. La unicitat de la partició dona la unicitat de la descomposició en C-components. ■

Demostració de 2.31. Primer, pel fet que $\mathbf{Y} \cap \mathbf{Z} = \emptyset$ i aplicant probabilitats totals sobre la distribució intervinguda $P(\mathcal{V} \mid \text{do}(\mathbf{x}))$, tenim que:

$$P(\mathbf{Y} \mid \text{do}(\mathbf{x})) = \mathbb{E}_{\mathbf{Z} \mid \text{do}(\mathbf{x})}[P(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{Z})].$$

El subíndex $\mathbb{E}_{\mathbf{Z} \mid \text{do}(\mathbf{x})}$ indica que l'esperança es calcula respecte a la distribució intervinguda. En el cas discret, aquesta esperança és la suma $\sum_{\mathbf{z}} p(\mathbf{y} \mid \text{do}(\mathbf{x}), \mathbf{z}) \cdot p(\mathbf{z} \mid \text{do}(\mathbf{x}))$. Ara:

$$\mathbb{E}_{\mathbf{Z} \mid \text{do}(\mathbf{x})}[P(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{Z})] = \mathbb{E}_{\mathbf{Z} \mid \text{do}(\mathbf{x})}[P(\mathbf{Y} \mid \mathbf{x}, \mathbf{Z})].$$

Si $\mathbf{Z}$ satisfà el criteri de *back-door*, tots els camins entre $\mathbf{X}$ i $\mathbf{Y}$ en $\mathcal{G}_{\mathbf{X}}$ han de començar entrant a $\mathbf{X}$ ²⁵. Per tant, qualsevol camí que connecti $\mathbf{X}$ i $\mathbf{Y}$ en $\mathcal{G}_{\mathbf{X}}$ és un *back-door path*. El criteri de *back-door* diu justament que aquests camins queden bloquejats per $\mathbf{Z}$, i per això $\mathbf{Y} \perp\!\!\!\perp_{\mathcal{G}_{\mathbf{X}}} \mathbf{X} \mid \mathbf{Z}$. La segona regla del do-calculus

²⁴ Aquest teorema diu que si R és una relació d'equivalència en un conjunt X , aleshores X es pot dividir de manera única en una partició de classes d'equivalència: per a cada $a \in X$ tenim que a pertany exactament a una classe d'equivalència A . Per a qualsevol parell de classes d'equivalència A, A' , tenim que $A = A'$ o bé $A \cap A' = \emptyset$.

²⁵ Com que $\mathcal{G}_{\mathbf{X}}$ elimina totes les arestes que surten d' $\mathbf{X}$, cap camí d'aquest graf pot sortir inicialment d' $\mathbf{X}$.

dona

$$P(Y | \text{do}(x), Z) = P(Y | x, Z).$$

Finalment, cal veure que:

$$\mathbb{E}_{Z|\text{do}(x)}[P(Y | x, Z)] = \mathbb{E}_Z[P(Y | x, Z)]. \quad (\text{A.4})$$

Per això hem de justificar que intervenir sobre X no altera la distribució de Z . Aplicarem la tercera regla del do-calculus; cal veure que $Z \perp_{\mathcal{G}_{\bar{X}}} X$, on $\mathcal{G}_{\bar{X}}$ és el graf amb les arestes entrants a X eliminades. En aquest graf, qualsevol camí entre X i Z ha de començar amb una aresta sortint de X . Aleshores només hi ha dues possibilitats:

1. Si el camí és dirigit de X cap a Z , algun node de Z és descendent de X , cosa que contradiu la hipòtesi del criteri *back-door*.
2. Si el camí no és dirigit de X cap a Z , en algun punt la direcció de les arestes ha de canviar i apareix un col·lisionador. Com que en aquest pas no condicionem per cap variable, aquests camins queden bloquejats (cf. 1.63).

Així doncs, $Z \perp_{\mathcal{G}_{\bar{X}}} X$ i la tercera regla del do-calculus implica que $P(Z | \text{do}(x)) = P(Z)$, per la qual cosa queda demostrada l'equació (A.4). ■

Demostració de 2.36. Idea: Primer identifiquem $p(z | \text{do}(x))$ com $p(z | x)$; després identifiquem $p(y | \text{do}(z))$ ajustant per X . Finalment, combinem les dues peces. Comencem escrivint:

$$P(Y | \text{do}(x)) = \mathbb{E}_{Z|\text{do}(x)}[P(Y | \text{do}(x), Z)].$$

Aquesta igualtat és la llei de la probabilitat total aplicada dins la distribució postintervenció $P(\mathcal{V} | \text{do}(x))$, com a la demostració de l'ajustament *back-door*. Volem aplicar la regla 2 del do-calculus per transformar la distribució de Z sota la intervenció $\text{do}(x)$: n'hi ha prou amb verificar que $Z \perp_{\mathcal{G}_X} X$. En efecte, en el graf $\mathcal{G}_X$, obtingut eliminant les arestes sortints de X , qualsevol camí entre X i Z és necessàriament un *back-door path*, i tots aquests camins estan bloquejats per $\emptyset$ per la segona condició del criteri *front-door*. Per tant, $P(Z | \text{do}(x)) = P(Z | x)$ i obtenim:

$$P(Y | \text{do}(x)) = \mathbb{E}_{Z|x}[P(Y | \text{do}(x), Z)].$$

Considerem ara un valor concret z de Z . Volem identificar el terme $P(Y | \text{do}(x), Z = z)$. Primer, tots els *back-door paths* de Z a Y queden bloquejats per X gràcies a la tercera condició de *front-door*. Aleshores, per la segona regla del do-calculus:

$$Y \perp_{\mathcal{G}_{\bar{X}, Z}} Z | X \implies P(Y | \text{do}(x), Z = z) = P(Y | \text{do}(x), \text{do}(z)).$$

Segon, apliquem la tercera regla del do-calculus per eliminar la intervenció sobre X . En efecte, en $\mathcal{G}_{\bar{Z}, \bar{X}}$, qualsevol camí de X a Y és o bé dirigit i queda tallat perquè la primera condició *front-door* obliga que passi per Z , o bé conté un col·lisionador; en tots dos casos queda bloquejat quan condicionem per Z .

$$Y \perp_{\mathcal{G}_{\bar{Z}, \bar{X}}} X | Z \implies P(Y | \text{do}(x), \text{do}(z)) = P(Y | \text{do}(z)).$$

Així, $P(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{Z} = \mathbf{z}) = P(\mathbf{Y} \mid \text{do}(\mathbf{z}))$. Ara marginalitzem sobre $\mathbf{X}$ dins la distribució $P(\mathcal{V} \mid \text{do}(\mathbf{z}))$:

$$P(\mathbf{Y} \mid \text{do}(\mathbf{z})) = \mathbb{E}_{\mathbf{X} \mid \text{do}(\mathbf{z})} [P(\mathbf{Y} \mid \text{do}(\mathbf{z}), \mathbf{X})].$$

Aplicarem de nou la tercera regla del do-calculus. En $\mathcal{G}_{\bar{\mathbf{Z}}}$ s'han eliminat les arestes entrants a $\mathbf{Z}$. Els camins de porta del darrere entre $\mathbf{X}$ i $\mathbf{Z}$ queden bloquejats per $\emptyset$ per la segona condició del criteri *front-door*; els camins dirigits de $\mathbf{X}$ cap a $\mathbf{Z}$ no poden romandre en $\mathcal{G}_{\bar{\mathbf{Z}}}$, perquè entrarien a $\mathbf{Z}$ per una aresta eliminada; i els camins restants queden bloquejats per col·lisionadors no condicionats. Així, $\mathbf{X} \perp_{\mathcal{G}_{\bar{\mathbf{Z}}}} \mathbf{Z}$ i la tercera regla dona $P(\mathbf{X} \mid \text{do}(\mathbf{z})) = P(\mathbf{X})$, com en [Par23, Sec. 1.1.4, Th. 2]. Finalment, per la regla 2, $P(\mathbf{Y} \mid \text{do}(\mathbf{z}), \mathbf{X}) = P(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z} = \mathbf{z})$ ja que $\mathbf{Y} \perp_{\mathcal{G}_{\bar{\mathbf{Z}}}} \mathbf{Z} \mid \mathbf{X}$, i això és perquè la tercera condició del criteri *front-door* garanteix que $\mathbf{X}$ bloqueja tots els *back-door paths* de $\mathbf{Z}$ a $\mathbf{Y}$. Combinant tots els passos anteriors, obtenim

$$P(\mathbf{Y} \mid \text{do}(\mathbf{x}), \mathbf{Z} = \mathbf{z}) = \mathbb{E}_{\mathbf{X}} [P(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z} = \mathbf{z})] \implies P(\mathbf{Y} \mid \text{do}(\mathbf{x})) = \mathbb{E}_{\mathbf{Z} \mid \mathbf{x}} [\mathbb{E}_{\mathbf{X}} [P(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z})]] .$$

Aquesta és exactament la fórmula d'ajustament *front-door*. ■

Observació. Una manera alternativa i més compacta de demostrar el teorema anterior és pensar-ho com una composició de dos ajustaments. En el cas discret, $p(\mathbf{y} \mid \text{do}(\mathbf{x}))$ és:

$$\sum_{\mathbf{z}} p(\mathbf{y} \mid \text{do}(\mathbf{x}), \mathbf{z}) p(\mathbf{z} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{z}} p(\mathbf{y} \mid \text{do}(\mathbf{x}), \text{do}(\mathbf{z})) p(\mathbf{z} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{z}} p(\mathbf{y} \mid \text{do}(\mathbf{z})) p(\mathbf{z} \mid \mathbf{x}).$$

El segon pas es justifica per la segona del do-calculus, usant la tercera condició del criteri *front-door*, i el tercer per la tercera regla del do-calculus, usant que $\mathbf{Z}$ intercepta tots els camins dirigits de $\mathbf{X}$ a $\mathbf{Y}$; a més, la identitat $p(\mathbf{z} \mid \text{do}(\mathbf{x})) = p(\mathbf{z} \mid \mathbf{x})$ també s'obté per la segona regla del do-calculus i gràcies a la segona condició. Un cop arribats aquí, l'efecte $p(\mathbf{y} \mid \text{do}(\mathbf{z}))$ s'identifica per *back-door* ajustant per $\mathbf{X}$, és a dir,

$$p(\mathbf{y} \mid \text{do}(\mathbf{z})) = \sum_{\mathbf{x}'} p(\mathbf{y} \mid \mathbf{z}, \mathbf{x}') p(\mathbf{x}').$$

Substituint aquesta expressió a l'equació anterior queda la fórmula d'ajustament *front-door*:

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \mathbf{x}) \sum_{\mathbf{x}'} p(\mathbf{y} \mid \mathbf{z}, \mathbf{x}') p(\mathbf{x}').$$

A.3

MODEL DE RESULTATS POTENCIALS

Demostració de 3.42. Per individualitat, existeix una funció $q : \mathcal{X}_{\text{cov}} \times \mathcal{Y}^2 \longrightarrow [0, 1]$ tal que:

$$\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q(\mathbf{X}_i, Y_i(0), Y_i(1)).$$

Per no-confusió, el mecanisme no depèn dels resultats potencials condicionadament a les covariants; per tant, existeix una funció $\tilde{q} : \mathcal{X}_{\text{cov}} \longrightarrow [0, 1]$ tal que aquesta probabilitat marginal es pot escriure com

$\tilde{q}(\mathbf{X}_i)$. Si $\mathbf{X}_i = \mathbf{x}$, aleshores el valor de propensió es pot expressar com:

$$e_{\text{fin}}(\mathbf{x}) = \frac{1}{N(\mathbf{x})} \sum_{i|\mathbf{X}_i=\mathbf{x}} \tilde{q}(\mathbf{X}_i) = \tilde{q}(\mathbf{x}).$$

Finalment, $\tilde{q}(\mathbf{x})$ és precisament la probabilitat individual d'assignació d'una unitat amb covariant $\mathbf{X}_i = \mathbf{x}$, és a dir, $\tilde{q}(\mathbf{x}) = \pi_i(\mathbf{X})$. ■

Demostració de 3.53. Per consistència, si $W_i = w$ aleshores $Y_i^{\text{obs}} = Y_i(w)$. Per ignorabilitat condicional, la distribució de $Y_i(w)$ condicionada a $\mathbf{X}_i = \mathbf{x}$ no depèn de si observem unitats sota tractament o de control; així,

$$\mathbb{E}[Y_i(w) | \mathbf{X}_i = \mathbf{x}] = \mathbb{E}[Y_i(w) | W_i = w, \mathbf{X}_i = \mathbf{x}] = \mathbb{E}[Y_i^{\text{obs}} | W_i = w, \mathbf{X}_i = \mathbf{x}] = m_w(\mathbf{x}),$$

on la positivitat garanteix que les dues esperances condicionades observacionals estan ben definides. En la primera igualtat hem usat la ignorabilitat i, en la segona, la consistència. Per tant,

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[\mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i]] = \mathbb{E}[m_1(\mathbf{X}_i) - m_0(\mathbf{X}_i)].$$

El cas del PATT s'obté restringint l'esperança a la subpoblació amb $W_i = 1$ i substituint el resultat potencial no observat $Y_i(0)$ per $m_0(\mathbf{X}_i)$:

$$\mathbb{E}[Y_i(1) - Y_i(0) | W_i = 1] = \mathbb{E}[Y_i^{\text{obs}} - m_0(\mathbf{X}_i) | W_i = 1].$$
 ■

Demostració de 3.61. Prenem $W = X$ i la intervenció nul·la a 3.59. Si $X(\omega) = x$, llavors $X_0(\omega) = x$ i, per composició, obtenim $Y_{0,x}(\omega) = Y_0(\omega)$. Com que $Y_0 = Y$, queda $Y_x(\omega) = Y(\omega)$. ■

Demostració de 3.64. Seguim el comentari de [GP98, Sec. 3]. En un sistema recursiu, una de les dues variables precedeix l'altra en l'ordenació causal. Si Y precedeix W , intervenir sobre W no pot afectar Y , i per tant $Y_{a,w}(\omega) = Y_a(\omega)$. En aquest cas la reversibilitat es redueix a

$$(Y_a(\omega) = y) \wedge (W_{a,y}(\omega) = w) \implies Y_a(\omega) = y,$$

que és trivial. Si, en canvi, W precedeix Y , intervenir sobre Y no pot afectar W , i per tant $W_{a,y}(\omega) = W_a(\omega)$. La hipòtesi de reversibilitat queda

$$(Y_{a,w}(\omega) = y) \wedge (W_a(\omega) = w).$$

Per composició, $W_a(\omega) = w$ implica $Y_{a,w}(\omega) = Y_a(\omega)$; combinant-ho amb $Y_{a,w}(\omega) = y$ obtenim $Y_a(\omega) = y$. Això és exactament la reducció recursiva de la reversibilitat descrita a l'original. ■

Demostració de 3.65. La demostració directa és per unicitat de solucions, que és la forma que cobreix tant el cas recursiu com el cas no recursiu amb solució única. Siguin $Y, W \in \mathcal{V}$ i $\mathbf{A} \subseteq \mathcal{V} \setminus \{W\}$. Fixem un estat $\omega \in \Omega$ i posem $\mathbf{u} := \mathbf{U}(\omega)$.

Sigui $\mathbf{v}$ l'única solució del submodel $\mathcal{M}_a$ en el context $\mathbf{u}$. Per hipòtesi, la seva component W és w , és a dir, $W_a(\omega) = w$. Considerem ara el submodel $\mathcal{M}_{a,w}$. Les equacions de totes les variables que no són intervingudes són les mateixes que a $\mathcal{M}_a$, i les variables de $\mathbf{A}$ continuen fixades a $\mathbf{a}$; l'única diferència és

que l'equació de W queda substituïda per la constant w . Com que la solució $\mathbf{v}$ de $\mathcal{M}_{\mathbf{a}}$ ja satisfà $W = w$, també satisfà totes les equacions de $\mathcal{M}_{\mathbf{a},w}$.

Per unicitat de la solució de $\mathcal{M}_{\mathbf{a},w}$, aquesta solució ha de ser precisament $\mathbf{v}$. En particular, les components corresponents a Y coincideixen:

$$Y_{\mathbf{a},w}(\omega) = Y_{\mathbf{a}}(\omega). \quad \blacksquare$$

Observació. Despleguem ara la reducció funcional de [GP98, Sec. 3.1] en el cas recursiu; és més detallada que l'original, però no introdueix una via diferent. Fixem $\mathbf{u}$ i sigui $\mathcal{V}$ el conjunt de totes les variables endògenes del model. Amb la notació habitual d'aquesta reducció, només necessitem resoldre les variables que poden aparèixer en l'expressió dels parets residuals $\mathbf{Z}$ de Y . Definim, doncs,

$$\mathbf{B}_{\text{int}} = \text{An}_{\mathbf{Z}}^{\mathcal{G}^M} \setminus (\mathbf{A} \cup \{W, Y\}),$$

on incloem $\mathbf{Z}$ dins dels seus avantpassats. Com que $\mathbf{Z} \subseteq \text{Pa}_Y^{\mathcal{G}^M}$, cap variable de $\mathbf{B}_{\text{int}}$ no pot dependre de Y en un DAG: si Y fos avantpassat d'algun avantpassat de $\mathbf{Z}$, tindríem un cicle dirigit que acabaria en Y passant per $\mathbf{Z} \rightarrow Y$.

Atès que el diagrama causal induït és acíclic, el conjunt $\mathbf{B}_{\text{int}}$ admet un ordre topològic $<$. Indexem aquestes variables com $V_1, V_2, \dots, V_k$, $k \leq n$, seguint aquest ordre: si V_i és pare d'un V_j , aleshores $V_i < V_j$ i, per tant, $i < j$. En el submodel $\mathcal{M}_{\mathbf{a}}$, l'equació estructural genèrica per a qualsevol $V_i \in \mathbf{B}_{\text{int}}$ ve donada per una funció f_{V_i} on els parets pertanyents a $\mathbf{A}$ ja han estat avaluats a la constant $\mathbf{a}$:

$$V_i = f_{V_i}((\text{Pa}_{V_i} \cap \mathbf{B}_{\text{int}}), \mathbf{a}, W, \mathbf{u}_i^c)$$

Resolem aquest subsistema de k equacions iterativament mitjançant substitució cap endavant, tractant W com w :

$$\begin{aligned} V_1 &= f_{V_1}(\mathbf{a}, w, \mathbf{u}_1^c) \equiv g_1(\mathbf{a}, w, \mathbf{u}) \\ V_2 &= f_{V_2}(\mathbf{a}, g_1(\mathbf{a}, w, \mathbf{u}), w, \mathbf{u}_2^c) \equiv g_2(\mathbf{a}, w, \mathbf{u}) \\ &\vdots \\ V_k &= f_{V_k}(\mathbf{a}, g_1(\cdot), \dots, g_{k-1}(\cdot), w, \mathbf{u}_k^c) \equiv g_k(\mathbf{a}, w, \mathbf{u}) \end{aligned}$$

Com a resultat d'aquest procés, eliminem les variables intermèdies com a arguments. Atès que $\mathbf{Z} \subseteq \mathbf{B}_{\text{int}}$, solament dependrà de les constants retintudes i de les exògenes: $\mathbf{Z} = \mathbf{g}_{\mathbf{Z}}(\mathbf{a}, w, \mathbf{u})$. Finalment, en $\mathcal{M}_{\mathbf{a}}$, l'equació estructural per a Y és $Y = f_Y(\mathbf{a}, \mathbf{Z}, W, \mathbf{u}_Y^c)$. Substituint-hi la funció vectorial resolta $\mathbf{g}_{\mathbf{Z}}$ i avaluant per a $W = w$, obtenim:

$$Y = f_Y(\mathbf{a}, \mathbf{g}_{\mathbf{Z}}(\mathbf{a}, w, \mathbf{u}), w, \mathbf{u}_Y^c)$$

Atès que l'interior de l'expressió ja no depèn de cap variable endògena intermèdia, podem col·lapsar la composició en una única funció reduïda f , obtenint l'expressió final $Y = f(\mathbf{a}, w, \mathbf{u})$.

Demostració de 3.66. Per definició de submodel, $\mathcal{M}_{x,w}$ substitueix l'equació estructural de X per l'equació

constant $X = x$. Per tant, qualsevol solució del submodel assigna a X el valor imposat per la intervenció, independentment de ω i de la resta d'equacions. Així, $X_{x,w}(\omega) = x$.

La versió en bloc s'obté pel mateix argument aplicat a cada variable intervinguda: en $\mathcal{M}_a$, l'equació de cada $A_j \in \mathbf{A}$ queda substituïda per $A_j = a_j$. ■

Demostració de 3.67. Siguin $Y, W \in \mathcal{V}$ dues variables diferents i sigui $\mathbf{A} \subseteq \mathcal{V} \setminus \{Y, W\}$. Fixem un estat $\omega \in \Omega$ i posem $\mathbf{u} := \mathbf{U}(\omega)$. Seguim la reducció funcional de [GP98, Sec. 3.1] fins al sistema reduït. El pas final d'unicitat està detallat més que a l'original, seguint l'argument de [Hal00, Ap. A, C5]; és una justificació explícita del mateix pas, no una prova per una via diferent. En el submodel $\mathcal{M}_a$, l'equació estructural per a Y pren la forma $Y = f_Y(\mathbf{a}, \mathbf{Z}, W, \mathbf{u}_Y^c)$, on $\mathbf{Z} = \text{Pa}_Y^{\mathcal{G}_M} \setminus (\mathbf{A} \cup \{W\})$.

Substituïm totes les variables excepte les de $\mathbf{A}$, Y i W , de manera que $\mathbf{Z}$ queda expressat com una funció de $\mathbf{a}$, w i $\mathbf{u}$. En compondre aquesta solució amb f_Y , obtenim una funció reduïda f :

$$f_Y(\mathbf{a}, \mathbf{Z}(\mathbf{a}, w, \mathbf{u}), w, \mathbf{u}_Y^c) \equiv f(\mathbf{a}, w, \mathbf{u}).$$

Com que hem mantingut W com a paràmetre, en el submodel $\mathcal{M}_{a,w}$ obtenim:

$$Y_{a,w}(\omega) = f(\mathbf{a}, w, \mathbf{u}).$$

Anàlogament, mantenint Y com a paràmetre en la reducció de l'equació de W , obtenim una funció g tal que:

$$W_{a,y}(\omega) = g(\mathbf{a}, y, \mathbf{u}).$$

Per tant, resoldre simultàniament $Y_a(\omega)$ i $W_a(\omega)$ en el submodel $\mathcal{M}_a$ equival a resoldre el sistema reduït:

$$y = f(\mathbf{a}, w, \mathbf{u}), \quad w = g(\mathbf{a}, y, \mathbf{u}).$$

Suposem ara que $Y_{a,w}(\omega) = y$ i $W_{a,y}(\omega) = w$. Per les dues igualtats anteriors, el parell (y, w) satisfà aquest sistema reduït. Per justificar que aquest parell coincideix amb la solució de $\mathcal{M}_a$, usem l'argument d'unicitat de [Hal00, Ap. A, C5]. Siguin $\mathbf{v}_1$ i $\mathbf{v}_2$ les solucions úniques de $\mathcal{M}_{a,w}$ i $\mathcal{M}_{a,y}$, respectivament. Per hipòtesi, totes dues assignacions coincideixen en les components indexades per $\mathbf{A} \cup \{Y, W\}$, amb valors $\mathbf{a}$, y i w . En el submodel $\mathcal{M}_{a,w,y}$, les equacions de qualsevol variable fora de $\mathbf{A} \cup \{Y, W\}$ són les mateixes que en els dos submodels anteriors, amb $Y = y$ i $W = w$ substituïts. Per tant, $\mathbf{v}_1$ i $\mathbf{v}_2$ satisfan també les equacions de $\mathcal{M}_{a,w,y}$; per unicitat, $\mathbf{v}_1 = \mathbf{v}_2$.

Aquesta solució comuna satisfà les equacions de $\mathcal{M}_a$: l'equació de Y es compleix perquè $\mathbf{v}_1$ satisfà $\mathcal{M}_{a,w}$, l'equació de W es compleix perquè $\mathbf{v}_2$ satisfà $\mathcal{M}_{a,y}$, i la resta d'equacions no canvien. Com que $\mathcal{M}_a$ té solució única, la seva component Y és y . Això prova la reversibilitat. ■

Demostració de 3.72. Volem demostrar que tot enunciat que és conseqüència semàntica d'un conjunt $\mathbf{S}$ també és derivable sintàcticament a partir de $\mathbf{S}$; és a dir, que $\mathbf{S} \models \sigma$ implica $\mathbf{S} \vdash_{AC} \sigma$.

Aquesta demostració segueix [GP98, Sec. 3.2], però explicita més passos lògics que l'original. Només en aquesta demostració recuperarem parcialment la sintaxi habitual de Galles–Pearl. Escriurem $\tilde{Y}_z(\mathbf{u})$ per

a la resposta estructural de Y en el submodel $\mathcal{M}_Z$ quan el context exogen és $\mathbf{u}$. Quan no hi hagi risc de confusió, escriurem simplement $Y_Z(\mathbf{u})$, com és habitual. La variable aleatòria emprada al text principal és $Y_Z = \tilde{Y}_Z \circ \mathbf{U}$, és a dir, $Y_Z(\omega) = \tilde{Y}_Z(\mathbf{U}(\omega))$.

N'hi ha prou amb demostrar que tot conjunt d'enunciats $\mathbf{S}$ que és consistent amb A_C té un model. En efecte, suposem que aquesta propietat es compleix i, tanmateix, A_C no és complet. Aleshores existeixen un conjunt $\mathbf{S}$ i un enunciat $p : X_Z(\mathbf{u}) = x$ tals que $\mathbf{S} \models p$, però $\mathbf{S} \not\models_{A_C} p$. Com que el domini de X és finit, la definitud diu que algun dels enunciats $X_Z(\mathbf{u}) = x''$ s'ha de satisfer. Si totes les alternatives $x'' \neq x$ fossin incompatibles amb $\mathbf{S}$, la definitud i la unicitat forçarien sintàcticament p , contradint $\mathbf{S} \not\models_{A_C} p$. Per tant, existeix algun valor $x' \neq x$ tal que $p' : X_Z(\mathbf{u}) = x'$ és compatible amb $\mathbf{S}$, és a dir, $\mathbf{S} \cup \{p'\}$ és consistent amb A_C . Per hipòtesi, aquest conjunt hauria de tenir un model. Però això és impossible, ja que $\mathbf{S} \models p$ implica que en qualsevol model de $\mathbf{S}$ s'ha de satisfer $X_Z(\mathbf{u}) = x$, i per tant cap model pot satisfer simultàniament $\mathbf{S}$ i $X_Z(\mathbf{u}) = x'$ amb $x' \neq x$. Aquesta contradicció prova que, si tot conjunt consistent té model, aleshores A_C és complet.

Fixem ara un conjunt $\mathbf{S}$ consistent amb A_C . Per un argument estàndard d'extensió de Lindenbaum²⁶, podem estendre $\mathbf{S}$ a un conjunt màximament consistent $\mathbf{S}^*$. Demostrarem que $\mathbf{S}^*$ té un model causal $\mathcal{M}$; en particular, aquest model satisfarà també tots els enunciats de $\mathbf{S}$.

Procedim per inducció sobre el nombre de variables endògenes $|\mathcal{V}|$ que apareixen a $\mathbf{S}^*$.

1. **Cas base:** Si $|\mathcal{V}| = 1$, diguem $\mathcal{V} = \{X\}$, aleshores els enunciats de la forma $X(\mathbf{u}) = x$ que pertanyen a $\mathbf{S}^*$ determinen completament la funció estructural de X . En efecte, per màxima consistència i definitud, per a cada $\mathbf{u}$ hi ha un únic valor x tal que $X(\mathbf{u}) = x \in \mathbf{S}^*$. Definim $f_X(\mathbf{u}) = x$ si i només si $X(\mathbf{u}) = x \in \mathbf{S}^*$. Per efectivitat, en el submodel $\mathcal{M}_X$ es compleix $X_X(\mathbf{u}) = x$ per a tot $x \in \mathcal{X}_X$. Per tant, el model així construït satisfà tots els enunciats de $\mathbf{S}^*$.
2. **Pas inductiu:** Suposem el resultat cert per a qualsevol conjunt màximament consistent que involucri estrictament menys de $|\mathcal{V}|$ variables, i sigui ara $\mathbf{S}^*$ un conjunt màximament consistent sobre $\mathcal{V}$.

Sigui $Y \in \mathcal{V}$ el darrer element d'un ordenament causal. Considerem el conjunt $\mathbf{S}'$ obtingut a partir de $\mathbf{S}^*$ eliminant tots els enunciats de la forma $Y_Z(\mathbf{u}) = y$ i $X_{y,z}(\mathbf{u}) = x$, amb $X \neq Y$. El conjunt $\mathbf{S}'$ continua essent consistent amb A_C , perquè qualsevol contradicció derivable de $\mathbf{S}'$ també ho seria de $\mathbf{S}^*$. Per tant, es pot estendre a un conjunt màximament consistent $\mathbf{S}'^*$ sobre el llenguatge reduït. Per la hipòtesi inductiva, existeix un model causal $\mathcal{M}'$ que satisfà tots els enunciats de $\mathbf{S}'^*$. Construïm ara un model causal $\mathcal{M}$ que satisfà $\mathbf{S}^*$. Com que Y és el darrer element de l'ordenació causal, no és pare de cap altra variable. Per tant, per a tota variable $X \neq Y$, podem mantenir la mateixa funció estructural que a $\mathcal{M}'$, és a dir, definim $f_X^{\mathcal{M}} := f_X^{\mathcal{M}'}$.

Només cal definir f_Y . Per a cada assignació $\mathbf{z}$ de les variables de $\mathcal{V} \setminus \{Y\}$ i cada $\mathbf{u}$, per màxima consistència i definitud hi ha un únic valor y tal que $Y_Z(\mathbf{u}) = y \in \mathbf{S}^*$. Definim, doncs, $f_Y(\mathbf{z}, \mathbf{u}) = y$ si i només si $Y_Z(\mathbf{u}) = y \in \mathbf{S}^*$. Aquesta definició és correcta perquè la definitud garanteix la unicitat del valor y per a cada parell $(\mathbf{z}, \mathbf{u})$.

²⁶ El lema de Lindenbaum garanteix que, donat un llenguatge formal, qualsevol conjunt consistent de fórmules es pot estendre a un conjunt màximament consistent; cf. [Men15, Prop. 2.14]. Per a l'ús específic d'aquest argument en la demostració de completesa de la lògica contrafactual sobre models causals, cf. [Hal00, Th. 3.2 i Ap. A].

Veiem ara que $\mathcal{M}$ satisfà tots els enunciats de $\mathbf{S}^*$. En primer lloc, satisfà tots els enunciats que no involucren Y , ja que sobre aquestes variables el model $\mathcal{M}$ coincideix amb $\mathcal{M}'$. A més, com que Y és terminal en l'ordre causal, intervenir sobre Y no altera el valor de cap altra variable. Per tant, si $X \neq Y$, en $\mathcal{M}$ es compleix $X_{y,z}(\mathbf{u}) = X_z(\mathbf{u})$. Si $X_{y,z}(\mathbf{u}) = x$ pertanyia a $\mathbf{S}^*$, l'axioma d'ordenació dona $X_{y,z}(\mathbf{u}) = X_z(\mathbf{u})$ i, pel tancament deductiu de $\mathbf{S}^*$ i la unicitat, també tenim $X_z(\mathbf{u}) = x$ en $\mathbf{S}^*$. Aquest darrer enunciat no ha estat eliminat, de manera que el satisfà $\mathcal{M}'$, i per tant també $\mathcal{M}$. Només resta verificar que $\mathcal{M}$ satisfà tots els enunciats de la forma $Y_z(\mathbf{u}) = y$ que apareixen a $\mathbf{S}^*$. Ho demostrarem per inducció sobre la quantitat $|\mathcal{V}| - |\mathcal{Z}|$.

Casos base:

1. Si $Y \in \mathcal{Z}$, aleshores z imposa algun valor y_0 a Y i, per efectivitat, $Y_z(\mathbf{u}) = y_0$. Com que $Y_z(\mathbf{u}) = y$ pertany a $\mathbf{S}^*$ i $\mathbf{S}^*$ és consistent amb unicitat, necessàriament $y = y_0$.
2. Si $|\mathcal{V}| - |\mathcal{Z}| = 1$, aleshores z imposa valors a totes les variables excepte Y . En aquest cas, la satisfacció de $Y_z(\mathbf{u}) = y$ ve directament de la definició de f_Y .

Pas inductiu: Suposem ara que $|\mathcal{V}| - |\mathcal{Z}| = k > 1$ i que el resultat és cert per a tots els conjunts d'intervenció de mida estrictament més gran. Sigui $Y_z(\mathbf{u}) = y \in \mathbf{S}^*$. Triem una variable $X \notin \mathcal{Z} \cup \{Y\}$. Per màxima consistència i definitud, existeix un únic valor x tal que $X_z(\mathbf{u}) = x \in \mathbf{S}^*$. De la mateixa manera, existeix un únic valor y' tal que $Y_{z,x}(\mathbf{u}) = y' \in \mathbf{S}^*$. Com que $|\mathcal{V}| - |\mathcal{Z} \cup \{X\}| = k - 1$, per la hipòtesi inductiva tenim que $\mathcal{M}$ satisfà $Y_{z,x}(\mathbf{u}) = y'$. A més, ja hem vist que $\mathcal{M}$ satisfà $X_z(\mathbf{u}) = x$. Aplicant composició dins de $\mathcal{M}$, obtenim que $\mathcal{M}$ satisfà $Y_z(\mathbf{u}) = y'$. El mateix pas també és una derivació sintàctica dins de $\mathbf{S}^*$: de $X_z(\mathbf{u}) = x$ i $Y_{z,x}(\mathbf{u}) = y'$, per composició es deriva $Y_z(\mathbf{u}) = y'$. Com que $Y_z(\mathbf{u}) = y \in \mathbf{S}^*$ i $\mathbf{S}^*$ és consistent amb unicitat, tenim $y = y'$. Per tant, $\mathcal{M}$ satisfà $Y_z(\mathbf{u}) = y$.

Això conclou el pas inductiu i, per tant, la construcció d'un model causal que satisfà qualsevol conjunt màximament consistent $\mathbf{S}^*$. En conseqüència, tot conjunt consistent $\mathbf{S}$ té model, i d'aquí se segueix la completesa d' A_C . ■

A.4

ESTUDIS EXPERIMENTALS I OBSERVACIONALS

Demostració de 4.3. La verificació desenvolupa directament el mecanisme d'assignació d'un experiment completament aleatoritzat de [IR15, Def. 4.2 i Sec. 4.4]. Per definició d'experiment completament aleatoritzat, la probabilitat d'un vector d'assignació $\mathbf{w}$ només depèn de si $\sum_i w_i = n_t$ o no. Per tant, no depèn ni de $\mathbf{X}$ ni de $\mathbf{Y}(0), \mathbf{Y}(1)$, i obtenim directament:

$$\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \pi(\mathbf{w}).$$

Per calcular la probabilitat marginal que una unitat concreta i sigui tractada, comptem els vectors admissibles amb $w_i = 1$. Cal triar les $n_t - 1$ unitats tractades restants entre les altres $n - 1$ unitats, de manera que:

$$\pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \frac{\binom{n-1}{n_t-1}}{\binom{n}{n_t}} = \frac{n_t}{n} \text{ i } 1 - \pi_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \frac{\binom{n-1}{n_t}}{\binom{n}{n_t}} = \frac{n_c}{n}. \quad (\text{A.5})$$

Com que $0 < n_t < n$, totes dues probabilitats marginals són estrictament positives i, per tant, el mecanisme és probabilístic. D'altra banda, el càlcul marginal de (A.5) no intervé en la no-confusió: aquesta se segueix de la igualtat anterior $\pi(\mathbf{w} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \pi(\mathbf{w})$, ja que el mecanisme conjunt no depèn dels resultats potencials. Així, el mecanisme és no-confós en el sentit de 3.39. ■

Demostració de 4.7. La demostració segueix el càlcul de [IR15, Sec. 6.3, Th. 6.1]. Notem que $Y_i^{\text{obs}} = Y_i(1)$ si $W_i = 1$ i $Y_i^{\text{obs}} = Y_i(0)$ si $W_i = 0$. Així doncs, podem reescriure l'estimador com:

$$\hat{\tau}_{\text{dif}} = \frac{1}{n_t} \sum_{i=1}^n W_i Y_i(1) - \frac{1}{n_c} \sum_{i=1}^n (1 - W_i) Y_i(0).$$

Com que condicionem a $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$, els resultats potencials són fixos i l'única aleatorietat prové de $\mathbf{W}$. En un experiment completament aleatoritzat amb n_t unitats tractades, tenim per (A.5):

$$\mathbb{E}_{\mathbf{W}}[W_i \mid \mathbf{Y}(0), \mathbf{Y}(1)] = \mathbb{E}_{\mathbf{W}}[W_i] = \frac{n_t}{n}, \quad \mathbb{E}_{\mathbf{W}}[1 - W_i \mid \mathbf{Y}(0), \mathbf{Y}(1)] = \mathbb{E}_{\mathbf{W}}[1 - W_i] = \frac{n_c}{n}.$$

Per la linealitat de l'esperança (cf. B.63),

$$\begin{aligned} \mathbb{E}_{\mathbf{W}}[\hat{\tau}_{\text{dif}} \mid \mathbf{Y}(0), \mathbf{Y}(1)] &= \frac{1}{n_t} \sum_{i=1}^n \mathbb{E}_{\mathbf{W}}[W_i] \cdot Y_i(1) - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}_{\mathbf{W}}[1 - W_i] \cdot Y_i(0) \\ &= \frac{1}{n} \sum_{i=1}^n Y_i(1) - \frac{1}{n} \sum_{i=1}^n Y_i(0) = \frac{1}{n} \sum_{i=1}^n (Y_i(1) - Y_i(0)), \end{aligned}$$

i això és justament el SATE. Per tant, el biaix és zero respecte de la distribució d'aleatorització. ■

Demostració de 4.9. Donem una derivació equivalent del resultat de [IR15, Sec. 6.4 i Apèndix A del cap. 6], basada en els mateixos moments de l'assignació sota aleatorització completa. Condicionem a $\mathbf{Y}(0)$ i $\mathbf{Y}(1)$, de manera que aquests vectors són fixos i l'única aleatorietat és la de $\mathbf{W}$. Posem $A = \frac{1}{n_t} \sum_{i=1}^n W_i Y_i(1)$ i $B = \frac{1}{n_c} \sum_{i=1}^n (1 - W_i) Y_i(0)$, de manera que $\hat{\tau}_{\text{dif}} = A - B$. Escrivim $a_i = Y_i(1) - \bar{Y}(1)$ i $b_i = Y_i(0) - \bar{Y}(0)$ i definim la covariància de mostra finita entre els dos vectors de resultats potencials:

$$S_{10} = \frac{1}{n-1} \sum_{i=1}^n a_i b_i.$$

En un experiment completament aleatoritzat, els moments dels indicadors d'assignació són

$$\mathbb{E}_{\mathbf{W}}[W_i] = \frac{n_t}{n}, \quad \mathbb{E}_{\mathbf{W}}[W_i W_j] = \frac{n_t(n_t - 1)}{n(n - 1)}, \quad \mathbb{E}_{\mathbf{W}}[(1 - W_i)(1 - W_j)] = \frac{n_c(n_c - 1)}{n(n - 1)} \quad (\text{A.6})$$

per a $i \neq j$. També, per a $i = j$,

$$\mathbb{E}_{\mathbf{W}}[W_i(1 - W_j)] = \frac{\binom{n-2}{n_t-1}}{\binom{n}{n_t}} = \frac{n_t}{n} \frac{n_c}{n-1},$$

i, en canvi, $W_i(1 - W_i) = 0$. Com que $\sum_i W_i = n_t$ sempre, tenim $A - \bar{Y}(1) = n_t^{-1} \sum_i W_i a_i$. Usant (A.6) i

$$\sum_i a_i = 0,$$

$$\begin{aligned} \text{Var}_{\mathbf{W}}(A) &= \mathbb{E}_{\mathbf{W}} \left[\left(\frac{1}{n_t} \sum_{i=1}^n W_i a_i \right)^2 \right] = \frac{1}{n_t^2} \left[\frac{n_t}{n} \sum_{i=1}^n a_i^2 + \frac{n_t(n_t-1)}{n(n-1)} \sum_{i \neq j} a_i a_j \right] \\ &= \frac{1}{n_t^2} \left[\frac{n_t}{n} (n-1) S_1^2 - \frac{n_t(n_t-1)}{n} S_1^2 \right] = \left(\frac{1}{n_t} - \frac{1}{n} \right) S_1^2. \end{aligned}$$

El mateix càlcul amb els indicadors $1 - W_i$, ara usant $\sum_i (1 - W_i) = n_c$ i el tercer moment de (A.6), dona

$$\text{Var}_{\mathbf{W}}(B) = \left(\frac{1}{n_c} - \frac{1}{n} \right) S_0^2.$$

Per a la covariància, com que $\sum_i a_i = \sum_i b_i = 0$,

$$\begin{aligned} \text{Cov}_{\mathbf{W}}(A, B) &= \mathbb{E}_{\mathbf{W}} \left[\left(\frac{1}{n_t} \sum_{i=1}^n W_i a_i \right) \left(\frac{1}{n_c} \sum_{j=1}^n (1 - W_j) b_j \right) \right] = \frac{1}{n_t n_c} \sum_{i \neq j} \frac{n_t n_c}{n(n-1)} a_i b_j \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n a_i \sum_{j \neq i} b_j = -\frac{1}{n(n-1)} \sum_{i=1}^n a_i b_i = -\frac{S_{10}}{n}. \end{aligned}$$

Així,

$$\text{Var}_{\mathbf{W}}(\hat{\tau}_{\text{dif}}) = \text{Var}_{\mathbf{W}}(A - B) = \left(\frac{1}{n_t} - \frac{1}{n} \right) S_1^2 + \left(\frac{1}{n_c} - \frac{1}{n} \right) S_0^2 + \frac{2S_{10}}{n} = \frac{S_1^2}{n_t} + \frac{S_0^2}{n_c} - \frac{S_1^2 + S_0^2 - 2S_{10}}{n}.$$

Finalment,

$$S_{\tau}^2 = \frac{1}{n-1} \sum_{i=1}^n ((Y_i(1) - Y_i(0)) - (\bar{Y}(1) - \bar{Y}(0)))^2 = S_1^2 + S_0^2 - 2S_{10}. \quad \blacksquare$$

Demostració de 4.11. A [IR15, Sec. 6.5 i Apèndix A del cap. 6] es mostra que les variàncies mostrals observades són estimadors no esbiaixats de S_1^2 i S_0^2 sota aleatorització completa. En donem el càlcul per a una mostra sense reemplaçament, considerablement més simple. Sigui $z_1, \dots, z_n$ valors fixos, amb mitjana $\bar{z}$ i variància finita $S_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2$. Sigui $R_i \in \{0, 1\}$ els indicadors d'una mostra aleatòria uniforme de mida fixa m , i posem $\bar{z}_R = m^{-1} \sum_i R_i z_i$ i:

$$s_R^2 = \frac{1}{m-1} \sum_{i|R_i=1} (z_i - \bar{z}_R)^2 \implies s_R^2 = \frac{1}{m-1} \left[\sum_{i=1}^n R_i (z_i - \bar{z})^2 - m(\bar{z}_R - \bar{z})^2 \right].$$

Prenent esperances, usant $\mathbb{E}[R_i] = \frac{m}{n}$ i el càlcul de variància de la demostració anterior aplicat a la mitjana mostral $\bar{z}_R$,

$$\mathbb{E}[s_R^2] = \frac{1}{m-1} \left[\frac{m}{n} (n-1) S_z^2 - m \left(\frac{1}{m} - \frac{1}{n} \right) S_z^2 \right] = S_z^2.$$

Aplicant aquesta identitat amb $z_i = Y_i(1)$ i $R_i = W_i$, i després amb $z_i = Y_i(0)$ i $R_i = 1 - W_i$, obtenim

$$\mathbb{E}_{\mathbf{W}}[s_1^2 \mid \mathbf{Y}(0), \mathbf{Y}(1)] = S_1^2, \quad \mathbb{E}_{\mathbf{W}}[s_0^2 \mid \mathbf{Y}(0), \mathbf{Y}(1)] = S_0^2.$$

Així,

$$\mathbb{E}_{\mathbf{W}} \left[\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}_{\text{dif}} \mid \mathbf{Y}(0), \mathbf{Y}(1)) \right] = \frac{S_1^2}{n_t} + \frac{S_0^2}{n_c}.$$

Si l'efecte causal és constant, $S_\tau^2 = 0$ i aquesta esperança coincideix amb la variància de 4.9. En general, combinant l'expressió anterior amb 4.9,

$$\mathbb{E}_{\mathbf{W}} \left[\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}_{\text{dif}} \mid \mathbf{Y}(0), \mathbf{Y}(1)) \right] - \text{Var}_{\mathbf{W}}(\hat{\tau}_{\text{dif}}) = \frac{S_\tau^2}{n} \geq 0. \quad \blacksquare$$

Demostració de 4.14. Aquest és el resultat de conservació de la ignorabilitat i la positivitat en condicionar per un valor d'equilibratge de [RR83, Th. 3] i [IR15, Sec. 12.3, Lema 12.2]. Fixem versions regulars de les probabilitats condicionals; totes les igualtats següents s'entenen quasi segurament. Escrivim $\mathbf{Y}_i^* = (Y_i(0), Y_i(1))$ i $I_w = \mathbb{1}_{\{W_i=w\}}$, per a $w \in \{0, 1\}$. Com que I_w és la indicadora de l'esdeveniment $\{W_i = w\}$, les probabilitats condicionals que segueixen són esperances condicionades d' I_w . Per B.62, condicionant primer per $\mathbf{Y}_i^*, \mathbf{X}_i$ i $b(\mathbf{X}_i)$ i després per $\mathbf{Y}_i^*$ i $b(\mathbf{X}_i)$, obtenim:

$$P(W_i = w \mid \mathbf{Y}_i^*, b(\mathbf{X}_i)) = \mathbb{E}[\mathbb{E}[I_w \mid \mathbf{Y}_i^*, \mathbf{X}_i, b(\mathbf{X}_i)] \mid \mathbf{Y}_i^*, b(\mathbf{X}_i)] = \mathbb{E}[P(W_i = w \mid \mathbf{X}_i, b(\mathbf{X}_i)) \mid \mathbf{Y}_i^*, b(\mathbf{X}_i)],$$

on hem usat la ignorabilitat condicional donada $\mathbf{X}_i$ i el fet que $b(\mathbf{X}_i)$ és funció de $\mathbf{X}_i$. Com que $b(\mathbf{X}_i)$ és un valor d'equilibratge,

$$P(W_i = w \mid \mathbf{X}_i, b(\mathbf{X}_i)) = P(W_i = w \mid b(\mathbf{X}_i)).$$

Substituint dins de l'esperança anterior i usant que $P(W_i = w \mid b(\mathbf{X}_i))$ és mesurable respecte de $b(\mathbf{X}_i)$, queda

$$P(W_i = w \mid \mathbf{Y}_i^*, b(\mathbf{X}_i)) = P(W_i = w \mid b(\mathbf{X}_i)),$$

i això prova $\mathbf{Y}_i^* \perp\!\!\!\perp W_i \mid b(\mathbf{X}_i)$. Finalment,

$$P(W_i = 1 \mid b(\mathbf{X}_i)) = \mathbb{E}[\mathbb{E}[W_i \mid \mathbf{X}_i] \mid b(\mathbf{X}_i)] = \mathbb{E}[e(\mathbf{X}_i) \mid b(\mathbf{X}_i)].$$

A la primera igualtat hem tornat a aplicar B.62. Per tant, si $0 < e(\mathbf{X}_i) < 1$ quasi segurament, la darrera esperança condicionada també és estrictament entre 0 i 1 quasi segurament: si fos 0 o 1 en un esdeveniment de probabilitat positiva, la integral condicionada contradiria respectivament $e(\mathbf{X}_i) > 0$ o $1 - e(\mathbf{X}_i) > 0$ quasi segurament. $\blacksquare$

Demostració de 4.15. La primera part és [RR83, Th. 1] i [IR15, Sec. 12.3, Lema 12.1]: el valor de propensió és un valor d'equilibratge. Fixem versions regulars de les probabilitats condicionals; les igualtats següents valen quasi segurament. Com que W_i és una variable indicadora, segueix una Bernoulli i, per tant, $P(W_i = 1 \mid \cdot) = \mathbb{E}[W_i \mid \cdot]$. A més, $e(\mathbf{X}_i)$ és funció de $\mathbf{X}_i$, de manera que condicionar per $\mathbf{X}_i$ ja determina també $e(\mathbf{X}_i)$. Per tant,

$$P(W_i = 1 \mid \mathbf{X}_i, e(\mathbf{X}_i)) = \mathbb{E}[W_i \mid \mathbf{X}_i, e(\mathbf{X}_i)] = \mathbb{E}[W_i \mid \mathbf{X}_i] = e(\mathbf{X}_i).$$

D'altra banda, per esperances iterades (cf. B.62),

$$P(W_i = 1 \mid e(\mathbf{X}_i)) = \mathbb{E}[W_i \mid e(\mathbf{X}_i)] = \mathbb{E}[\mathbb{E}[W_i \mid \mathbf{X}_i, e(\mathbf{X}_i)] \mid e(\mathbf{X}_i)] = \mathbb{E}[e(\mathbf{X}_i) \mid e(\mathbf{X}_i)] = e(\mathbf{X}_i).$$

La mateixa igualtat per a $W_i = 0$ s'obté substituint $e(\mathbf{X}_i)$ per $1 - e(\mathbf{X}_i)$. Per tant, la distribució condicional de W_i donada $\mathbf{X}_i$ i $e(\mathbf{X}_i)$ coincideix amb la distribució condicional de W_i donada només $e(\mathbf{X}_i)$, i això prova $W_i \perp\!\!\!\perp \mathbf{X}_i \mid e(\mathbf{X}_i)$.

La darrera afirmació s'obté aplicant 4.14 amb $b(\mathbf{X}_i) = e(\mathbf{X}_i)$. ■

Demostració de 4.16. Aquest és el resultat de *coarseness* de [RR83, Th. 2], recollit també a [IR15, Sec. 12.3, Lema 12.3], però amb una demostració lleugerament diferent a les originals. Sigui $b(\mathbf{X}_i)$ un valor d'equilibratge. Com que $b(\mathbf{X}_i)$ és funció de $\mathbf{X}_i$, condicionar per $\mathbf{X}_i$ ja determina també $b(\mathbf{X}_i)$, i per tant

$$e(\mathbf{X}_i) = P(W_i = 1 \mid \mathbf{X}_i) = P(W_i = 1 \mid \mathbf{X}_i, b(\mathbf{X}_i)).$$

Això és cert perquè la σ -àlgebra generada per $\mathbf{X}_i$ conté la σ -àlgebra generada per $b(\mathbf{X}_i)$ (ja que b és una funció mesurable de $\mathbf{X}_i$). Com que $b(\mathbf{X}_i)$ és un valor d'equilibratge, $W_i \perp\!\!\!\perp \mathbf{X}_i \mid b(\mathbf{X}_i)$, de manera que

$$P(W_i = 1 \mid \mathbf{X}_i, b(\mathbf{X}_i)) = P(W_i = 1 \mid b(\mathbf{X}_i)).$$

Triem una versió regular de la probabilitat condicionada $P(W_i = 1 \mid b(\mathbf{X}_i))$ i definim $f(t)$ com el valor d'aquesta versió en t sobre la imatge de b ; fora d'aquesta imatge podem definir f arbitràriament. Com que les probabilitats condicionals només són úniques llevat de conjunts nuls (cf. B.61), obtenim $e(\mathbf{X}_i) = f(b(\mathbf{X}_i))$ quasi segurament. ■

Demostració de 4.18. La prova segueix les dues igualtats de [IR15, Sec. 17.8], que són una aplicació de l'estimador de Horvitz–Thompson [HT52] al cas del valor de propensió. Una formulació equivalent en termes d'IP weighting i estandardització es troba a [HR20, Technical Point 2.3]. Escrivim $\mathbb{E}$ per a l'esperança conjunta, respecte del mostreig de superpoblació i del mecanisme d'assignació. Seguint la notació del treball, reservem $\mathbb{E}_{\text{sp}}$ per al mostreig de superpoblació i $\mathbb{E}_{\mathbf{W}}$ per a l'assignació, condicionada a la mostra finita; en aquesta prova només en fem servir la marginal de W_i . Com que les unitats són les còpies i.i.d. introduïdes a 3.24, tenen la mateixa distribució. Per linealitat, n'hi ha prou de provar les dues igualtats:

$$\mathbb{E} \left[\frac{W_i Y_i^{\text{obs}}}{e(\mathbf{X}_i)} \right] = \mathbb{E}_{\text{sp}}[Y_i(1)], \quad \mathbb{E} \left[\frac{(1 - W_i) Y_i^{\text{obs}}}{1 - e(\mathbf{X}_i)} \right] = \mathbb{E}_{\text{sp}}[Y_i(0)].$$

Per positivitat, els denominadors són no nuls quasi segurament. A més, la integrabilitat de $Y_i(0)$ i $Y_i(1)$ garanteix que els termes ponderats tenen esperança finita; per exemple, usant ignorabilitat condicional,

$$\mathbb{E} \left[\frac{W_i |Y_i^{\text{obs}}|}{e(\mathbf{X}_i)} \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{W_i |Y_i(1)|}{e(\mathbf{X}_i)} \mid \mathbf{X}_i, Y_i(1) \right] \right] = \mathbb{E}_{\text{sp}} \left[\frac{|Y_i(1)| \mathbb{E}_{\mathbf{W}}[W_i \mid \mathbf{X}_i, Y_i(1)]}{e(\mathbf{X}_i)} \right] = \mathbb{E}_{\text{sp}}[|Y_i(1)|].$$

En l'últim pas hem usat que $\mathbb{E}_{\mathbf{W}}[W_i \mid \mathbf{X}_i, Y_i(1)] = e(\mathbf{X}_i)$ quasi segurament. Provem la primera igualtat; la segona és anàloga. Com hem dit a la demostració de 4.7, $Y_i^{\text{obs}} = Y_i(1)$ si $W_i = 1$ i $Y_i^{\text{obs}} = Y_i(0)$ si $W_i = 0$. Per tant, $W_i Y_i^{\text{obs}} = W_i Y_i(1)$. Aplicant B.62 per condicionar respecte de $\mathbf{X}_i$,

$$\mathbb{E} \left[\frac{W_i Y_i^{\text{obs}}}{e(\mathbf{X}_i)} \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{W_i Y_i(1)}{e(\mathbf{X}_i)} \mid \mathbf{X}_i \right] \right].$$

Per ignorabilitat condicional, $W_i \perp\!\!\!\perp Y_i(1) \mid \mathbf{X}_i$. Per tant, dins l'esperança condicionada podem separar la part d'assignació de la part de superpoblació:

$$\mathbb{E} \left[\frac{W_i Y_i(1)}{e(\mathbf{X}_i)} \mid \mathbf{X}_i \right] = \frac{\mathbb{E}_W[W_i \mid \mathbf{X}_i] \mathbb{E}_{\text{sp}}[Y_i(1) \mid \mathbf{X}_i]}{e(\mathbf{X}_i)} = \mathbb{E}_{\text{sp}}[Y_i(1) \mid \mathbf{X}_i],$$

on hem usat que $\mathbb{E}_W[W_i \mid \mathbf{X}_i] = e(\mathbf{X}_i)$ i la positivitats garanteix que el denominador no s'anul·la quasi segurament. Prenent esperança i aplicant de nou les esperances iterades,

$$\mathbb{E} \left[\frac{W_i Y_i^{\text{obs}}}{e(\mathbf{X}_i)} \right] = \mathbb{E}_{\text{sp}} [\mathbb{E}_{\text{sp}}[Y_i(1) \mid \mathbf{X}_i]] = \mathbb{E}_{\text{sp}}[Y_i(1)].$$

De manera anàloga,

$$\mathbb{E} \left[\frac{(1 - W_i) Y_i^{\text{obs}}}{1 - e(\mathbf{X}_i)} \right] = \mathbb{E}_{\text{sp}}[Y_i(0)].$$

Finalment, per la linealitat de l'esperança i perquè les unitats tenen la mateixa llei,

$$\mathbb{E}[\hat{\tau}_{\text{IPW}}] = \mathbb{E}_{\text{sp}}[Y_i(1)] - \mathbb{E}_{\text{sp}}[Y_i(0)] = \text{PATE}.$$

■

A.5

CAS PRÀCTIC

En les demostracions d'aquesta secció fem servir la convenció d'assignacions de 1.2 i la convenció operativa de 5.1 per a restriccions, diferències d'assignacions, intervencions i marginalitzacions.

Demostració de 5.10. **Nota:** Seguim la demostració de [Bar+20, Ap. D, Th. 11], en lloc de l'original [TP02, Lem. 1].

Fixem l'ordre topològic $<$ de 5.1 i escrivim $\mathcal{V}_{>i} := \mathcal{V} \setminus (\mathcal{V}^{(i-1)} \cup \{V_i\})$, $\mathbf{v}_{>i} := \mathbf{v}_{\mathcal{V}_{>i}}$ i altres expressions similars (aquesta és només una abreviatura per no carregar la notació). Per la regla de la cadena aplicada a la distribució observacional,

$$p(\mathbf{v}) = \prod_{i=1}^n p(v_i \mid \mathbf{v}^{(i-1)}). \quad (\text{A.7})$$

Ara introduïm una intervenció sobre totes les variables posteriors a V_i . En el graf $\mathcal{G}_{\overline{\mathcal{V}_{>i}}}$, eliminem les arestes entrants a $\mathcal{V}_{>i}$; com que l'ordre $<$ és topològic, no hi pot haver cap camí actiu de V_i cap a una variable de $\mathcal{V}_{>i}$ un cop condicionem per $\mathcal{V}^{(i-1)}$. Per tant, $V_i \perp\!\!\!\perp_{\mathcal{G}_{\overline{\mathcal{V}_{>i}}}} \mathcal{V}_{>i} \mid \mathcal{V}^{(i-1)}$. Aplicant la tercera regla del do-calculus a cada factor $p(v_i \mid \mathbf{v}^{(i-1)})$ de (A.7), obtenim:

$$p(\mathbf{v}) = \prod_{i=1}^n p(v_i \mid \text{do}(\mathbf{v}_{>i}), \mathbf{v}^{(i-1)}). \quad (\text{A.8})$$

Fixem ara un índex i . Sigui $\mathbf{C}_i$ la C-component maximal de $\mathcal{G}$ que conté V_i i siguin també $\mathbf{C}_{<i} := \mathbf{C}_i \cap \mathcal{V}^{(i-1)}$ i $\mathbf{c}_{<i} := \mathbf{v}_{\mathbf{C}_{<i}}$. També posem $\mathbf{W}_i := \mathcal{V}^{(i-1)} \setminus \mathbf{C}_{<i}$. Volem substituir l'observació de les variables de $\mathbf{W}_i$ per una intervenció. Per la segona regla del do-calculus, amb $\mathbf{X} = \mathcal{V}_{>i}$, $\mathbf{Z} = \mathbf{W}_i$ i $\mathbf{W} = \mathbf{C}_{<i}$, n'hi ha prou amb

provar

$$V_i \perp_{\mathcal{G}_{\overline{\mathcal{V}_{>i}, \mathbf{W}_i}}} \mathbf{W}_i \mid \mathcal{V}_{>i}, \mathbf{C}_{<i}. \quad (\text{A.9})$$

Suposem, per contradicció, que hi ha un camí actiu p en $\mathcal{G}_{\overline{\mathcal{V}_{>i}, \mathbf{W}_i}}$ entre algun $W \in \mathbf{W}_i$ i V_i , condicionant per $\mathcal{V}_{>i} \cup \mathbf{C}_{<i}$. Com que les arestes sortints de $\mathbf{W}_i$ han estat eliminades, el camí només pot sortir de W per una aresta que entra a W . Si aquesta primera aresta fos dirigida, anomenem-la $T \rightarrow W$, aleshores $T < W < V_i$, de manera que $T \in \mathcal{V}^{(i-1)}$. Si $T \in \mathbf{W}_i$, l'aresta $T \rightarrow W$ hauria estat eliminada; si $T \in \mathbf{C}_{<i}$, el camí quedaria blocat perquè condicionem per $\mathbf{C}_{<i}$. Així, la primera aresta no pot ser dirigida i ha de ser bidirigida.

El camí p tampoc no pot estar format totalment per arestes bidirigides, ja que això faria que W i V_i pertanyessin a la mateixa C-component, contradient $W \notin \mathbf{C}_{<i}$. Per tant, en algun punt del camí hi ha una primera aresta dirigida després d'una successió d'arestes bidirigides. Només hi ha dues possibilitats, $\dots \leftrightarrow Z \rightarrow T$ o $\dots \leftrightarrow Z \leftarrow T$:

- Si $\dots \leftrightarrow Z \rightarrow T$, aleshores l'aresta $Z \rightarrow T$ només pot romandre en $\mathcal{G}_{\overline{\mathcal{V}_{>i}, \mathbf{W}_i}}$ si $Z \notin \mathbf{W}_i$. A més, Z no pot ser a $\mathcal{V}_{>i}$, perquè Z seria un no-col·lisionador observat²⁷. Per tant, ha de ser $Z \in \mathbf{C}_{<i}$, però el camí queda blocat perquè condicionem per $\mathbf{C}_{<i}$.
- Si $\dots \leftrightarrow Z \leftarrow T$, llavors Z no pot ser a $\mathcal{V}_{>i}$, perquè estem a $\mathcal{G}_{\overline{\mathcal{V}_{>i}, \mathbf{W}_i}}$. Així, $Z \in \mathcal{V}^{(i-1)}$, i l'ordre topològic força també $T \in \mathcal{V}^{(i-1)}$. Perquè l'aresta $T \rightarrow Z$ no hagi estat eliminada per $\mathbf{W}_i$, cal que $T \notin \mathbf{W}_i$ i, per tant, $T \in \mathbf{C}_{<i}$. Però aleshores el camí queda blocat en T , que és un no-col·lisionador observat.

En tots els casos obtenim una contradicció. Això prova (A.9). Aplicant, ara sí, la segona regla del do-calculus a (A.8), obtenim:

$$p(v_i \mid \text{do}(\mathbf{v}_{>i}), \mathbf{v}^{(i-1)}) = p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{c}_{<i} \setminus v_i), \mathbf{c}_{<i}). \quad (\text{A.10})$$

Eliminem de la intervenció les variables de la mateixa C-component que venen després de V_i . Sigui $\mathbf{C}_{>i} := \mathbf{C}_i \cap \mathcal{V}_{>i}$. En el terme de la dreta de (A.10), les variables de $\mathcal{V} \setminus \mathbf{C}_i$ i de $\mathbf{C}_{>i}$ estan intervingudes, mentre que les de $\mathbf{C}_{<i}$ estan observades. Per la tercera regla del do-calculus, volem provar

$$V_i \perp_{\mathcal{G}_{\overline{\mathcal{V} \setminus \mathbf{C}_i, \mathbf{C}_{>i}}}} \mathbf{C}_{>i} \mid \mathcal{V} \setminus \mathbf{C}_i, \mathbf{C}_{<i}.$$

Aquí $\mathbf{C}_{>i}$ no conté avantpassats de $\mathbf{C}_{<i}$ en $\mathcal{G}_{\overline{\mathcal{V} \setminus \mathbf{C}_i, \mathbf{C}_{>i}}}$, perquè l'ordre $<$ és topològic. Per veure la separació, suposem que hi ha un camí actiu p entre V_i i algun $T \in \mathbf{C}_{>i}$ en el graf anterior, condicionant per $\mathcal{V} \setminus \mathbf{C}_i$ i $\mathbf{C}_{<i}$. El camí no pot entrar a T per una aresta dirigida, ja que totes les arestes entrants a $\mathbf{C}_{>i}$ han estat eliminades. Tampoc no pot sortir de la C-component $\mathbf{C}_i$: qualsevol node de $\mathcal{V} \setminus \mathbf{C}_i$ que aparegui com a no-col·lisionador està condicionat, i qualsevol desviació que torni a entrar a $\mathbf{C}_i$ ho ha de fer travessant una aresta dirigida o bidirigida que deixa un d'aquests nodes com a no-col·lisionador observat.

Finalment, si el camí roman dins de $\mathbf{C}_i$, per arribar d'un node no posterior V_i a un node posterior T sense usar una aresta entrant a $\mathbf{C}_{>i}$ ha de passar per algun predecessor de V_i de la mateixa C-component;

²⁷ Recordem de 1.64 que si observem els mediadors d'una cadena o la causa comuna d'una bifurcació, el camí que els uneix queda bloquejat.

aquest predecessor pertany a $\mathbf{C}_{<i}$ i està condicionat. En tots els casos el camí queda blocat, contradient que fos actiu. Així, podem aplicar la tercera regla:

$$p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{c}_{<i} \setminus v_i), \mathbf{c}_{<i}) = p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{c}_i), \mathbf{c}_{<i}), \quad (\text{A.11})$$

on $\mathbf{c}_i = \mathbf{v}_{\mathbf{C}_i}$. Substituint (A.10) i (A.11) a (A.8), obtenim

$$p(\mathbf{v}) = \prod_{i=1}^n p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{c}_i), \mathbf{c}_{<i}).$$

Ara reagrupem el producte segons la partició $C(\mathcal{G}) = \{\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]\}$. Si $V_i \in \mathbf{S}_j$, aleshores $\mathbf{C}_i = \mathbf{S}_j$ i $\mathbf{c}_{<i} = \mathbf{s}_j^{(i-1)}$. Així,

$$p(\mathbf{v}) = \prod_{j=1}^k \prod_{\{i \mid V_i \in \mathbf{S}_j\}} p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j), \mathbf{s}_j^{(i-1)}). \quad (\text{A.12})$$

Per a cada j , la intervenció $\text{do}(\mathbf{v} \setminus \mathbf{s}_j)$ imposa valors a totes les variables de fora de $\mathbf{S}_j$. Per tant, la regla de la cadena dins de la distribució intervinguda dona

$$\prod_{\{i \mid V_i \in \mathbf{S}_j\}} p(v_i \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j), \mathbf{s}_j^{(i-1)}) = p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)). \quad (\text{A.13})$$

Combinant (A.12) i (A.13), concloem

$$p(\mathbf{v}) = \prod_{j=1}^k p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)).$$

A més, els passos factor a factor que condueixen a (A.8), (A.10) i (A.11) identifiquen els condicionals intervinguts de (A.13) amb els condicionals observacionals de la regla de la cadena. Per tant, cada factor és identificable a partir de la distribució observacional:

$$p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)) = \prod_{\{i \mid V_i \in \mathbf{S}_j\}} p(v_i \mid \mathbf{v}^{(i-1)}).$$

Finalment, la formulació amb parets esteses és exactament la segona part de [TP02, Lem. 1]. En la seva notació, si T_i és la C-component de V_i en $\mathcal{G}[\mathcal{V}^{(i)}]$ i $\text{Pa}(T_i)$ designa T_i juntament amb els seus pares dirigits, aleshores

$$p(v_i \mid \mathbf{v}^{(i-1)}) = p(v_i \mid \mathbf{v}_{\text{Pa}(T_i) \setminus \{V_i\}})$$

per als valors on el condicional està ben definit. Ara bé, amb la definició 2.23, $T_i = C^{(i)}$ i

$$\text{Pa}(T_i) \setminus \{V_i\} = \text{Pa}_i^+.$$

Per tant, per a qualsevol versió fixada dels condicionals sobre el suport corresponent,

$$p(v_i \mid \mathbf{v}^{(i-1)}) = p(v_i \mid \mathbf{v}_{\text{Pa}_i^+}),$$

que és la formulació equivalent de l'enunciat. ■

Demostració de 5.12. Posem $\mathbf{W} = \mathcal{V} \setminus \mathbf{A}$. El submodel $\mathcal{M}_{\mathbf{w}}$, obtingut intervenint $\mathbf{W}$, indueix el graf $\mathcal{G}[\mathcal{V} \setminus \mathbf{W}] = \mathcal{G}[\mathbf{A}]$. A més, com que els nodes de $\mathbf{W}$ no són avantpassats de $\mathbf{Y}$, intervenir-los no canvia la funció marginal sobre $\mathbf{A}$ $p(\mathbf{a} \mid \text{do}(\mathbf{w})) = p(\mathbf{a})$ ²⁸.

Escrivim $\mathbf{X} = \mathbf{X}_{\mathbf{A}} \sqcup \mathbf{X}_{\mathbf{W}}$, amb $\mathbf{X}_{\mathbf{W}} = \mathbf{X} \setminus \mathbf{A}$. Les variables de $\mathbf{X}_{\mathbf{W}}$ tampoc són avantpassades de $\mathbf{Y}$; per tant, intervenir-les no afecta. Notem que $\mathbf{Y} \perp_{\mathcal{G}_{\mathbf{X}_{\mathbf{A}}, \mathbf{X}_{\mathbf{W}}}} \mathbf{X}_{\mathbf{W}} \mid \mathbf{X}_{\mathbf{A}}$: en efecte, no hi ha camins dirigits des de $\mathbf{X}_{\mathbf{W}}$ cap a $\mathbf{Y}$, perquè $\mathbf{X}_{\mathbf{W}} \cap \text{An}_{\mathbf{Y}}^{\mathcal{G}} = \emptyset$, i els camins *back-door* queden tallats en considerar el graf mutilat $\mathcal{G}_{\mathbf{X}_{\mathbf{A}}, \mathbf{X}_{\mathbf{W}}}$. Per tant,

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = p(\mathbf{y} \mid \text{do}(\mathbf{x}_{\mathbf{A}}), \text{do}(\mathbf{x}_{\mathbf{W}})) = p(\mathbf{y} \mid \text{do}(\mathbf{x}_{\mathbf{A}})) = p(\mathbf{y} \mid \text{do}(\mathbf{x}_{\mathbf{A}}), \text{do}(\mathbf{w})).$$

La darrera igualtat torna a ser una aplicació de la regla 3, ara amb $\mathbf{Y} \perp_{\mathcal{G}_{\mathbf{X}_{\mathbf{A}}, \mathbf{W}}} \mathbf{W} \mid \mathbf{X}_{\mathbf{A}}$. El raonament és el mateix: com que $\mathbf{W} = \mathcal{V} \setminus \text{An}_{\mathbf{Y}}^{\mathcal{G}}$, no hi ha camins dirigits de $\mathbf{W}$ cap a $\mathbf{Y}$ i no hi ha camins *back-door* en $\mathcal{G}_{\mathbf{W}}$. La darrera expressió és precisament la consulta $p(\mathbf{y} \mid \text{do}(\mathbf{x}_{\mathbf{A}}))$ en el submodel $\mathcal{M}_{\mathbf{w}}$. ■

Demostració de 5.13. Per definició, cap node de $\mathbf{W}$ és avantpassat de $\mathbf{Y}$ en $\mathcal{G}_{\mathbf{X}}$. Volem veure que $\mathbf{Y} \perp_{\mathcal{G}_{\mathbf{X}, \mathbf{W}}} \mathbf{W} \mid \mathbf{X}$: en efecte, en $\mathcal{G}_{\mathbf{X}}$ no hi ha cap camí dirigit des de $\mathbf{W}$ fins a $\mathbf{Y}$, ja que això faria que algun node de $\mathbf{W}$ fos avantpassat de $\mathbf{Y}$ en aquest graf. D'altra banda, en passar a $\mathcal{G}_{\mathbf{X}, \mathbf{W}}$ s'eliminen les arestes entrants a $\mathbf{W}$, de manera que els camins *back-door* de $\mathbf{W}$ amb $\mathbf{Y}$ també queden tallats. La regla 3 del *do-calculus* ens dona:

$$p(\mathbf{y} \mid \text{do}(\mathbf{x}), \text{do}(\mathbf{w})) = p(\mathbf{y} \mid \text{do}(\mathbf{x})). \quad \blacksquare$$

Demostració de 5.14. Primer considerem el cas $\mathbf{X} = \emptyset$. Per a cada component $\mathbf{S}_j$, sigui $\mathbf{A}_j = \text{An}_{\mathbf{S}_j}^{\mathcal{G}} \setminus \mathbf{S}_j$, i escrivim $\mathbf{s}_j = \mathbf{v}_{\mathbf{S}_j}$ i $\mathbf{a}_j = \mathbf{v}_{\mathbf{A}_j}$. Seguint [SP06, Lem. 4], tenim la cadena d'identitats:

$$\begin{aligned} \prod_{i=1}^k p(\mathbf{s}_i \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_i)) &= \prod_{i=1}^k p_{\mathbf{v} \setminus \mathbf{s}_i}(\mathbf{s}_i) = \prod_{i=1}^k p(\mathbf{s}_i \mid \text{do}(\mathbf{a}_i)) = \prod_{i=1}^k \prod_{\{j \mid \mathbf{V}_j \in \mathbf{S}_i\}} p(v_j \mid \mathbf{v}^{(j-1)} \setminus \mathbf{a}_i, \text{do}(\mathbf{a}_i)) \\ &= \prod_{i=1}^k \prod_{\{j \mid \mathbf{V}_j \in \mathbf{S}_i\}} p(v_j \mid \mathbf{v}^{(j-1)}) = \prod_{i=1}^n p(v_i \mid \mathbf{v}^{(i-1)}) = p(\mathbf{v}). \end{aligned}$$

Si escrivim $\mathbf{B}_i = \mathcal{V} \setminus (\mathbf{S}_i \cup \mathbf{A}_i)$, volem provar que $\mathbf{S}_i \perp_{\mathcal{G}_{\mathbf{A}_i, \mathbf{B}_i}} \mathbf{B}_i \mid \mathbf{A}_i$: en efecte, els nodes de $\mathbf{B}_i$ no són avantpassats de $\mathbf{S}_i$ i, en el graf mutilat, els camins *back-door* que podrien sortir de $\mathbf{B}_i$ queden tallats. Estem, doncs, en condicions d'aplicar la tercera regla del *do-calculus* i eliminar de la intervenció les variables de $\mathcal{V} \setminus (\mathbf{S}_i \cup \mathbf{A}_i)$.

La identitat següent és la regla de la cadena dins de cada component, però aplicada sota la intervenció sobre $\mathbf{A}_i$; per això, en el condicionament de $\mathbf{V}_j$ només apareixen els predecessors que no han rebut un valor per intervenció. Per justificar el pas als condicionals observacionals, considerem els dos casos de [SP06, Lem. 4].

1. Si $A \in \mathbf{A}_i \setminus \mathcal{V}^{(j-1)}$, tenim que $\mathbf{V}_j \perp_{\mathcal{G}_{\mathbf{A}_i}} A \mid \mathcal{V}^{(j-1)}$; per tant, la intervenció sobre A s'elimina per la tercera regla del *do-calculus*.

²⁸ La reducció és anàloga a la de [SP06, Lem. 5]. Aquesta és una aplicació de la regla 3 del *do-calculus*: no hi ha cap camí dirigit de $\mathbf{W}$ cap a $\mathbf{A}$, i els camins *back-door* queden tallats en el graf mutilat corresponent.

2. Si $A \in \mathbf{A}_i \cap \mathcal{V}^{(j-1)}$, podem aplicar la segona regla del *do-calculus* perquè es dona $V_j \perp_{\mathcal{G}_{\mathbf{A}_i \setminus \{A\}, A}} A \mid \mathcal{V}^{(j-1)} \setminus \mathbf{A}_i$ ²⁹.

Aplicant aquest argument a cada $A \in \mathbf{A}_i$ s'obté $p(v_j \mid \mathbf{v}^{(j-1)})$. Les últimes identitats només reagrupen els termes i apliquen de nou la regla de la cadena. Per tant, en el cas $\mathbf{X} = \emptyset$ hem provat la factorització de $p(\mathbf{v})$; marginalitzant les variables de $\mathcal{V} \setminus \mathbf{Y}$ s'obté l'enunciat per a $p(\mathbf{y})$.

Ara passem al cas general, amb $\mathbf{X} \neq \emptyset$. El submodel $\mathcal{M}_{\mathbf{x}}$ indueix el graf $\mathcal{G}[\mathcal{V} \setminus \mathbf{X}]$ i la PMF intervinguda $p_{\mathbf{x}}(\mathbf{v} \setminus \mathbf{x})$; les seves C-components són precisament $\mathcal{G}[\mathbf{S}_1], \dots, \mathcal{G}[\mathbf{S}_k]$. Aplicant al submodel el cas sense intervencions que acabem de provar, obtenim primer

$$p_{\mathbf{x}}(\mathbf{v} \setminus \mathbf{x}) = \prod_{j=1}^k p_{\mathbf{x}}(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus (\mathbf{x} \cup \mathbf{s}_j))).$$

Com que $\mathbf{X} \subseteq \mathcal{V} \setminus \mathbf{S}_j$ per a tot j , $\text{do}(\mathbf{x})$ ja queda inclòs en el terme que imposa valors a totes les variables de $\mathcal{V} \setminus \mathbf{S}_j$. Per tant,

$$p(\mathbf{v} \setminus \mathbf{x} \mid \text{do}(\mathbf{x})) = \prod_{j=1}^k p(\mathbf{s}_j \mid \text{do}(\mathbf{v} \setminus \mathbf{s}_j)).$$

Finalment,

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{y} \cup \mathbf{x})} p(\mathbf{v} \setminus \mathbf{x} \mid \text{do}(\mathbf{x})),$$

i substituint-hi la factorització anterior obtenim l'expressió de l'enunciat. ■

Demostració de 5.15. Treballem en la crida recursiva concreta on s'executa la línia de fallada. Denotem per $\mathbf{Y}_\ell, \mathbf{X}_\ell, \mathcal{G}_\ell = (\mathcal{V}_\ell, \mathcal{E}_\ell)$ els conjunts i el graf d'aquesta crida, i per $\mathbf{S}_\ell$ el conjunt tal que la fallada retorna $\text{FAIL}(\mathcal{G}_\ell, \mathcal{G}_\ell[\mathbf{S}_\ell])$. Just abans de fallar tenim

$$C(\mathcal{G}_\ell[\mathcal{V}_\ell \setminus \mathbf{X}_\ell]) = \{\mathcal{G}_\ell[\mathbf{S}_\ell]\}, \quad C(\mathcal{G}_\ell) = \{\mathcal{G}_\ell\}.$$

Així, $\mathbf{S}_\ell = \mathcal{V}_\ell \setminus \mathbf{X}_\ell$, el subgraf $\mathcal{G}_\ell[\mathbf{S}_\ell]$ és una única C-component i $\mathcal{G}_\ell$ també ho és. A més, $\mathbf{X}_\ell \neq \emptyset$, perquè altrament s'hauria executat el cas base de la primera línia.

Com que no s'ha executat la reducció a avantpassats, tenim $\mathcal{V}_\ell = \text{An}_{\mathbf{Y}_\ell}^{\mathcal{G}_\ell}$. Per tant, cada node de $\mathbf{X}_\ell$ té un camí dirigit en $\mathcal{G}_\ell$ cap a algun node de $\mathbf{Y}_\ell$. D'altra banda, com que tampoc s'ha executat la línia que afegeix $\mathbf{W}$, es compleix

$$\mathbf{S}_\ell = \mathcal{V}_\ell \setminus \mathbf{X}_\ell \subseteq \text{An}_{\mathbf{Y}_\ell}^{(\mathcal{G}_\ell)_{\overline{\mathbf{X}_\ell}}}.$$

En particular, cada node de $\mathbf{S}_\ell \setminus \mathbf{Y}_\ell$ té un camí dirigit cap a $\mathbf{Y}_\ell$ en $(\mathcal{G}_\ell)_{\overline{\mathbf{X}_\ell}}$; com que les arestes entrants a $\mathbf{X}_\ell$ han estat eliminades, aquests camins romanen dins de $\mathbf{S}_\ell$.

Construïm ara un subgraf $\mathcal{F}$ d'arestes de $\mathcal{G}_\ell$. Conservem totes les arestes bidirigides. Per a cada node de $\mathbf{S}_\ell \setminus \mathbf{Y}_\ell$, conservem una única aresta dirigitada que sigui el primer pas d'un camí dirigit dins de $(\mathcal{G}_\ell)_{\overline{\mathbf{X}_\ell}}$.

²⁹ Per veure-la, considerem qualsevol camí de *back-door* des de A fins a V_j . Si el camí té un node que no es troba a $\mathcal{V}^{(j-1)}$, per recursivitat, ha de contenir un col·lisionador bloquejat; per tant, tal camí queda *d*-separat. A més, com que tots els nodes del camí estan condicionats o intervinguts, el camí només pot usar arestes bidirigides. Però $\mathbf{A}_i \cap \mathbf{S}_i = \emptyset$, de manera que aquests camins bidirigits també queden bloquejats.

i acabat en Y_ℓ . Per a cada node de X_ℓ , conservem una única aresta dirigida que sigui el primer pas d'un camí dirigit dins de $\mathcal{G}_\ell$ i acabat en Y_ℓ . Eliminem totes les altres arestes dirigides; en particular, no conservem cap aresta sortint dels nodes de Y_ℓ .

Com que hem conservat totes les arestes bidirigides i $\mathcal{G}_\ell$ és una única C-component, $\mathcal{F}$ continua sent una C-component. A més, cada node de $\mathcal{V}_\ell \setminus Y_\ell$ té exactament un fill dirigit en $\mathcal{F}$, mentre que els nodes de Y_ℓ no en tenen cap. Així, $\mathcal{F}$ és un C-bosc arrelat en Y_ℓ .

Definim $\mathcal{F}'$ com el subgraf d'arestes de $\mathcal{F}$ induït pels vèrtexs S_ℓ ; és a dir,

$$\mathcal{V}_{\mathcal{F}'} = S_\ell, \quad \mathcal{E}_{\mathcal{F}'} = \mathcal{E}_{\mathcal{F}} \cap \mathcal{E}(\mathcal{G}_\ell[S_\ell]).$$

Per construcció, les arestes dirigides que surten de nodes de $S_\ell \setminus Y_\ell$ ja havien estat triades dins de S_ℓ , i no hi ha arestes sortints dels nodes de Y_ℓ . Per tant, passar de $\mathcal{F}$ a $\mathcal{F}'$ no crea cap arrel nova fora de Y_ℓ . Com que $\mathcal{G}_\ell[S_\ell]$ és una C-component i només hem eliminat arestes dirigides, $\mathcal{F}'$ també és un C-bosc arrelat en el mateix conjunt d'arrels Y_ℓ . A més, $\mathcal{F}' \subseteq \mathcal{F}$.

Ara verifiquem les condicions de 5.6. Primer, $\mathcal{F}'$ no conté variables intervingudes, ja que $\mathcal{V}_{\mathcal{F}'} = S_\ell = \mathcal{V}_\ell \setminus X_\ell$; per tant, $\mathcal{V}_{\mathcal{F}'} \cap X_\ell = \emptyset$. Segon, $\mathcal{F}$ sí que conté variables de X_ℓ , perquè $\mathcal{V}_{\mathcal{F}} = \mathcal{V}_\ell$ i $X_\ell \neq \emptyset$. Finalment, el conjunt d'arrels és Y_ℓ , i cada node de Y_ℓ és avantpassat de si mateix en $(\mathcal{G}_\ell)_{\overline{X_\ell}}$; per això

$$Y_\ell \subseteq \text{An}_{Y_\ell}^{(\mathcal{G}_\ell)_{\overline{X_\ell}}}.$$

Per tant, $\mathcal{F}$ i $\mathcal{F}'$ formen un *hedge* per a la consulta local $p(y_\ell \mid \text{do}(x_\ell))$.

El pas de la consulta local a una subconsulta de la crida inicial requereix tenir en compte la propagació de la fallada al llarg de l'algorisme. En particular, la línia que afegeix W pot modificar el conjunt d'intervencions local, i les crides que descomponen en C-components poden substituir temporalment l'objectiu per una component que després es marginalitza. Per això no identifiquem simplement X_ℓ i Y_ℓ amb subconjunts de les variables inicials.

La part global de [SP06, Th. 6] prova exactament el que falta: qualsevol fallada local d'ID es pot rastrejar fins a una subconsulta de la consulta inicial. És a dir, existeixen $X' \subseteq X$ i $Y' \subseteq Y$ tals que els C-boscobos obtinguts a la crida de fallada formen un *hedge* per a $p(y' \mid \text{do}(x'))$. Amb això, la parella $\mathcal{F}, \mathcal{F}'$ construïda més amunt prova el resultat. ■

Demostració de 5.16. En aquestes línies de l'algorisme tenim que $S = \mathcal{V} \setminus X$ i, per tant, $\mathcal{V} = S \sqcup X$. També sabem que no s'ha activat la reducció a avantpassats; així, $\mathcal{V} = \text{An}_{\mathcal{V}}^{\mathcal{G}}$. Com que $Y \subseteq S$, això implica $X \subseteq \text{An}_S^{\mathcal{G}}$.

Ara fem servir que $S \subset S'$. Com que $\mathcal{V} = S \sqcup X$, els nodes que queden fora de S' són

$$\mathcal{V} \setminus S' = X \setminus S' = X_{\text{ext}}.$$

En intervenir sobre X_{ext} amb valors x_{ext} , les variables que resten en el submodel són $\mathcal{V} \setminus X_{\text{ext}} = S'$ i el graf corresponent és $\mathcal{G}[S']$, tal com estableix [SP06, Lem. 8].

La distribució d'aquest submodel sobre $\mathbf{S}'$ és precisament el factor que l'algorisme anomena $\tilde{q}$. En efecte, com que $\mathcal{G}[\mathbf{S}']$ és una C-component maximal de $\mathcal{G}$, el lema 5.10 dona

$$p(\mathbf{s}' \mid \text{do}(\mathbf{v} \setminus \mathbf{s}')) = \prod_{\{i \mid V_i \in \mathbf{S}'\}} p(v_i \mid \mathbf{s}'^{(i-1)}, \mathbf{v}^{(i-1)} \setminus \mathbf{s}').$$

En aquesta crida, les variables de $\mathcal{V} \setminus \mathbf{S}' = \mathbf{X}_{\text{ext}}$ ja han rebut valors per intervenció. En particular, $\mathcal{V}^{(i-1)} \setminus \mathbf{S}' \subseteq \mathbf{X}_{\text{ext}}$, i els seus valors són, amb la convenció notacional del capítol, $\mathbf{v}^{(i-1)} \setminus \mathbf{s}'$. Així obtenim la distribució inicial del problema reduït:

$$\tilde{q}(\mathbf{s}') = \prod_{\{i \mid V_i \in \mathbf{S}'\}} p(v_i \mid \mathbf{s}'^{(i-1)}, \mathbf{v}^{(i-1)} \setminus \mathbf{s}') = p(\mathbf{s}' \mid \text{do}(\mathbf{x}_{\text{ext}})).$$

En la notació de [SP06, Lem. 8], aquesta és la distribució P' amb què es fa la crida recursiva.

La intervenció original es descompon com $\mathbf{x} = (\mathbf{x}_{\text{ext}}, \mathbf{x}_{\text{int}})$. La part externa ja ha quedat incorporada en $\tilde{q}$, i la part interna és la que encara s'ha d'intervenir dins de $\mathcal{G}[\mathbf{S}']$. Per tant, la consulta original en $(\mathcal{G}, p)$ coincideix amb la consulta reduïda en $(\mathcal{G}[\mathbf{S}'], \tilde{q})$:

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \tilde{p}(\mathbf{y} \mid \text{do}(\mathbf{x}_{\text{int}})).$$

Escrivint aquesta última marginal de manera explícita, i usant que $\mathbf{S}' = \mathbf{S} \sqcup \mathbf{X}_{\text{int}}$ i $\mathbf{Y} \subseteq \mathbf{S}$, obtenim

$$p(\mathbf{y} \mid \text{do}(\mathbf{x})) = \sum_{\mathbf{s} \setminus \mathbf{y}} \tilde{p}(\mathbf{s} \mid \text{do}(\mathbf{x}_{\text{int}})).$$

Aquesta és exactament la marginal que calcula la crida $\text{ID}(\mathbf{Y}, \mathbf{X}_{\text{int}}, \tilde{q}, \mathcal{G}[\mathbf{S}'])$, que és la formulació de [SP06, Lem. 8]. ■

Demostració de 5.27. Siguin $\mathcal{Z}_+ = \{z \mid P(Z = z) > 0\}$. Per la llei de la probabilitat total sota cada intervenció,

$$\begin{aligned} P(S = 1 \mid \text{do}(T = 1)) &= \sum_{z \in \mathcal{Z}_+} P(S = 1 \mid \text{do}(T = 1), Z = z) P(Z = z \mid \text{do}(T = 1)), \\ P(S = 1 \mid \text{do}(T = 0)) &= \sum_{z \in \mathcal{Z}_+} P(S = 1 \mid \text{do}(T = 0), Z = z) P(Z = z \mid \text{do}(T = 0)). \end{aligned}$$

Per hipòtesi, $P(Z = z \mid \text{do}(T = 1)) = P(Z = z \mid \text{do}(T = 0)) = P(Z = z)$, per a tot z . Per tant, la diferència $P(S = 1 \mid \text{do}(T = 1)) - P(S = 1 \mid \text{do}(T = 0))$ ens dona:

$$\sum_{z \in \mathcal{Z}_+} \left[P(S = 1 \mid \text{do}(T = 1), Z = z) - P(S = 1 \mid \text{do}(T = 0), Z = z) \right] P(Z = z).$$

Per hipòtesi, cada diferència entre claudàtors és estrictament positiva per a tot $z \in \mathcal{Z}_+$, i els pesos $P(Z = z)$ són positius i sumen 1. Així, la suma és estrictament positiva. En conseqüència, $P(S = 1 \mid \text{do}(T = 1)) > P(S = 1 \mid \text{do}(T = 0))$. ■

MESURA I FUNCIONS MESURABLES

En aquesta secció es revisen els conceptes bàsics de mesura, σ -àlgebres i funcions mesurables. Aquest marc matemàtic és fonamental per definir espais de probabilitat i variables aleatòries després.

La presentació d'aquesta secció segueix principalment les parts de mesura de Lebesgue, mesurabilitat, funcions mesurables i integral de Lebesgue de [Vil23, Secs. 1–3].

Definició B.1 (Mesura i interval a $\mathbb{R}^n$ [Vil23, Sec. 1, Def. 1.1]). La mesura (longitud) d'un interval $[a, b]$ (o bé $[a, b)$, $(a, b]$, o (a, b)) a $\mathbb{R}$ és $|[a, b]| = b - a$. Diem *interval a $\mathbb{R}^n$* a qualsevol conjunt de la forma $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$, amb $a_1 < b_1, \dots, a_n < b_n$, o a qualsevol obtingut de manera semblant reemplaçant qualsevol dels intervals $[a_j, b_j]$ per $[a_j, b_j)$, $(a_j, b_j]$ o bé (a_j, b_j) .

L'interval és *obert* si els intervals a $\mathbb{R}$ que el produeixen són tots oberts; és a dir, si I té la forma $I = (a_1, b_1) \times \cdots \times (a_n, b_n)$. Anàlogament, I és *tancat* si és de la forma $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$.

Definició B.2 (Volum d'un interval). Sigui $I = (a_1, b_1) \times \cdots \times (a_n, b_n)$ un interval obert. Definim la *mesura exterior per intervals* o *volum d'un interval* i la denotem per $m^*(I)$ o $v(I)$ com el producte de les seves llargàries, és a dir:

$$m^*(I) = v(I) = |I|^* = (b_1 - a_1) \cdots (b_n - a_n).$$

A $\mathbb{R}^2$ un interval és un rectangle, i la seva mesura és l'àrea. A $\mathbb{R}^3$ un interval és un paral·lelepípede i la seva mesura és el volum.

Recordem que volem donar sentit a la mesura d'un conjunt arbitrari $A \subset \mathbb{R}^n$. El primer intent, que funciona gairebé sempre, és utilitzar els rectangles com a unitat de mesura, recobrint el conjunt A .

Definició B.3 (Mesura exterior de Lebesgue). Sigui $A \subset \mathbb{R}^n$. Diem *mesura exterior de Lebesgue d' A* a l'ínnim:

$$|A|^* = m^*(A) = \inf \left\{ \sum_{i=1}^{\infty} m(I_i) \mid A \subseteq \bigcup_{i=1}^{\infty} I_i, I_i \text{ interval obert} \right\}.$$

Per tant, de tots els possibles recobriments d' A per intervals, en calculem la suma de volums dels intervals i en prenem l'ínnim.

Propietat B.4.

1. La mesura exterior del conjunt buit és 0.
2. Si $A, B \subset \mathbb{R}^n$ i $A \subset B$, llavors $|A|^* \leq |B|^*$.
3. Si A és finit o numerable, $|A|^* = 0$.

Propietat B.5 (Invariant per translacions). *A més, m^* és invariant per translacions; això és, per a tot $A \subset \mathbb{R}^n$ i per a tot $x \in \mathbb{R}^n$ tenim $m^*(A + x) = m^*(A)$. Existeix un recobriment en el desplaçat si, i només si, existeix un recobriment en el mateix conjunt:*

$$A \subset \bigcup_{i=1}^{\infty} I_i \iff A + x \subset \bigcup_{i=1}^{\infty} (I_i + x)$$

També, el volum és el mateix, $v(I) = v(I + x)$.

Propietat B.6 (σ -subadditivitat). *m^* és σ -subadditiva:*

$$m^*\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} m^*(A_k), \quad \forall A_k \subset \mathbb{R}^n.$$

Nota B.7. En definitiva, existeixen $A, B \subset \mathbb{R}^n$ tal que $A \cap B = \emptyset$ i $m^*(A \cup B) \neq m^*(A) + m^*(B)$. En aquest sentit, si $d(A, B) > 0$, aleshores sí es dona que $m^*(A) + m^*(B) = m^*(A \cup B)$.

Sigui X un conjunt ambient (habitualment $X = \mathbb{R}^n$) i sigui $\mathcal{P}(X)$ el conjunt de parts d' X ; és a dir, el conjunt que té per elements els subconjunts de X .

Definició B.8 (σ -àlgebra). Si X és un espai $X = \mathbb{R}^n$, una família $\Sigma \subset \mathcal{P}(X)$ és una σ -àlgebra si:

1. $X \in \Sigma$.
2. Si $E \in \Sigma$, aleshores $E^C \in \Sigma$.
3. Si $\{E_k\}_{k=1}^{\infty} \subseteq \Sigma$, aleshores $\bigcup_{k=1}^{\infty} E_k \in \Sigma$.

Definició B.9 (Mesura positiva). Sigui X i sigui $\Sigma \subset \mathcal{P}(X)$ una σ -àlgebra. Una funció $\mu : \Sigma \longrightarrow [0, +\infty]$ és una *mesura positiva* si:

1. $\mu(\emptyset) = 0$.
2. μ és σ -additiva: si $E_j \in \Sigma$ són disjunts dos a dos, llavors:

$$\mu\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mu(E_j).$$

Definició B.10 (Continuïtat absoluta i mesura dominant [Dur19, Ap. A.4]; [Nag23]). Siguin μ i ν dues mesures positives sobre un mateix espai mesurable (X, Σ) . Diem que ν és *absolutament contínua* respecte de μ , i escrivim $\nu \ll \mu$, si per a tot $A \in \Sigma$,

$$\mu(A) = 0 \implies \nu(A) = 0.$$

Una mesura μ és σ -finita si existeixen $E_1, E_2, \dots \in \Sigma$ tals que $X = \bigcup_{n \geq 1} E_n$ i $\mu(E_n) < +\infty$ per a tot n . Donada una família de mesures $\mathcal{P}$ sobre (X, Σ) , una mesura σ -finita μ s'anomena *mesura dominant* si $P \ll \mu$ per a tota $P \in \mathcal{P}$.

Teorema B.11 (Radon–Nikodym [Dur19, Ap. A.4, Thm. A.4.8]; [Nag23]). Siguin μ i ν dues mesures positives σ -finites sobre (X, Σ) . Si $\nu \ll \mu$, aleshores existeix una funció mesurable $f : X \longrightarrow [0, +\infty]$, única

μ -gairebé pertot, tal que per a tot $A \in \Sigma$,

$$v(A) = \int_A f \, d\mu.$$

Aquesta funció s'anomena la derivada de Radon–Nikodym de v respecte de μ i s'escriu $f = \frac{dv}{d\mu}$.

Observació B.12. En probabilitat i estadística, quan una llei P està dominada per una mesura de referència μ , la seva densitat és precisament $p = \frac{dP}{d\mu}$. Així, una PMF és una densitat respecte de la mesura de comptatge, mentre que una PDF real usual és una densitat respecte de la mesura de Lebesgue [Vil24b, Sec. 2, Defs. 2.2.3 i 2.4.1].

Definició B.13 (Conjunt mesurable Lebesgue). Sigui $E \subset \mathbb{R}^n$. Diem que E és mesurable Lebesgue si per a tot $A \subset \mathbb{R}^n$ es dona el següent:

$$|A|^* = |A \cap E|^* + |A \cap E^c|^*.$$

Els conjunts E que tenen aquesta propietat formen una σ -àlgebra, denotada $\mathcal{M}(\mathbb{R}^n)$ i anomenada la σ -àlgebra de conjunts mesurables Lebesgue.

Alternativament, un conjunt A és mesurable Lebesgue si per a tot $\varepsilon > 0$ existeix un obert $\mathcal{U} \supset A$ tal que $m^*(\mathcal{U} \setminus A) \leq \varepsilon$. Si $A \subset \mathbb{R}^n$ és mesurable, definim $m(A) = |A| = m^*(A)$.

Nota B.14. Se segueix directament que qualsevol obert en $\mathbb{R}^n$ és mesurable, ja que $m^*(\mathcal{U} \setminus \mathcal{U}) = 0 \leq \varepsilon$.

Definició B.15 (Mesura de Lebesgue [Vil23, Sec. 1, Def. 1.11]). La mesura exterior restringida a $\mathcal{M}(\mathbb{R}^n)$ és una mesura positiva, anomenada *mesura de Lebesgue*.

$$\begin{aligned} |\cdot| : \mathcal{M}(\mathbb{R}^n) &\longrightarrow [0, +\infty] \\ E &\longmapsto |E| = m(E) \end{aligned}$$

Per a $E \in \mathcal{M}(\mathbb{R}^n)$ denotarem $|E|$ en lloc d' $|E|^*$ (tot i que siguin el mateix valor).

A partir d'aquí enunciem una sèrie de propietats teòriques de la mesura de Lebesgue. La majoria són força intuïtives, sobre tot si pensem que estem formalitzant la noció de volum d'un conjunt a $\mathbb{R}^n$.

Teorema B.16 (Teorema de Carathéodory).

1. $\mathcal{M}(\mathbb{R}^n)$ és una σ -àlgebra a $\mathbb{R}^n$.
2. $|\cdot|$ és una mesura positiva a $\mathcal{M}(\mathbb{R}^n)$.
3. La mesura $|\cdot|$ és completa si $z \in \mathcal{M}(\mathbb{R}^n)$ té mesura 0, $|z| = 0$ i $z' \subset z$, llavors també $z' \in \mathcal{M}(\mathbb{R}^n)$ i $|z'| = 0$.

Notació B.17 (Almost everywhere, a.e.). Quan una propietat es compleix llevat potser d'un conjunt excepcional de mesura zero, diem que es compleix *gairebé per a tot* o, en anglès, *almost everywhere*.

Propietat B.18.

1. Si $A, B \in \mathcal{M}(\mathbb{R}^n)$ i $A \subset B$, $|A| \leq |B|$.
2. Si $A, B \in \mathcal{M}(\mathbb{R}^n)$ i $A \subset B$, $|B \setminus A| = |B| - |A|$ sempre que $|A| < +\infty$.
3. Si $E_j \in \mathcal{M}(\mathbb{R}^n)$, $j \geq 1$, formen una successió ascendent, $E_j \subset E_{j+1}$, per a tot $j \geq 1$, llavors $|\bigcup_{j=1}^{\infty} E_j| = \lim_{j \rightarrow \infty} |E_j|$.

4. *Invariància per translacions i simetria respecte l'origen. Com ja hem vist per a mesures exteriors, B.5, si $E \in \mathcal{M}(\mathbb{R}^n)$, aleshores:*
- $x + E = \{x + y \mid y \in E\} \in \mathcal{M}(\mathbb{R}^n)$ i $|x + E| = |E|$.
 - $-E = \{-x \mid x \in E\} \in \mathcal{M}(\mathbb{R}^n)$ i $|-E| = |E|$ (la mesura positiva és una funció parella).
5. Si I és interval, $I \in \mathcal{M}(\mathbb{R}^n)$ i $|I| = v(I)$.

Proposició B.19.

1. Els conjunts mesurables de Lebesgue formen la σ -àlgebra $\mathcal{M}(\mathbb{R}^n)$ que conté els oberts de $\mathbb{R}^n$.
2. Tot conjunt de mesura exterior 0 és mesurable.

Teorema B.20. Sigui $E \subset \mathbb{R}^n$. Són equivalents:

1. $E \in \mathcal{M}(\mathbb{R}^n)$.
2. $\forall \varepsilon > 0$ existeix $\mathcal{T} \subset E \subset \mathcal{U}$, $\mathcal{T}$ tancat i $\mathcal{U}$ obert, tals que $m^*(\mathcal{U} \setminus \mathcal{T}) < \varepsilon$.
3. E és igual a $G \setminus Z$, on G és un conjunt G_σ i $|Z| = 0$.
4. E és igual a $F \cup Z$, on F és un conjunt F_σ i $|Z| = 0$.

Pel segon apartat, podem aproximar tot $E \in \mathcal{M}(\mathbb{R}^n)$ amb tancats per dins i amb oberts per fora.

Suposem, a partir d'ara, $E \in \mathcal{M}(\mathbb{R}^n)$; és a dir, E és un conjunt mesurable, i diem $\overline{\mathbb{R}} = [-\infty, +\infty]$. En general, tindrem funcions $f : E \rightarrow \overline{\mathbb{R}}$ i, per ser rigorosos, ens caldrà estendre la relació d'ordre de $\mathbb{R}$ a tot $\overline{\mathbb{R}}$.

$$-\infty < a < +\infty, \forall a \in \mathbb{R}.$$

També cal estendre-hi les operacions suma i multiplicació, de la manera que ja coneixem:

1. $a \pm \infty = \pm\infty + a = \pm\infty$.
2. $a \cdot (\pm\infty)$ és $+\infty$ si $a > 0$, $-\infty$ si $a < 0$ i, finalment, una indeterminació si $a = 0$.
3. La resta de casos se segueixen fàcilment prenent $\pm\infty$ en lloc d' a .

Proposició B.21. Sigui $f : E \rightarrow \overline{\mathbb{R}}$, $E \subset \mathbb{R}^n$ mesurable. Són equivalents:

1. $\forall \alpha \in \mathbb{R}$, el conjunt $f^{-1}(\alpha, +\infty] = \{x \in E \mid f(x) > \alpha\}$ és mesurable.
2. $\forall \alpha \in \mathbb{R}$, el conjunt $f^{-1}[\alpha, +\infty] = \{x \in E \mid f(x) \geq \alpha\}$ és mesurable.
3. $\forall \alpha \in \mathbb{R}$, el conjunt $f^{-1}[-\infty, \alpha) = \{x \in E \mid f(x) < \alpha\}$ és mesurable.
4. $\forall \alpha \in \mathbb{R}$, el conjunt $f^{-1}[-\infty, \alpha] = \{x \in E \mid f(x) \leq \alpha\}$ és mesurable.

Quan es compleixen aquestes condicions, també es dona que per a tot $\alpha \in \overline{\mathbb{R}}$, $f^{-1}(\{\alpha\}) = \{x \in E \mid f(x) = \alpha\}$ és mesurable.

Definició B.22 (Funció mesurable [Vil23, Sec. 2, Prop. 2.1 i Def. 2.2]). Una funció $f : E \rightarrow \overline{\mathbb{R}}$ és mesurable si E és mesurable; equivalentment, es compleixen tots els apartats de la proposició anterior.

Sigui $A \subset E$ un conjunt mesurable. Definim la funció característica $\chi_A : E \rightarrow \mathbb{R}$ mitjançant l'assignació següent:

$$x \mapsto \chi_A(x) := \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases}$$

Aleshores, fixat $\alpha \in \mathbb{R}$, els tres conjunts següents són mesurables:

$$\{x \in E \mid \chi_A(x) > \alpha\} = \begin{cases} \emptyset, & \text{si } \alpha \geq 1, \\ A, & \text{si } \alpha \in (0, 1), \\ E, & \text{si } \alpha \leq 0. \end{cases}$$

Si $f : E \rightarrow \mathbb{R}$ és contínua, $f^{-1}(\alpha, +\infty) = \{x \in E \mid f(x) > \alpha\}$ és l'antiimatge de l'obert $(\alpha, +\infty)$ i, per tant, també és un obert. En particular, és mesurable.

Proposició B.23. Si $g, f : \mathbb{R}^n \rightarrow \mathbb{R}$ són funcions mesurables i $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que $F \in C(\mathbb{R}^2)$, aleshores $F(f, g) : \mathbb{R}^n \rightarrow \mathbb{R}$ és mesurable.

Proposició B.24. Sigui $f : E \rightarrow \overline{\mathbb{R}}$ mesurable i sigui $g : E \rightarrow \overline{\mathbb{R}}$ tal que $g = f$ llevat d'un conjunt Z amb $|Z| = 0$ (és a dir, $g(x) = f(x)$ si $x \notin Z$). Llavors, g també és mesurable.

Propietat B.25. Siguin $c \in \mathbb{R}, E \in \mathcal{M}(\mathbb{R}^n)$ i $f, g : E \rightarrow \overline{\mathbb{R}}$ funcions mesurables. Aleshores:

$$cf, f + g, f \cdot g; \frac{f}{g}^{30}, \max\{f, g\}, |f|, f^+, f^-$$

són funcions mesurables.

Prenem una funció mesurable $f : E \rightarrow \overline{\mathbb{R}}$. Donada f amb signe arbitrari, escrivim $f = f^+ - f^-$. Com que $f^+ = \max\{f, 0\}$ i $f^- = \max\{-f, 0\}$, aquestes funcions són mesurables i, per tant, també ho és $|f| = f^+ + f^-$.

Ara definim la integral de f com la resta de les integrals de f^+ i f^- (cosa que sabem fer per la subsecció anterior), sempre que aquesta resta no porti a una indeterminació.

Definició B.26 (Integral de Lebesgue [Vil23, Sec. 3, Def. 3.8]). Sigui $E \in \mathcal{M}(\mathbb{R}^n)$ i sigui $f : E \rightarrow \overline{\mathbb{R}}$ mesurable tal que almenys un dels dos valors $\int_E f^+$ i $\int_E f^-$ sigui finit. Llavors, definim la *integral de Lebesgue* de f com:

$$\int_E f = \int_E f^+ - \int_E f^-.$$

Definició B.27 (Funció integrable [Vil23, Sec. 3, Def. 3.9]). Una funció $f : E \rightarrow \overline{\mathbb{R}}$ mesurable és integrable si:

$$\int_E |f| = \int_E f^+ + \int_E f^- < \infty.$$

Denotem per $\mathcal{L}^1(E)$ el conjunt de funcions integrables a E .

Proposició B.28. Sigui $E \in \mathcal{M}(\mathbb{R}^n)$, un conjunt mesurable.

1. $\mathcal{L}^1(E)$ és un espai vectorial sobre el qual la integral de Lebesgue és una aplicació lineal. És a dir, si $\alpha, \beta \in \mathbb{R}$ i $f, g \in \mathcal{L}^1(E)$, llavors $\alpha f + \beta g \in \mathcal{L}^1(E)$ i:

$$\int_E (\alpha f + \beta g) = \alpha \int_E f + \beta \int_E g.$$

³⁰ Cal suposar que $g(x) \neq 0$ almost everywhere.

2. Si $f, g \in \mathcal{L}^1(E)$ i $f \leq g$:

$$\int_E f \leq \int_E g \text{ i } \left| \int_E f \right| \leq \int_E |f|.$$

3. Invariància per translacions i simetries. Si $v \in \mathbb{R}^n$ és fix i $f \in \mathcal{L}^1(E)$, llavors $f(x-v) \in \mathcal{L}^1(E+v)$ i:

$$\int_{E+v} f(x-v) dx = \int_E f(x) dx.$$

De manera semblant, $f(-x) \in \mathcal{L}^1(-E)$ i:

$$\int_{-E} f(-x) dx = \int_E f(x) dx.$$

Aquí, $E+v = \{x+v \mid x \in E\}$ i $-E = \{-x \mid x \in E\}$.

Proposició B.29.

1. Si $f \in \mathcal{L}^1(E)$ i g és mesurable tal que $|g| \leq f$, llavors $g \in \mathcal{L}^1(E)$.
2. Si f és mesurable i acotada i $E \in \mathcal{M}(\mathbb{R}^n)$ tal que tenim $|E| < \infty$, llavors $f \in \mathcal{L}^1(E)$.

B.2

VARIABLES I VECTORS ALEATORIS

Aquesta secció fixa el llenguatge probabilístic bàsic que s'utilitza al llarg del treball. L'objectiu és recordar espais de probabilitat, variables aleatòries, lleis i distribucions conjuntes abans d'emprar-los en xarxes bayesianes i resultats potencials.

Definició B.30 (Àlgebra de Borel). La σ -àlgebra de Borel d'un espai topològic E és la σ -àlgebra generada pels seus conjunts oberts; és a dir, la més petita de les σ -àlgebres de parts de E que contenen tots els oberts d' E . La denotarem per $\mathcal{B}(E)$.

1. $E \in \mathcal{B}(E)$,
2. Si $A \in \mathcal{B}(E)$, aleshores $A^c \in \mathcal{B}(E)$,
3. Si $(A_n)_n \subset \mathcal{B}(E)$, aleshores $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}(E)$.

Definició B.31 (Espai de probabilitat [San99, Sec. 1.2]; [Vil24b, Sec. 1, Def. 1.1.1]). Un **espai de probabilitat** és una tripleta $(\Omega, \mathcal{F}, \mathbb{P})$, on:

1. Ω és un conjunt no buit anomenat *espai mostral*.
2. $\mathcal{F}$ és una σ -àlgebra de subconjunts de Ω , els elements de la qual s'anomenen *esdeveniments*.
3. $\mathbb{P}$ és una *mesura de probabilitat*, una funció $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ tal que:
 1. $\mathbb{P}(\Omega) = 1$,
 2. Si $(A_n)_n \subset \mathcal{F}$ disjunts 2 a 2, aleshores:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Treballarem ara amb un espai de probabilitat $(\Omega, \mathcal{F}, \mathbb{P})$. Tots els esdeveniments que apareguin s'assumiran tàcitament com a $\mathcal{F}$ -mesurables, totes les σ -àlgebres que apareguin s'assumiran tàcitament com a subàlgebres de $\mathcal{F}$. Al llarg del treball, quan no s'indiqui el contrari, els espais de valors de les variables aleatòries seran espais de Borel estàndard; els casos reals $\mathbb{R}^m$ que apareixen en aquest annex en són el cas principal.

Definició B.32 (Variable aleatòria [Vil24b, Sec. 2, Def. 2.1.2]). Sigui $(\Omega, \mathcal{F}, \mathbb{P})$ un espai de probabilitat i $(E, \mathcal{E})$ un espai mesurable. Una variable aleatòria amb valors en E és una aplicació mesurable $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$. És a dir, per a tot $B \in \mathcal{E}$, la preimatge $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$ és un esdeveniment, $X^{-1}(B) \in \mathcal{F}$.

Definició B.33 (Llei d'una variable aleatòria [San99, Sec. 2.2, Def. 2.2]). Sigui $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ una variable aleatòria. La llei d' X és la probabilitat P_X sobre $(E, \mathcal{E})$ definida, per a tot $B \in \mathcal{E}$, cf. [Vil24b, Sec. 2, Def. 2.1.3]:

$$P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}).$$

És clar que $P_X(E) = 1$, de manera que és una probabilitat. Finalment, sigui $\{B_n\}_{n \geq 1}$ una col·lecció d'elements disjunts de $\mathcal{E}$. Aleshores:

$$X^{-1}\left(\bigcup_{n \geq 1} B_n\right) = \bigcup_{n \geq 1} X^{-1}(B_n).$$

Definició B.34 (Variable aleatòria real). Una variable aleatòria real és una variable aleatòria amb valors en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proposició B.35. Sigui X, Y dues variables aleatòries reals definides en l'espai de probabilitat $(\Omega, \mathcal{F}, \mathbb{P})$ i $a, b \in \mathbb{R}$. Aleshores:

1. $aX + bY$ i XY són també variables aleatòries reals.
2. $\max(X, Y)$ i $\min(X, Y)$ són variables aleatòries reals.
3. Si $Y : \Omega \rightarrow (0, +\infty)$, el quocient és una variable aleatòria real.

Definició B.36 (Variable aleatòria discreta). Una variable aleatòria $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ direm que és discreta si la seva llei està concentrada en un conjunt numerable i mesurable $E_0 \in \mathcal{E}$, és a dir, si $P_X(E_0) = 1$.

Definició B.37 (Funcions de distribució). La funció de distribució associada (**funció de distribució acumulada**, CDF) a una variable aleatòria real X és la funció $F : \mathbb{R} \rightarrow [0, 1]$ definida per:

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P} \circ X^{-1}((-\infty, x]). \end{aligned}$$

És a dir, considerem la família de les semirrectes $(-\infty, x]$, $x \in \mathbb{R}$ i el valor que pren $\mathbb{P} \circ X^{-1}$ sobre aquesta família. F és una funció de distribució d'una variable aleatòria real X si, i només si:

1. F és creixent.

2. F és contínua per la dreta, és a dir, $\lim_{x_n \rightarrow x^+} F(x_n) = F(x)$.
3. $\lim_{x \rightarrow +\infty} F(x) = 1$ i $\lim_{x \rightarrow -\infty} F(x) = 0$.

Propietat B.38 (Propietats de les CDF).

1. $\mathbb{P}(X < a) = F(a^-)$,
2. $\mathbb{P}(a \leq X \leq b) = F(b) - F(a^-)$,
3. $\mathbb{P}(a < X \leq b) = F(b) - F(a)$,
4. $\mathbb{P}(a \leq X < b) = F(b^-) - F(a^-)$,
5. $\mathbb{P}(a < X < b) = F(b^-) - F(a)$.
6. $\mathbb{P}(X = a) = F(a) - F(a^-) \geq 0$.

Definició B.39 (PMF i PDF [Vil24b, Sec. 2, Defs. 2.2.3, 2.4.1 i 2.4.2]). En el cas real, la distribució d'una variable aleatòria real X pot ser caracteritzada per:

- Una **funció de massa de probabilitat** (PMF) si X és discreta, $p_X(x) = \mathbb{P}(X = x)$, que assigna una probabilitat a cada valor possible de X . Equivalentment, $p_X = \frac{dP_X}{d\kappa}$, on κ és la mesura de comptatge sobre el suport de X .
- Una funció mesurable $f : \mathbb{R} \rightarrow [0, +\infty]$ anomenada **funció de densitat de probabilitat** (PDF) si $P_X \ll \lambda_1$, on λ_1 és la mesura de Lebesgue a $\mathbb{R}$, i
 1. $f = \frac{dP_X}{d\lambda_1}$ en el sentit del teorema de Radon–Nikodym B.11,
 2. $\int_{\mathbb{R}} f d\lambda_1 = 1$,
 3. per a tot $A \in \mathcal{B}(\mathbb{R})$, $P_X(A) = \int_A f d\lambda_1$.

Definició B.40 (Variable aleatòria absolutament contínua). Es diu que una variable aleatòria real X és absolutament contínua amb PDF f si la seva funció de distribució F es pot escriure com:

$$F(x) = \int_{(-\infty, x]} f d\lambda_1,$$

per a tot $x \in \mathbb{R}$, on la funció f satisfà les condicions de la definició B.39. Equivalentment, la llei P_X és absolutament contínua respecte de la mesura de Lebesgue, $P_X \ll \lambda_1$.

Definició B.41 (Vector aleatori real). Un vector aleatori real m -dimensional és una aplicació:

$$\begin{aligned} X : (\Omega, \mathcal{F}) &\longrightarrow (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)) \\ \omega &\longmapsto X(\omega) = (X_1(\omega), \dots, X_m(\omega)) \end{aligned}$$

tal que cadascuna de les components $X_i : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ és una variable aleatòria real.

Definició B.42 (Distribució conjunta i marginal). La **distribució de probabilitat conjunta** d'un vector aleatori real X és la seva llei $P_X = \mathbb{P} \circ X^{-1}$ sobre $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Es pot descriure per la seva CDF conjunta, $F_X(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$. Si $P_X \ll \lambda_m$, on λ_m és la mesura de Lebesgue a $\mathbb{R}^m$, es descriu per la seva PDF conjunta $f_X = \frac{dP_X}{d\lambda_m}$.

$$F_X(x_1, \dots, x_m) = \int_{(-\infty, x_1] \times \dots \times (-\infty, x_m]} f_X d\lambda_m.$$

Quan F_X és prou regular, f_X coincideix λ_m -gairebé pertot amb la derivada mixta $\frac{\partial^m F_X}{\partial x_1 \cdots \partial x_m}$. La **distribució marginal** d'un subconjunt de variables s'obté integrant (o sumant) la distribució conjunta sobre els valors de les altres variables. La PDF marginal de X_i és:

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{m-1}} f_X(x_1, \dots, x_m) d\lambda_{m-1}(x_{-i}),$$

on x_{-i} denota totes les coordenades excepte x_i .

B.3

INDEPENDÈNCIA I INDEPENDÈNCIA CONDICIONAL

Aquí es revisen les nocions elementals d'independència, probabilitat condicionada i marginalització. L'objectiu és preparar les eines que després permeten interpretar factoritzacions i criteris d'ajustament causal. Arribarem als models d'independència, semigrafoïdes i grafoïdes que sustenten la lectura gràfica de les independències.

Definició B.43 (Independència estadística). Siguin $X_1, \dots, X_m$ variables aleatòries definides en un espai de probabilitat $(\Omega, \mathcal{F}, \mathbb{P})$, amb valors respectius en espais mesurables $(E_i, \mathcal{E}_i)$. Direm que són *independents* si, per a conjunts qualssevol $B_i \in \mathcal{E}_i$ es compleix que:

$$\mathbb{P}(\{X_1 \in B_1, \dots, X_m \in B_m\}) = \prod_{i=1}^m \mathbb{P}(\{X_i \in B_i\});$$

en altres paraules, si els esdeveniments $\{X_1 \in B_1\}, \dots, \{X_m \in B_m\}$ són independents. Un conjunt $\{X_n \mid n \geq 1\}$ infinit té esdeveniments independents si qualsevol subcol·lecció finita ho és.

Teorema B.44. Considerem un vector aleatori real $X = (X_1, \dots, X_m)$. Denotem per F la seva funció de distribució i per $F_i, i = 1, \dots, m$ les funcions de distribució respectives de les variables aleatòries components. Suposem que les variables aleatòries $X_1, \dots, X_m$ són independents. Aleshores, es compleix que $F(x_1, \dots, x_m) = F_1(x_1) \cdots F_m(x_m)$, per a tot $x = (x_1, \dots, x_m) \in \mathbb{R}^m$. El recíproc també és cert.

Definició B.45 (Probabilitat condicionada). Donats dos esdeveniments $A, B \in \mathcal{F}$ amb $\mathbb{P}(B) > 0$, la probabilitat condicionada d' A donat B es defineix com:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Teorema B.46 (Distribució condicional regular). Sigui $(\Omega, \mathcal{F}, \mathbb{P})$ un espai de probabilitat i siguin X, Y variables aleatòries amb valors en espais de Borel estàndard. Aleshores, existeix una versió regular de la distribució condicional d' Y donada X [San99, Sec. 1.5]; [Dur19, Sec. 4.1, Thm. 4.1.17].

Teorema B.47 (Probabilitats compostes). Considerem n esdeveniments $A_1, \dots, A_n$ tals que $\mathbb{P}(A_1 \cap \dots \cap$

$A_{n-1}) > 0$ de l'espai mostral Ω . Aleshores:

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 | A_1) \cdots \mathbb{P}(A_n | A_1 \cap \cdots \cap A_{n-1}) = \prod_{i=1}^n \mathbb{P}(A_i | A_1 \cap \cdots \cap A_{i-1}).$$

Teorema B.48 (Probabilitats totals). Si un conjunt d'esdeveniments $\{B_1, B_2, \dots, B_n\}$ forma una partició de l'espai mostral Ω , llavors per a qualsevol esdeveniment $A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \mathbb{P}(A | B_i) \mathbb{P}(B_i)$$

Aquest procés s'anomena **marginalització**. Per a variables aleatòries reals contínues X i Y , la densitat marginal de X s'obté a partir de la densitat conjunta i la condicionada:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x | y) f_Y(y) dy$$

Teorema B.49 (Regla de Bayes). Donats dos esdeveniments A i B amb $\mathbb{P}(A), \mathbb{P}(B) > 0$, la relació entre les seves probabilitats condicionades és:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B)}$$

Per a variables aleatòries reals contínues X i Y , la regla de Bayes per a les seves densitats és:

$$f_{X|Y}(x | y) = \frac{f_{Y|X}(y | x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y | x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y | x') f_X(x') dx'}$$

Definició B.50 (Independència condicional, [Lau96, Sec. 3.1, pp. 28–29]). Dos esdeveniments A, B són condicionalment independents donada una sub- σ -àlgebra $\mathcal{H} \subseteq \mathcal{F}$ si $\mathbb{P}(A \cap B | \mathcal{H}) = \mathbb{P}(A | \mathcal{H}) \cdot \mathbb{P}(B | \mathcal{H})$, quasi segurament. Normalment escriurem $(A \perp\!\!\!\perp B) | \mathcal{H}$.

Anàlogament, diem que $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ són condicionalment independents donada $\mathcal{H}$ si $(A \perp\!\!\!\perp B) | \mathcal{H}$ per a tot $A \in \mathcal{A}, B \in \mathcal{B}$.

Observació B.51. Suposem que $\mathcal{A}, \mathcal{B}$ i $\mathcal{H}$ són independents. Llavors, per a $A \in \mathcal{A}$ observem que $\mathbb{P}(A | \mathcal{H}) = \mathbb{P}(A)$ quasi segurament, ja que per a $H \in \mathcal{H}$ es compleix la relació

$$\int_H \mathbb{P}(A) d\mathbb{P} = \mathbb{P}(A) \mathbb{P}(H) = \mathbb{P}(A \cap H).$$

Així doncs, utilitzant la independència entre $\mathcal{A}$ i $\mathcal{B}$, obtenim

$$\mathbb{P}(A | \mathcal{H}) \mathbb{P}(B | \mathcal{H}) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(A \cap B | \mathcal{H}) \quad \text{quasi segurament.}$$

En la darrera igualtat hem utilitzat que $A \cap B \perp\!\!\!\perp \mathcal{H}$, ja que tant A com B són independents de $\mathcal{H}$. Concloem que $\mathcal{A}$ i $\mathcal{B}$ també són independents condicionadament a $\mathcal{H}$. Tot i així, val a dir que acostumarem a treballar amb distribucions discretes.

Teorema B.52. Siguin $\mathcal{A}, \mathcal{B}$ i $\mathcal{H}$ sub- σ -àlgebres de $\mathcal{F}$. Aleshores,

$$(\mathcal{A} \perp\!\!\!\perp \mathcal{B}) \mid \mathcal{H} \iff \mathbb{P}(A \mid \mathcal{B} \vee \mathcal{H}) = \mathbb{P}(A \mid \mathcal{H}) \text{ per a tot } A \in \mathcal{A}.$$

$\mathcal{B} \vee \mathcal{H}$ es refereix a la σ -àlgebra més petita que conté a $\mathcal{B}$ i $\mathcal{H}$.

Teorema B.53. Siguin $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ sub- σ -àlgebres de $\mathcal{F}$. Aleshores es verifiquen les següents propietats:

1. $\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C} \iff \mathcal{B} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{C}$ (simetria);
2. $\mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{D}) \mid \mathcal{C} \implies \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C}$ (reducció);
3. $\mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C}) \mid \mathcal{D} \implies \mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid (\mathcal{C} \vee \mathcal{D})$ (unió feble);
4. $(\mathcal{A} \perp\!\!\!\perp \mathcal{C} \mid \mathcal{B}) \wedge (\mathcal{A} \perp\!\!\!\perp \mathcal{D} \mid \mathcal{B} \vee \mathcal{C}) \implies \mathcal{A} \perp\!\!\!\perp (\mathcal{C} \vee \mathcal{D}) \mid \mathcal{B}$ (contracció);

Si aquestes relacions provenen d'una distribució amb densitat estrictament positiva respecte d'una mesura dominant producte, també es verifica la propietat d'intersecció:

$$(\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{C} \vee \mathcal{D}) \wedge (\mathcal{A} \perp\!\!\!\perp \mathcal{C} \mid \mathcal{B} \vee \mathcal{D}) \implies \mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C}) \mid \mathcal{D}.$$

Corol·lari B.54. Siguin X, Y, Z, W variables aleatòries sobre Ω i $\phi : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ una funció mesurable. Aleshores es compleixen les propietats següents:

1. $X \perp\!\!\!\perp Y \mid Z \implies Y \perp\!\!\!\perp X \mid Z$ (simetria);
2. $(X \perp\!\!\!\perp Y \mid Z) \wedge (W = \phi(Y)) \implies X \perp\!\!\!\perp W \mid Z$ (reducció);
3. $X \perp\!\!\!\perp (Y, Z) \mid W \implies X \perp\!\!\!\perp Y \mid (Z, W)$ (unió feble);
4. $(X \perp\!\!\!\perp Z \mid Y) \wedge (X \perp\!\!\!\perp W \mid (Y, Z)) \implies X \perp\!\!\!\perp (Z, W) \mid Y$ (contracció);
5. Si la llei conjunta de (X, Y, Z) admet una densitat estrictament positiva respecte d'una mesura dominant producte, llavors $(X \perp\!\!\!\perp Y \mid Z) \wedge (X \perp\!\!\!\perp Z \mid Y) \implies X \perp\!\!\!\perp (Y, Z)$ (intersecció).

Definició B.55 (Model d'independència, [Lau96, Sec. 3.1, pp. 30–31]). Sigui $I = \{1, \dots, n\}$. Un model d'independència $\perp_\sigma$ sobre I és una relació ternària sobre subconjunts de I . El model d'independència és un *semigrafoide* si es compleix per a tots els subconjunts disjunts A, B, C, D :

- Si $A \perp_\sigma B \mid C$, aleshores $B \perp_\sigma A \mid C$ (simetria);
- Si $A \perp_\sigma (B \cup D) \mid C$, aleshores $A \perp_\sigma B \mid C$ i $A \perp_\sigma D \mid C$ (descomposició);
- Si $A \perp_\sigma (B \cup D) \mid C$, aleshores $A \perp_\sigma B \mid (C \cup D)$ (unió feble);
- Si $A \perp_\sigma B \mid C$ i $A \perp_\sigma D \mid (B \cup C)$, aleshores $A \perp_\sigma (B \cup D) \mid C$ (contracció).

El model d'independència és un *grafoide* si a més també satisfà:

- Si $A \perp_\sigma B \mid (C \cup D)$ i $A \perp_\sigma C \mid (B \cup D)$, aleshores $A \perp_\sigma (B \cup C) \mid D$ (intersecció).

Finalment, el grafoide és *composicional* si també:

- Si $A \perp_\sigma B \mid C$ i $A \perp_\sigma D \mid C$, aleshores $A \perp_\sigma (B \cup D) \mid C$ (composició).

Nota B.56. Pel que fa al corol·lari anterior, apartat 5., a [Lau96, Sec. 3.1, Prop. 3.1] es demostra que si $p > 0$, llavors $\perp$ és un grafoide, però aquesta no és una condició necessària. En aquest sentit, [PJS17,

App. A.1] recorda que l'existència d'una densitat estrictament positiva és una condició suficient perquè valgui la propietat d'intersecció, però no necessària. No l'assumim globalment.

Definició B.57 (Models d'independència probabilística, [Lau96, Sec. 3.1, pp. 28–31]). Si $\mathcal{V} = \{V_i\}_{i \in I}$ té llei conjunta P , podem definir un model d'independència $\perp\!\!\!\perp_P$ com $A \perp\!\!\!\perp_P B \mid C \iff \mathbf{A} \perp\!\!\!\perp_P \mathbf{B} \mid \mathbf{C}$, on $\mathbf{A} = \{V_i \in \mathcal{V} \mid i \in A\}$, $\mathbf{B} = \{V_i \in \mathcal{V} \mid i \in B\}$ i $\mathbf{C} = \{V_i \in \mathcal{V} \mid i \in C\}$. El subíndex P indica la llei sota la qual es verifica la independència; reservem $\mathbb{P}$ per a la probabilitat de l'espai mostral subjacent.

Notació B.58. Si $\mathcal{V} = \{V_i\}_{i \in I}$ i $A, B, C \subseteq I$, la relació sobre índexs $A \perp\!\!\!\perp_P B \mid C$ s'ha d'entendre com la relació probabilística entre els subconjunts corresponents de variables, $\mathbf{A} \perp\!\!\!\perp_P \mathbf{B} \mid \mathbf{C}$. Quan calgui parlar d'arestes del graf, les escriurem sempre entre variables, per exemple $V_i \rightleftharpoons V_j$ o $V_i \rightarrow V_j$; les expressions indexades com $i \rightleftharpoons j$ només són una abreviació auxiliar després d'haver fixat explícitament la correspondència $i \mapsto V_i$.

B.4

ESPERANCES I MOMENTS

Aquesta secció recull les propietats bàsiques de l'esperança, la variància i la covariància. L'objectiu és disposar del vocabulari necessari per parlar d'efectes mitjans, estimadors i comparacions entre resultats potencials.

Proposició B.59. Siguin X, Y variables aleatòries reals simples i $a, b \in \mathbb{R}$. Aleshores:

1. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$,
2. Si $X, Y \geq 0$ i $X \leq Y$, es té $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Definició B.60 (Esperança matemàtica [Vil24b, Sec. 3, Def. 3.1.6]). Donada una variable aleatòria real X qualsevol, direm que es té esperança matemàtica finita si, i només si, $\mathbb{E}(|X|) < \infty$ i, en aquest cas, definim $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$.

Definició B.61 (Esperança condicionada [Dur19, Sec. 4.1]). Sigui $Y \in \mathcal{L}^1(\Omega)$ i sigui $\mathcal{G} \subseteq \mathcal{F}$ una sub- σ -àlgebra. L'esperança condicionada de Y donada $\mathcal{G}$, denotada per $\mathbb{E}[Y \mid \mathcal{G}]$, és qualsevol variable aleatòria $\mathcal{G}$ -mesurable i integrable tal que, per a tot $A \in \mathcal{G}$,

$$\int_A \mathbb{E}[Y \mid \mathcal{G}] d\mathbb{P} = \int_A Y d\mathbb{P}.$$

Si X és una variable aleatòria o un vector aleatori, escriurem $\mathbb{E}[Y \mid X]$ en lloc de $\mathbb{E}[Y \mid \sigma(X)]$. Aquesta variable és única quasi segurament; per això, les igualtats entre esperances condicionades s'entenen en aquest sentit.

Proposició B.62 (Esperances iterades [Dur19, Sec. 4.1, Thm. 4.1.13]). Sigui $Y \in \mathcal{L}^1(\Omega)$ i siguin $\mathcal{G} \subseteq \mathcal{H} \subseteq \mathcal{F}$ dues sub- σ -àlgebres. Aleshores,

$$\mathbb{E}[\mathbb{E}[Y \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[Y \mid \mathcal{G}] \quad \text{quasi segurament.}$$

En particular, si X i Z són variables o vectors aleatoris, llavors

$$\mathbb{E}[Y | X] = \mathbb{E}[\mathbb{E}[Y | X, Z] | X] \quad \text{quasi segurament.}$$

Demostració, cas general. Per definició d'esperança condicionada, per a tot $A \in \mathcal{G}$,

$$\int_A \mathbb{E}[\mathbb{E}[Y | \mathcal{H}] | \mathcal{G}] d\mathbb{P} = \int_A \mathbb{E}[Y | \mathcal{H}] d\mathbb{P}.$$

Com que $\mathcal{G} \subseteq \mathcal{H}$, també $A \in \mathcal{H}$, i per tant

$$\int_A \mathbb{E}[Y | \mathcal{H}] d\mathbb{P} = \int_A Y d\mathbb{P}.$$

Així, $\mathbb{E}[\mathbb{E}[Y | \mathcal{H}] | \mathcal{G}]$ satisfà la propietat característica de $\mathbb{E}[Y | \mathcal{G}]$. La igualtat se segueix de la unicitat quasi segura de l'esperança condicionada. ■

Demostració, cas discret. Suposem, per simplicitat, que X , Z i Y prenen un nombre finit de valors; el mateix càlcul s'estén al cas numerable quan Y és integrable. Fixem un valor x tal que $\mathbb{P}(X = x) > 0$. Aleshores, amb la convenció que els termes amb probabilitat condicional nul·la no contribueixen a la suma,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y | X, Z] | X = x] &= \sum_z \mathbb{E}[Y | X = x, Z = z] \mathbb{P}(Z = z | X = x) \\ &= \sum_z \sum_y y \mathbb{P}(Y = y | X = x, Z = z) \mathbb{P}(Z = z | X = x) \\ &= \sum_y y \sum_z \mathbb{P}(Y = y | X = x, Z = z) \mathbb{P}(Z = z | X = x) \\ &= \sum_y y \mathbb{P}(Y = y | X = x) = \mathbb{E}[Y | X = x]. \end{aligned}$$

La penúltima igualtat és la fórmula de la probabilitat total aplicada condicionadament a l'esdeveniment $\{X = x\}$. Com que la igualtat val per a tot x amb $\mathbb{P}(X = x) > 0$, obtenim

$$\mathbb{E}[\mathbb{E}[Y | X, Z] | X] = \mathbb{E}[Y | X] \quad \text{quasi segurament.} \quad \blacksquare$$

Proposició B.63. *Siguin $X, Y \in \mathcal{L}^1(\Omega)$ variables aleatòries reals, $a, b \in \mathbb{R}$.*

1. *La variable aleatòria $aX + bY \in \mathcal{L}^1(\Omega)$ i $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.*
2. *Si $X \leq Y$, aleshores $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*
3. *Propietat de convexitat. Es compleix que $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

Proposició B.64 (Esperança d'una variable absolutament contínua). *Una variable aleatòria real X amb densitat f pertany a $\mathcal{L}^1(\Omega)$ si, i només si, $\int_{-\infty}^{+\infty} |x| \cdot f(x) dx < +\infty$. En aquest cas, $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$.*

Procés B.65 (Càlcul de moments [Vil24b, Sec. 3, Defs. 3.2.1 i 3.2.4]). *Si existeix $\mu_1 = \mathbb{E}(X)$, els altres moments es calculen tenint en compte la definició d'esperança d'una funció d'una variable aleatòria; així, existeix el*

moment no centrat d'ordre k , donat per:

$$m_k = \begin{cases} \sum_{x \in E_X} x^k \mathbb{P}(X = x), & \text{si } X \text{ és discreta.} \\ \int_{\mathbb{R}} x^k f_X(x) dx, & \text{si } X \text{ és contínua.} \end{cases}$$

Proposició B.66 (Desigualtat de Jensen). Sigui g una funció real convexa i X una variable aleatòria real de $\mathcal{L}^1(\Omega)$ tal que $g(X) \in \mathcal{L}^1(\Omega)$. Aleshores, $g(\mathbb{E}(X)) \leq \mathbb{E}(g(X))$.

Proposició B.67. Siguin X, Y dues variables aleatòries reals independents amb esperança finita. Aleshores, la variable aleatòria producte XY té també esperança finita i $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proposició B.68. Siguin X, Y dues variables aleatòries reals independents amb moments de segon ordre finits. Aleshores, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Definició B.69 (Covariància). Per mesurar el grau de dependència entre dues variables aleatòries reals X i Y s'introdueix la noció de covariància. Suposem que les variables tenen moments de segon ordre finits. La covariància entre X i Y es defineix per

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

B.5

MODELS ESTADÍSTICS I ESTIMACIÓ PUNTUAL

Aquesta secció recull només el llenguatge estadístic que es necessita per passar dels estimands causals a quantitats calculables amb dades observades. L'objectiu és fixar què entenem per model paramètric, estadístic i estimador.

En aquesta secció, Ω denota l'espai mostral estadístic o espai d'observacions. No s'ha de confondre amb l'espai mostral subjacent d'un espai de probabilitat com el de la definició B.31: formalment, les observacions poden venir d'un vector aleatori $X : (\tilde{\Omega}, \tilde{\mathcal{F}}) \rightarrow (\Omega, \mathcal{A})$, i cada P_θ és la llei induïda per X sobre $(\Omega, \mathcal{A})$.

Definició B.70 (Model estadístic [Vil24a, Sec. 1.1, Def. 1.1.1]). És una terna $(\Omega, \mathcal{A}, \mathcal{P})$ en què Ω és l'espai d'observacions, $\mathcal{A}$ és una σ -àlgebra sobre Ω i $\mathcal{P}$ és una família de probabilitats sobre l'espai mesurable $(\Omega, \mathcal{A})$ [JM11, Cap. 4, Def. 4.1.1].

Definició B.71 (Model paramètric [Vil24a, Sec. 1.1, Def. 1.1.2]). Un model estadístic $(\Omega, \mathcal{A}, \mathcal{P})$ és **paramètric** si podem escriure la família de probabilitats com

$$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\},$$

on $\Theta \subset \mathbb{R}^d$, $d \geq 1$ [JM11, Cap. 4, Def. 4.1.2]. El conjunt Θ s'anomena espai de paràmetres.

Considerem un model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ i un vector aleatori $X = (X_1, \dots, X_n) : \tilde{\Omega} \rightarrow \Omega$, on $\Omega \subset \mathbb{R}^n$ és l'espai de les observacions. Recordem que P_θ és la llei de X en $(\Omega, \mathcal{A})$.

Definició B.72 (Funció de versemblança [Vil24a, Sec. 1.2, Def. 1.2.1]). Suposem que el model és dominat per una mesura μ sobre $(\Omega, \mathcal{A})$ en el sentit de la definició B.10. Pel teorema de Radon–Nikodym B.11, per a cada $\theta \in \Theta$ existeix una densitat

$$\mathcal{L}(\cdot, \theta) = \frac{dP_\theta}{d\mu}$$

definida μ -gairebé pertot [JM11, Cap. 4, Def. 4.3.1]. La **funció de versemblança** és l'aplicació

$$\mathcal{L} : \Omega \times \Theta \longrightarrow \mathbb{R}^+, \quad (x, \theta) \longmapsto \mathcal{L}(x, \theta),$$

considerada com a funció de θ quan l'observació x és fixada.

Observació B.73. La definició anterior recupera els dos casos elementals habituals:

1. Si Ω és numerable i μ és la mesura comptadora, $\mathcal{L}(x, \theta) = P_\theta(\{x\})$.
2. Si $\Omega \subset \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\Omega)$ i el model és dominat per la mesura de Lebesgue, $\mathcal{L}(x, \theta) = f_\theta(x)$, on f_θ és una densitat de P_θ .

Sovint treballarem amb mostres aleatòries simples de mida n ; és a dir, amb variables aleatòries $X_1, \dots, X_n$ independents i idènticament distribuïdes. Si la llei marginal és dominada per una mesura ν i té densitat $f_\theta = dP_\theta^{(1)}/d\nu$, la llei conjunta és dominada per $\nu^{\otimes n}$ i, per a $x = (x_1, \dots, x_n)$, la versemblança és

$$\mathcal{L}(x, \theta) = \prod_{i=1}^n f_\theta(x_i).$$

En particular, en el cas discret $f_\theta(x_i) = P_\theta^{(1)}(\{x_i\})$, mentre que en el cas absolutament continu respecte de Lebesgue f_θ és la densitat usual.

Definició B.74 (Model exponencial [Vil24a, Sec. 1.2, Def. 1.2.4]). Direm que un model estadístic paramètric dominat $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ és exponencial si existeixen funcions reals mesurables $\Psi, \Phi_1, \dots, \Phi_r : \Omega \longrightarrow \mathbb{R}$ i funcions reals $D, C_1, \dots, C_r : \Theta \longrightarrow \mathbb{R}$ tals que podem expressar la funció de versemblança com [JM11, Cap. 4, Sec. 4.4]:

$$\mathcal{L}(x, \theta) = \exp \left\{ \sum_{k=1}^r C_k(\theta) \Phi_k(x) + D(\theta) + \Psi(x) \right\}.$$

Equivalentment, podem escriure

$$\mathcal{L}(x, \theta) = \tilde{D}(\theta) \tilde{\Psi}(x) \exp \left\{ \sum_{k=1}^r C_k(\theta) \Phi_k(x) \right\},$$

amb $\tilde{D}(\theta), \tilde{\Psi}(x) > 0$. Les funcions que apareixen en aquesta representació no són úniques. En el cas particular que l'espai de paràmetres Θ sigui unidimensional, la versemblança es pot escriure com

$$\mathcal{L}(x, \theta) = \exp \{C(\theta)\Phi(x) + D(\theta) + \Psi(x)\}.$$

Definició B.75 (Estadístic [Vil24a, Sec. 2.1]). Considerem el model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$. Un **estadístic** és una aplicació mesurable [JM11, Cap. 5, Sec. 5.1]:

$$T : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E}),$$

on $(E, \mathcal{E})$ és un espai mesurable. La mitjana mostral, la variància mostral, el màxim i el mínim són exemples habituals d'estadístics reals.

Definició B.76 (Estadístic suficient [Vil24a, Sec. 1.3, Def. 1.3.3]). Un estadístic $T : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ és **suficient** per al model si la distribució condicional de l'observació X donat $T(X)$ no depèn de θ [JM11, Cap. 5, Def. 5.1.2]. Intuïtivament, la mostra completa no aporta informació addicional sobre θ un cop conegut el valor de T .

Teorema B.77 (Factorització de Neyman–Fisher [Vil24a, Sec. 1.3, Thm. 1.3.4]). Sigui $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ un model dominat per μ , amb versemblança $\mathcal{L} = dP_\theta/d\mu$, i sigui $T : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ un estadístic. Aleshores, T és suficient si, i només si, existeixen funcions mesurables no negatives $\psi : E \times \Theta \longrightarrow \mathbb{R}^+$ i $h : \Omega \longrightarrow \mathbb{R}^+$ tals que

$$\mathcal{L}(x, \theta) = \psi(T(x), \theta)h(x),$$

per a tot $\theta \in \Theta$ i per a μ -gairebé tot $x \in \Omega$ [JM11, Cap. 5, Thm. 5.1.3].

Definició B.78 (Model regular [Vil24a, Sec. 1.3, Def. 1.3.6]). Un model paramètric dominat unidimensional és **regular** si es compleixen les condicions següents [JM11, Cap. 5, Def. 5.2.1]:

1. El suport $\{x \in \Omega \mid \mathcal{L}(x, \theta) > 0\}$ no depèn de θ .
2. Per a cada x del suport, existeixen les derivades $\partial_\theta \log \mathcal{L}(x, \theta)$ i $\partial_\theta^2 \log \mathcal{L}(x, \theta)$.
3. Les operacions de derivació respecte de θ i d'integració respecte de la mesura dominant es poden intercanviar localment, per exemple mitjançant una funció integrable que domini les derivades corresponents.
4. La informació de Fisher és finita i estrictament positiva per a tot $\theta \in \Theta$.

Definició B.79 (Informació de Fisher [Vil24a, Sec. 1.4, Def. 1.4.1]). En un model regular unidimensional, la **informació de Fisher** és [JM11, Cap. 5, Def. 5.2.3]:

$$I_F(\theta) = \mathbb{E}_\theta \left[(\partial_\theta \log \mathcal{L}(X, \theta))^2 \right].$$

Considerem ara un model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^d$, on el paràmetre θ és desconegut, i una funció paramètrica $g : \Theta \longrightarrow \mathbb{R}^k$.

Definició B.80 (Estimador [Vil24a, Sec. 2.1, Def. 2.1.1]). Un **estimador** de $g(\theta)$ és un estadístic [JM11, Cap. 7]:

$$T : \Omega \longrightarrow g(\Theta)$$

que s'utilitza per aproximar el valor desconegut de $g(\theta)$ a partir de les observacions.

Definició B.81 (Funció de pèrdua [Vil24a, Sec. 2.1]). Considerem una funció mesurable [JM11, Cap. 6]:

$$\begin{aligned} W : \mathbb{R}^k \times \Theta &\longrightarrow \mathbb{R}^+ \\ (t, \theta) &\longmapsto W(t, \theta) \end{aligned}$$

que representa l'error que cometem en proposar t com a valor estimat de $g(\theta)$. Aquesta funció és una **funció de pèrdua** i ha de complir $W(g(\theta), \theta) = 0$.

Definició B.82 (Funció de risc [Vil24a, Sec. 2.1, Def. 2.1.3]). Donats una funció de pèrdua W i un estimador T , la **funció de risc** associada a T és la funció que amitjana les pèrdues [JM11, Cap. 6]:

$$\begin{aligned} R_T : \Theta &\longrightarrow \mathbb{R} \\ \theta &\longrightarrow \mathbb{E}_\theta[W(T, \theta)]. \end{aligned}$$

Aquesta funció està definida només per aquells $\theta \in \Theta$ tals que $R_T(\theta) < \infty$.

Definició B.83 (Estadístic integrable [Vil24a, Sec. 2.1, Def. 2.1.6]). Un estadístic T és integrable si té esperança finita per a tot θ ; és a dir, $\mathbb{E}_\theta(|T|) < \infty$, $\forall \theta \in \Theta$.

B.6

BIAIX, VARIÀNCIA I MITJANES MOSTRALS

Aquesta secció fixa les propietats bàsiques dels estimadors que s'utilitzen al capítol d'estudis: absència de biaix, variància i comparació de mitjanes observades.

Definició B.84 (Biaix i absència de biaix [Vil24a, Defs. 2.1.7–2.1.8]). Si T és un estimador integrable de $g(\theta)$, el seu **biaix** és [JM11, Cap. 7, Defs. 7.1.2–7.1.3]:

$$b_T(\theta) = \mathbb{E}_\theta(T) - g(\theta).$$

Direm que T és **sense biaix** si $b_T(\theta) = 0$ per a tot $\theta \in \Theta$, és a dir, si $\mathbb{E}_\theta(T) = g(\theta)$.

Definició B.85 (Variància i risc quadràtic [Vil24a, Defs. 2.1.3 i 2.1.8]). Si T és de quadrat integrable, la seva variància respecte al model és [JM11, Cap. 7, Sec. 7.1]:

$$\text{Var}_\theta(T) = \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right].$$

Per a la pèrdua quadràtica $W(t, \theta) = (t - g(\theta))^2$, el risc és

$$R_T(\theta) = \mathbb{E}_\theta \left[(T - g(\theta))^2 \right] = \text{Var}_\theta(T) + b_T(\theta)^2.$$

En particular, quan T és sense biaix per a $g(\theta)$, el risc quadràtic coincideix amb $\text{Var}_\theta(T)$.

Corol·lari B.86 (Mitjana mostral [Vil24a, Sec. 2.1, Cor. 2.1.13]). Si $X_1, \dots, X_n$ són variables aleatòries reals

idènticament distribuïdes i integrables sota P_θ , la mitjana mostral [JM11, Cap. 7, Cor. 7.1.8]:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

és un estimador sense biaix de la mitjana poblacional: $\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}_\theta(X_1)$.

Corol·lari B.87 (Variància mostral corregida [Vil24a, Prop. 2.1.14 i Cor. 2.1.15]). Si $X_1, \dots, X_n$ són idènticament distribuïdes, independents i tenen moment de segon ordre sota P_θ , escrivim [JM11, Cap. 7, Cor. 7.1.10]:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad \widehat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Aleshores S_n^2 és esbiaixat per a la variància poblacional, amb

$$\mathbb{E}_\theta(S_n^2) = \frac{n-1}{n} \text{Var}_\theta(X_1),$$

i la variància mostral corregida $\widehat{S}_n^2$ no té biaix:

$$\mathbb{E}_\theta(\widehat{S}_n^2) = \text{Var}_\theta(X_1).$$

Definició B.88 (UMV [Vil24a, Sec. 2.2, Def. 2.2.1]). Per a una funció real $g : \Theta \rightarrow \mathbb{R}$, denotem per

$$\mathcal{E}_g = \{T \mid T \text{ és un estimador integrable i sense biaix de } g(\theta)\}$$

la classe d'estimadors sense biaix de $g(\theta)$. Un estimador $T^* \in \mathcal{E}_g$ és **uniformement de mínima variància** (UMV) si, per a tot $T \in \mathcal{E}_g$ i per a tot $\theta \in \Theta$ [JM11, Cap. 7, Def. 7.2.1]:

$$\text{Var}_\theta(T^*) \leq \text{Var}_\theta(T).$$

Dues preguntes naturals que hom es podria plantejar són les següents:

1. *Aquests UMV són únics?* La resposta és que sí.
2. *Existència dels UMV (procés de construcció)?* La resposta a aquesta pregunta es dona en dos teoremes que queden fora de l'abast d'aquesta secció: Rao–Blackwell i Lehmann–Scheffé.

Definició B.89 (Estadístic complet [Vil24a, Sec. 2.2, Def. 2.2.2]). Direm que un estadístic $S : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ és **complet** si per a tota funció mesurable $f : E \rightarrow \mathbb{R}$ tal que $f(S)$ és integrable es compleix [JM11, Cap. 7, Sec. 7.2]:

$$\mathbb{E}_\theta(f(S)) = 0, \forall \theta \in \Theta \implies P_\theta(f(S) = 0) = 1, \forall \theta \in \Theta.$$

CONFIANÇA I TESTOS

Aquesta secció passa de l'estimació d'un valor concret a la quantificació de la incertesa i la presa de decisions. L'objectiu és introduir regions de confiança, funcions pivotants i tests d'hipòtesis, fins arribar al criteri de Neyman-Pearson per controlar errors.

L'estimació puntual ens proporciona un estimador T que interpretem com una aproximació al valor de la funció $g(\theta)$. Ara desenvoluparem un mètode que consisteix a trobar un subconjunt de $g(\Theta)$ depenent de les observacions i que contingui el valor del paràmetre amb una probabilitat fixada a priori. Considerem un model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^d$, amb paràmetre θ desconegut, i una funció $g : \Theta \rightarrow \mathbb{R}^k$. Fixem un nivell de confiança $\gamma \in (0, 1)$.

Definició B.90 (Regió de confiança [Vil24a, Sec. 3.1, Def. 3.1.1]). Una **regió de confiança** per a $g(\theta)$ al nivell de confiança γ és una família de parts de $g(\Theta)$, $\{S(x) \subset g(\Theta) \mid x \in \Omega\}$, que compleix les dues propietats següents [JM11, Cap. 9, Def. 9.1.1]:

1. Per a tot $\theta \in \Theta$, el conjunt $\{x \in \Omega \mid g(\theta) \in S(x)\}$ és mesurable.
2. Per a tot $\theta \in \Theta$,

$$P_\theta(g(\theta) \in S(X)) \geq \gamma.$$

Si repetíssim moltes vegades l'experiència aleatòria, la regió $S(X)$ construïda a partir de les observacions contindria el veritable valor de $g(\theta)$ aproximadament en almenys un $100\gamma\%$ dels casos. Quan g pren valors reals ($k = 1$) i les regions $S(x)$ són intervals, les regions de confiança s'anomenen **intervals de confiança**.

Definició B.91 (Funció pivotant [Vil24a, Sec. 3.1, Def. 3.1.2]). Una funció $\pi : \Omega \times g(\Theta) \rightarrow \mathbb{R}^k$ s'anomena *funció pivotant* per a $g(\theta) \in g(\Theta)$ si compleix [JM11, Cap. 9, Def. 9.2.1]:

1. Per a tot $\theta \in \Theta$, $\pi(\cdot, g(\theta))$ és mesurable.
2. La llei de $\pi(\cdot, g(\theta))$ no depèn de θ .

Si la mida de la mostra és prou gran, podem buscar un interval aproximat mitjançant el comportament asimptòtic d'algun estadístic associat al model. Sigui $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}$, i sigui $(T_n)_n$ una successió d'estadístics. Suposem:

1. $\text{Var}_\theta(T_n) < \infty$,
2. $\mathbb{E}_\theta(T_n) = g(\theta)$,
3. i el comportament asimptòtic següent:

$$\frac{T_n - g(\theta)}{\sqrt{\text{Var}_\theta(T_n)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

Sigui $\eta_\gamma > 0$ tal que, per a $Z \sim N(0, 1)$, $P(-\eta_\gamma \leq Z \leq \eta_\gamma) = \gamma$. Per a n gran,

$$P_\theta \left(-\eta_\gamma \leq \frac{T_n - g(\theta)}{\sqrt{\text{Var}_\theta(T_n)}} \leq \eta_\gamma \right) \approx \gamma,$$

i, per tant, obtenim l'interval asimptòtic aproximat

$$\left[T_n - \eta_\gamma \sqrt{\text{Var}_\theta(T_n)}, T_n + \eta_\gamma \sqrt{\text{Var}_\theta(T_n)} \right].$$

Si $\text{Var}_\theta(T_n)$ depèn de paràmetres desconeguts, en aplicacions se substitueix per un estimador consistent.

Considerem un model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$. Qüestionar-se alguna cosa del paràmetre és equivalent a fer una partició de l'espai de paràmetres $\Theta = \Theta_0 \cup \Theta_1$, amb $\Theta_0 \cap \Theta_1 = \emptyset$, i demanar-se si θ pertany o no a Θ_0 .

Definició B.92 (Hipòtesis [Vil24a, Sec. 4.1, Defs. 4.1.2–4.1.4]). La **hipòtesi nul·la** és $H_0 : \theta \in \Theta_0$, i la **hipòtesi alternativa** és $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$ [JM11, Cap. 10, Sec. 10.1]. Quan un dels dos conjunts, Θ_0 o Θ_1 , consta d'un sol element, direm que la hipòtesi corresponent és simple; altrament, direm que és composta.

Definició B.93 (Test, potència i errors [Vil24a, Sec. 4.1, Defs. 4.1.9–4.1.11]). Un **test** és una funció mesurable $\varphi : \Omega \rightarrow [0, 1]$, on $\varphi(x)$ s'interpreta com la probabilitat de rebutjar H_0 quan s'observa x [JM11, Cap. 10, Sec. 10.1]. Si $\varphi = \mathbb{1}_{A_1}$ per a una regió crítica $A_1 \in \mathcal{A}$, el test és no aleatoritzat i $A_0 = \Omega \setminus A_1$ és la regió d'acceptació.

La **funció de potència** del test és

$$\Phi_\varphi(\theta) = \mathbb{E}_\theta[\varphi(X)].$$

Per a $\theta \in \Theta_0$, l'error de primera espècie és $\alpha_1(\theta) = \Phi_\varphi(\theta)$; per a $\theta \in \Theta_1$, l'error de segona espècie és $\alpha_2(\theta) = 1 - \Phi_\varphi(\theta)$. El test té **nivell de significació** $\alpha \in (0, 1)$ si

$$\sup_{\theta \in \Theta_0} \Phi_\varphi(\theta) \leq \alpha.$$

La restricció de Φ_φ a Θ_1 s'anomena potència contra l'alternativa.

	Acceptar H_0	Rebutjar H_0
H_0 certa	Decisió correcta	Error de primera espècie
H_0 falsa	Error de segona espècie	Decisió correcta

Observació B.94 (Intervals de confiança i tests [Vil24a, Sec. 4.1, Obs. 4.1.12]). Considerem el model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ i sigui $\theta_0 \in \Theta$. A partir de les regions de confiança podem construir tests d'hipòtesis a un nivell de significació α donat per contrastar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. En efecte, suposem que $[T_1, T_2]$ és un interval de confiança al nivell de confiança γ per al paràmetre θ és a dir,

$$P_\theta(\theta \in [T_1(X), T_2(X)]) \geq \gamma, \quad \forall \theta \in \Theta.$$

Aleshores, $A_0 = \{x \in \Omega \mid T_1(x) \leq \theta_0 \leq T_2(x)\}$ és la regió d'acceptació d'un test $\varphi = \mathbb{1}_{A_0^c}$ al nivell de

significació $\alpha = 1 - \gamma$, ja que $P_{\theta_0}(A_0^c) \leq 1 - \gamma = \alpha$ [JM11, Cap. 10, Obs. 10.1.3]. Aquest mètode té l'inconvenient de no controlar l'error de segona espècie, o el que és el mateix, la potència del test.

Hem de fixar una estratègia per determinar un test que controli els dos errors. Dels dos errors, el de primera espècie serà el que tindrà pitjors conseqüències i, entre les decisions possibles, considerarem únicament les que tinguin aquest error afitat per una quantitat fixada a priori; després, dins d'aquesta classe de tests amb error de primera espècie afitat, provarem de trobar el que minimitzi l'error de segona espècie i maximitzi, doncs, la potència del test.

Definició B.95 (UMP [Vil24a, Sec. 4.1, Def. 4.1.13]). Designarem per $\mathcal{D}_\alpha$ el conjunt de tots els tests de nivell de significació α . Donats dos tests $\varphi, \varphi' \in \mathcal{D}_\alpha$, direm que φ és més potent que φ' si [JM11, Cap. 10, Sec. 10.1]:

$$\Phi_{\varphi'}(\theta) \leq \Phi_{\varphi}(\theta), \quad \forall \theta \in \Theta_1.$$

Fixat un nivell de significació α , direm que un test de la classe $\mathcal{D}_\alpha$ és **uniformement de màxima potència** (UMP) si és més potent que qualsevol altre test de $\mathcal{D}_\alpha$.

Observació B.96 (Existència i asimetria dels tests UMP [Vil24a, Sec. 4.1, Obs. 4.1.14]). 1. En general pot ocórrer que no existeixi un test UMP o, en cas de no existir, que no sigui únic. Més endavant donarem resultats en aquesta direcció.

2. En aquesta estratègia les dues hipòtesis no són tractades de la mateixa manera i, per tant, el tractament no és pas simètric. Les dues hipòtesis no són intercanviables.

Considerem un model estadístic paramètric dominat $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ amb funció de versemblança associada $\mathcal{L}(x, \theta)$.

Lema B.97 (Neyman–Pearson [Vil24a, Sec. 4.1, Lema 4.1.15]). Fixat $\alpha \in (0, 1)$, considerem el contrast d'hipòtesis simples $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$. Suposem que existeix una constant positiva c_α tal que [JM11, Cap. 10, Sec. 10.2, Thm. 10.2.1]:

$$A_0 = \{x \in \Omega \mid \mathcal{L}(x, \theta_1) \leq c_\alpha \mathcal{L}(x, \theta_0)\}$$

satisfà $P_{\theta_0}(A_0^c) = \alpha$. Aleshores, el test $\varphi(x) = \mathbb{1}_{A_0^c}(x)$ té nivell de significació α i és de màxima potència entre tots els tests de nivell α per contrastar H_0 contra H_1 . En aquest contrast simple, és un test UMP. Si $\mathcal{L}(x, \theta_0) > 0$ sobre el suport comú del model, aleshores la regió d'acceptació es pot escriure com

$$A_0 = \left\{ x \in \Omega \mid \frac{\mathcal{L}(x, \theta_1)}{\mathcal{L}(x, \theta_0)} \leq c_\alpha \right\}.$$

TAULES DE CONTINGÈNCIA I MESURES D'ASSOCIACIÓ

Aquesta secció recull les eines estadístiques bàsiques per analitzar variables categòriques: taules de contingència, contrast de Pearson d'independència, odds ratio i resum estratificat de Mantel-Haenszel.

La notació d'aquesta secció és la que farem servir en el cas pràctic de Berkeley: O_{ij} per als recomptes observats, E_{ij} per als esperats sota independència, i a, b, c, d per a les cel·les d'una taula 2×2 . Les fonts citades usen notacions diferents; per exemple, Julia de Ferran i Márquez-Carreras [JM11, Cap. 12, Sec. 12.3] escriu els recomptes com n_{kj} i els marges com $n_{k\cdot}, n_{\cdot j}$, mentre que Vilar [Vil24a, Sec. 4.3, Defs. 4.3.1 i 4.3.5] presenta el contrast de Pearson en forma multinomial. Per tant, les cites d'aquesta secció s'han d'entendre com a referències al test o resultat estadístic subjacent, no com una adopció literal de la mateixa notació.

Siguin X i Y dues variables categòriques amb valors possibles $x_1, \dots, x_r$ i $y_1, \dots, y_s$. A partir d'una mostra de mida n , escriurem O_{ij} per al nombre d'observacions amb $X = x_i$ i $Y = y_j$. Els totals marginals seran

$$O_{i\cdot} = \sum_{j=1}^s O_{ij}, \quad O_{\cdot j} = \sum_{i=1}^r O_{ij}, \quad n = \sum_{i,j} O_{ij}.$$

La matriu $O = (O_{ij})$ és la **taula de contingència** observada. Sota la hipòtesi nul·la d'independència entre X i Y , els recomptes esperats s'estimen amb els marges:

$$E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}. \quad (\text{B.1})$$

Definició B.98 (Contrast de Pearson d'independència). Amb la notació anterior, el contrast χ^2 de Pearson d'independència en una taula $r \times s$ utilitza l'estadístic

$$X^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

on els E_{ij} són els recomptes esperats de (B.1). Aquesta és la forma del test d'independència de Pearson per a taules de contingència [JM11, Cap. 12, Sec. 12.3]; també es pot veure com una especialització del contrast χ^2 de Pearson per a models multinomials [Vil24a, Sec. 4.3, Defs. 4.3.1 i 4.3.5]. Sota la hipòtesi nul·la i les condicions asimptòtiques habituals, X^2 convergeix en llei a una $\chi^2_{(r-1)(s-1)}$. Com a criteri pràctic d'ús de l'aproximació, demanarem que els recomptes esperats no siguin inferiors a 5, en paral·lel amb la regla $np_{0j} \geq 5$ del contrast χ^2 multinomial [Vil24a, Obs. 4.3.3].

El contrast anterior detecta associació estadística entre variables observades. Rebutjar independència no és una afirmació causal: només diu que la distribució conjunta observada no factoritza com el producte dels marginals.

Definició B.99 (Odds ratio). Per a una taula 2×2 amb recomptes positius,

	$Y = 1$	$Y = 0$
$X = 1$	a	b
$X = 0$	c	d

la raó d'avantatges o *odds ratio* és

$$\widehat{\text{OR}} = \frac{a/b}{c/d} = \frac{ad}{bc}.$$

El valor $\widehat{\text{OR}} = 1$ equival a $ad - bc = 0$; per tant, en una taula 2×2 , el signe de $\log \widehat{\text{OR}}$ coincideix amb el signe de $ad - bc$.

Definició B.100 (Estimador de Mantel–Haenszel [MH59]). Suposem que disposem de K taules 2×2 estratificades. Per a l'estrat k , escrivim

	$Y = 1$	$Y = 0$	
$X = 1$	a_k	b_k	$n_k = a_k + b_k + c_k + d_k.$
$X = 0$	c_k	d_k	

L'odds ratio de l'estrat és $\widehat{\text{OR}}_k = \frac{a_k d_k}{b_k c_k}$, i l'estimador de Mantel–Haenszel d'una odds ratio comuna és

$$\widehat{\text{OR}}_{MH} = \frac{\sum_{k=1}^K a_k d_k / n_k}{\sum_{k=1}^K b_k c_k / n_k}. \quad (\text{B.2})$$

Definició B.101 (Contrast de Cochran–Mantel–Haenszel [Coc54; MH59]). Amb la notació de la definició B.100, el contrast de Cochran–Mantel–Haenszel de no-associació comuna contrasta $H_0 : \text{OR}_k = 1$ per a tot estrat k . Condicionem pels marges de cada estrat i escrivim A_k per a la variable aleatòria que dona el recompte de la cel·la superior esquerra de l'estrat k ; en les dades observades, $A_k = a_k$. Sota H_0 tenim

$$\mathbb{E}_0(A_k) = \frac{(a_k + b_k)(a_k + c_k)}{n_k}, \quad \text{Var}_0(A_k) = \frac{(a_k + b_k)(c_k + d_k)(a_k + c_k)(b_k + d_k)}{n_k^2(n_k - 1)}.$$

L'estadístic

$$M^2 = \frac{\left[\sum_{k=1}^K (a_k - \mathbb{E}_0(A_k)) \right]^2}{\sum_{k=1}^K \text{Var}_0(A_k)}$$

és asimptòticament χ^2 amb un grau de llibertat.

Observació B.102. L'estimador (B.2) és un resum estratificat útil quan té sentit aproximar les taules per una odds ratio comuna. La hipòtesi d'homogeneïtat $\text{OR}_1 = \dots = \text{OR}_K$ de les odds ratios estratificades és una pregunta diferent de $H_0 : \text{OR}_k = 1$; habitualment es pot contrastar amb proves de tipus Breslow–Day [BD80].

B.9

DISTRIBUCIONS

Aquesta secció funciona com a taula de consulta de distribucions discretes i contínues habituals. L'objectiu és concentrar en un únic lloc les expressions de densitat, esperança i variància que poden aparèixer en exemples o càlculs auxiliars.

Distribució	PMF/PDF/CDF	Esperança	Variància	φ
Bernoulli(p) $p \in (0, 1)$	$\mathbb{P}(X = 1) = p$ $\mathbb{P}(X = 0) = q = 1 - p$	p	pq	$pe^{it} + q$
$B(n, p)$ $n \in \mathbb{N}, p \in (0, 1)$	$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$	np	npq	$(pe^{it} + q)^n$
Geom(p) $p \in (0, 1)$	$\mathbb{P}(X = k) = p \cdot q^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-q \cdot e^{it}}$
Poiss(λ) $\lambda \in (0, +\infty)$	$\mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$	λ	λ	$\exp\{\lambda(e^{it} - 1)\}$
$U(a, b)$ $a < b$	$f(x) = \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(x)$ $F(x) = \frac{x-a}{b-a} \cdot \mathbb{1}_{(a,b)}(x)$	$m_1 = \frac{a+b}{2}$ $m_n = \frac{\sum_{i=0}^n a^i b^{n-i}}{n+1}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Exp(λ) $\lambda \in (0, +\infty)$	$f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0,+\infty)}(x)$ $F(x) = (1 - e^{-\lambda x}) \cdot \mathbb{1}_{(0,+\infty)}(x)$	$m_1 = \frac{1}{\lambda}$ $m_n = \frac{n!}{\lambda^n}$	$\frac{1}{\lambda^2}$	$\left(1 - \frac{it}{\lambda}\right)^{-1}$
Gamma(α, β) $\alpha, \beta \in (0, +\infty)$	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)} \cdot x^{\beta-1} \cdot e^{-\alpha x} \cdot \mathbb{1}_{(0,+\infty)}(x)$	$\frac{\beta}{\alpha}$	$\frac{\beta}{\alpha^2}$	$\left(1 - \frac{it}{\alpha}\right)^{-\beta}$

Figura B.1: Distribucions discretes i absolutament contínues.

Teoria de grafs

C.1

GRAFS I SUBGRAFS

Aquesta secció introdueix el vocabulari elemental de teoria de grafs que es farà servir després en els DAGs causals. L'objectiu és fixar què són vèrtexs, arestes, subgrafs, camins, connexió i cicles.

GRAFS I GRAFS SIMPLES

Es fixa la definició abstracta de graf i la notació bàsica de vèrtexs, arestes i incidència. També es distingeixen els grafs simples, que seran el cas de referència abans de passar als DAGs.

Moltes situacions del món real es poden descriure de manera convenient mitjançant un diagrama format per un conjunt de punts i línies que uneixen certs parells d'aquests punts. Per exemple, els punts podrien representar persones, i les línies, relacions d'amistat; o els punts podrien ser centres de comunicació, i les línies, els enllaços de comunicació. En aquests diagrames, l'interès principal rau en si dos punts estan connectats o no, més que en la forma específica de la connexió. L'abstracció matemàtica d'aquest tipus de situacions dona lloc al concepte de graf.

Definició C.1 (Graf, [BM08]). Un **graf** $\mathcal{G}$ és un triplet ordenat $(\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}), \psi_{\mathcal{G}})$ format per:

1. Un conjunt no buit $\mathcal{V}(\mathcal{G})$ de **vèrtexs**.
2. Un conjunt $\mathcal{E}(\mathcal{G})$ d'**arestes**, disjunt de $\mathcal{V}(\mathcal{G})$.
3. Una **funció d'incidència** $\psi_{\mathcal{G}}$ que associa a cada aresta de $\mathcal{G}$ un parell no ordenat de vèrtexs de $\mathcal{G}$ (no necessàriament diferents).

Si e és una aresta i u, v són vèrtexs tals que $\psi_{\mathcal{G}}(e) = uv$, es diu que e **uneix** u i v . Els vèrtexs u i v s'anomenen els **extrems** de e .

Els grafs es denominen així perquè es poden representar gràficament, i aquesta representació ens ajuda a entendre moltes de les seves propietats. Cada vèrtex s'indica amb un punt i cada aresta amb una línia que uneix els punts que representen els seus extrems.

És important destacar que la representació gràfica d'un graf no és única; la posició relativa dels vèrtexs i la forma de les arestes no tenen cap significat. Un diagrama d'un graf simplement representa la relació d'incidència entre els seus vèrtexs i arestes.

Definició C.2 (Conceptes bàsics).

- Dos vèrtexs que són extrems d'una mateixa aresta s'anomenen **adjacents**.

- Dues arestes que comparteixen un vèrtex s'anomenen **adjacents**.
- Una aresta amb extrems idèntics s'anomena **llaç**.
- Una aresta amb extrems diferents s'anomena **enllaç**.

Definició C.3 (Graf simple i complet). Un graf és **simple** si no té llaços i no hi ha més d'una aresta que uneixi el mateix parell de vèrtexs. Un graf simple on cada parell de vèrtexs diferents està unit per una aresta s'anomena **graf complet**.

Nota C.4 (Notació). Utilitzarem els símbols $v(\mathcal{G})$ i $\varepsilon(\mathcal{G})$ per denotar el nombre de vèrtexs i arestes d'un graf $\mathcal{G}$, respectivament. En contextos on només es discuteix un graf, ometrem la $\mathcal{G}$ i escriurem simplement $\mathcal{V}, \mathcal{E}, v, \varepsilon$.

SUBGRAFS

S'introdueixen les maneres bàsiques de restringir un graf a part dels seus vèrtexs o arestes. La notació de subgrafs i eliminació serà necessària per descriure grafs mutilats.

Definició C.5 (Subgraf). Un graf $\mathcal{H}$ és un **subgraf** de $\mathcal{G}$ (escrit $\mathcal{H} \subseteq \mathcal{G}$) si $\mathcal{V}(\mathcal{H}) \subseteq \mathcal{V}(\mathcal{G})$, $\mathcal{E}(\mathcal{H}) \subseteq \mathcal{E}(\mathcal{G})$, i $\psi_{\mathcal{H}}$ és la restricció de $\psi_{\mathcal{G}}$ a $\mathcal{E}(\mathcal{H})$. Si $\mathcal{H} \subseteq \mathcal{G}$ però $\mathcal{H} \neq \mathcal{G}$, $\mathcal{H}$ és un **subgraf propi** de $\mathcal{G}$. Si $\mathcal{H}$ és un subgraf de $\mathcal{G}$, llavors $\mathcal{G}$ és un **supergraf** de $\mathcal{H}$. Un subgraf $\mathcal{H}$ és un **subgraf generador** (o *spanning subgraph*) de $\mathcal{G}$ si $\mathcal{V}(\mathcal{H}) = \mathcal{V}(\mathcal{G})$.

Definició C.6 (Graf simple subjacent). El **graf simple subjacent** d'un graf $\mathcal{G}$ és el graf simple obtingut eliminant tots els llaços i, per a cada parell de vèrtexs, totes les arestes menys una que els uneixi.

Definició C.7 (Subgraf induït).

- El **subgraf induït** per un subconjunt de vèrtexs $\mathcal{V}' \subseteq \mathcal{V}$, denotat per $\mathcal{G}[\mathcal{V}']$, és el graf que té $\mathcal{V}'$ com a conjunt de vèrtexs i el conjunt d'arestes format per totes les arestes de $\mathcal{G}$ que tenen tots dos extrems a $\mathcal{V}'$.
- El **subgraf induït** per un subconjunt d'arestes $\mathcal{E}' \subseteq \mathcal{E}$, denotat per $\mathcal{G}[\mathcal{E}']$, és el graf que té $\mathcal{E}'$ com a conjunt d'arestes i el conjunt de vèrtexs format per tots els extrems de les arestes a $\mathcal{E}'$.

Nota C.8 (Notació d'eliminació). L'operació d'eliminar vèrtexs o arestes es denota amb el signe menys.

- $\mathcal{G} - \mathcal{V}'$ és el subgraf $\mathcal{G}[\mathcal{V} \setminus \mathcal{V}']$. Si $\mathcal{V}' = \{v\}$, escrivim $\mathcal{G} - v$.
- $\mathcal{G} - \mathcal{E}'$ és el subgraf generador de $\mathcal{G}$ amb conjunt d'arestes $\mathcal{E} \setminus \mathcal{E}'$. Si $\mathcal{E}' = \{e\}$, escrivim $\mathcal{G} - e$.

CAMINS I CONNEXIÓ

Es defineixen passeigs, recorreguts i camins com a formes de moure's dins d'un graf. A partir d'això es fixa la noció de connexió i de component.

Per a analitzar l'estructura d'un graf, és fonamental entendre com es connecten els seus vèrtexs. Això ens porta a les nocions de passeig, recorregut i camí.

Definició C.9 (Passeig, recorregut i camí [BM08]).

- Un **passeig** (*walk*) en un graf $\mathcal{G}$ és una seqüència finita no nul·la $W = v_0 e_1 v_1 e_2 \dots e_k v_k$, els termes de la qual són alternativament vèrtexs i arestes, de manera que per a $1 \leq i \leq k$, els extrems de l'aresta e_i són v_{i-1} i v_i .
- Els vèrtexs v_0 i v_k s'anomenen **origen** i **termini** de W , respectivament. Els vèrtexs $v_1, \dots, v_{k-1}$ són els **vèrtexs interns**. L'enter k és la **longitud** del passeig.
- Si totes les arestes $e_1, \dots, e_k$ d'un passeig són diferents, el passeig s'anomena **recorregut** (*trail*).
- Si, a més, tots els vèrtexs $v_0, \dots, v_k$ són diferents, el recorregut s'anomena **camí** (*path*).

Observació C.10 (Notació de camins). En un graf simple, un passeig queda determinat unívocament per la seva seqüència de vèrtexs $v_0 v_1 \dots v_k$. Per tant, sovint especificarem un camí simplement llistant els seus vèrtexs.

Definició C.11 (Connexió i components).

- Dos vèrtexs u i v d'un graf $\mathcal{G}$ estan **connectats** si existeix un camí (u, v) a $\mathcal{G}$. La relació de connexió és una relació d'equivalència sobre el conjunt de vèrtexs $\mathcal{V}$.
- Les classes d'equivalència d'aquesta relació s'anomenen **components** de $\mathcal{G}$.
- Un graf $\mathcal{G}$ és **connex** si té exactament una component. Altrament, és **desconnex**. El nombre de components d'un graf $\mathcal{G}$ es denota per $\omega(\mathcal{G})$.

CICLES

S'aïllen els cicles com els passeigs tancats que permeten distingir estructures acícliques de les que no ho són. Aquesta noció serà el punt de partida per definir els DAGs.

Definició C.12 (Cicle). Un **cicle** és un passeig tancat (origen i termini coincideixen) que és un recorregut on l'únic vèrtex repetit és l'origen com a termini. Un cicle de longitud k s'anomena **k -cicle**. Un 3-cicle s'anomena sovint **triangle**. Un cicle és **senar** o **parell** segons si la seva longitud és senar o parell.

GRAFS DIRIGITS

Aquí passem dels grafs no dirigits als digrafs, que són l'estructura natural per representar dependències causals. L'objectiu és definir pares, fills, avantpassats, descendents i DAGs, i justificar el paper de l'ordre topològic.

Fins ara hem considerat grafs on les arestes són bidireccionals. No obstant això, en moltes aplicacions, com ara xarxes de trànsit unidireccional o diagrames de flux, és crucial considerar la direcció. Això ens porta al concepte de graf dirigit.

Definició C.13 (Graf dirigit). Un **graf dirigit** o **digraf** $\mathcal{D}$ és un triplet $(\mathcal{V}(\mathcal{D}), A(\mathcal{D}), \psi_{\mathcal{D}})$ format per un conjunt no buit de vèrtexs $\mathcal{V}(\mathcal{D})$, un conjunt d'arcs $A(\mathcal{D})$ disjunt de $\mathcal{V}(\mathcal{D})$, i una funció d'incidència $\psi_{\mathcal{D}}$ que associa a cada arc un parell **ordenat** de vèrtexs de $\mathcal{D}$. Si $\psi_{\mathcal{D}}(a) = (u, v)$, es diu que u és la **cua** i v és el **cap** de l'arc a .

Observació C.14 (Tipus d'arestes i grafs mixtos). En el context de la inferència causal, és comú treballar amb grafs que poden contenir diferents tipus d'arestes entre dos vèrtexs u i v : **no dirigides** ($u - v$), **dirigides** ($u \rightarrow v$ o $u \leftarrow v$) i **bidirigides** ($u \leftrightarrow v$). Un graf que conté tant arestes dirigides com no dirigides o bidirigides s'anomena **graf mixt**.

Definició C.15 (Conceptes de digrafs).

- El **graf subjacent** d'un digraf $\mathcal{D}$ és el graf obtingut substituint cada arc per una aresta amb els mateixos extrems.
- Una **orientació** d'un graf $\mathcal{G}$ és un digraf obtingut assignant una direcció a cada aresta de $\mathcal{G}$.
- Un **passeig dirigit** és una seqüència $v_0 a_1 v_1 \dots a_k v_k$ on $a_i = (v_{i-1}, v_i)$ per a tot i . Els conceptes de **recorregut dirigit**, **camí dirigit** i **cicle dirigit** es defineixen de manera anàloga.

Definició C.16 (Relacions familiars en digrafs). Donat un arc (u, v) en un digraf:

- El vèrtex u s'anomena **pare** de v .
- El vèrtex v s'anomena **fill** de u .

Aquestes relacions es generalitzen a través de camins dirigits:

- Si existeix un camí dirigit de u a v , es diu que u és un **avantpassat** de v .
- De manera equivalent, v és un **descendent** de u .

Aquesta analogia familiar és especialment útil: els pares són avantpassats dels seus fills, dels fills dels seus fills, i així successivament.

Notació C.17. Denotem els conjunts de pares, fills, avantpassats i descendents d'un vèrtex V com a Pa_V , Ch_V , An_V i De_V , respectivament. Si els vèrtexs estan indexats com a V_k , utilitzem la notació simplificada $\text{Pa}_k := \text{Pa}_{V_k}$ (i anàlogament per a la resta).

Nota C.18 (Convenció causal). En aquest annex mantenim la notació habitual, com ara $\mathcal{V}(\mathcal{G})$ o, si no hi ha ambigüitat, $\mathcal{V}$. En els capítols causals, en canvi, escriurem $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ amb $\mathcal{V} = \{V_i\}_{i \in I}$ i

$I = \{1, \dots, n\}$. Així, en la nostra notació, una aresta causal s'escriu entre variables, per exemple $V_i \rightarrow V_j$, i no només entre índexs.

Nota C.19. Si $\mathcal{D} = (\mathcal{V}, A)$ és un digraf i $S \subseteq \mathcal{V}$, llavors $S \subseteq \text{An}_S$, $\text{Pa}_S \subseteq \text{An}_S$, $S \subseteq \text{De}_S$ i $\text{Ch}_S \subseteq \text{De}_S$.

Una classe especialment important de digrafs són aquells que no contenen cicles dirigits. Aquests grafs són fonamentals en l'estudi de planificació de tasques, anàlisi de dependències i molts altres problemes algorísmics.

Definició C.20 (DAG i ordre topològic).

- Un **graf acíclic dirigit** (DAG, de l'anglès *Directed Acyclic Graph*) és un digraf que no conté cap cicle dirigit.
- Un **ordre topològic** dels vèrtexs d'un digraf finit $\mathcal{D}$ és un ordre total $<$ sobre $\mathcal{V}(\mathcal{D})$ compatible amb els arcs: si $u \rightarrow v$ és un arc de $\mathcal{D}$, aleshores $u < v$. Equivalentment, si $m = v(\mathcal{D})$, podem enumerar els vèrtexs com $v_1, \dots, v_m$ de manera que tot arc $v_i \rightarrow v_j$ satisfaci $i < j$.

Una propietat fonamental dels DAGs finits és que sempre admeten un ordre topològic.

Teorema C.21. Un digraf finit $\mathcal{D}$ té un ordre topològic si i només si és un DAG.

Demostració.

$\Rightarrow$ Si $\mathcal{D}$ té un ordre topològic i l'escrivim com una enumeració compatible $v_1, \dots, v_m$, no pot tenir cap cicle dirigit. Si existís un cicle $v_{i_1} \rightarrow v_{i_2} \rightarrow \dots \rightarrow v_{i_k} \rightarrow v_{i_1}$, hauria de ser $i_1 < i_2 < \dots < i_k < i_1$, la qual cosa és una contradicció.

$\Leftarrow$ Si $\mathcal{D}$ és un DAG, ha de contenir almenys un vèrtex sense arcs entrants (un vèrtex «font»). Si no fos així, podríem construir un camí dirigit cap enrere indefinidament, i com que el graf és finit, acabariem repetint un vèrtex, formant un cicle. L'algorisme per a trobar un ordre topològic és el següent:

1. Trobem un vèrtex v_1 sense arcs entrants i el posem com a primer element de l'ordre.
2. Considerem el graf $\mathcal{D} - v_1$. Aquest també és un DAG. Trobem aleshores un vèrtex v_2 sense arcs entrants a $\mathcal{D} - v_1$.
3. Repetim el procés fins que s'hagin ordenat tots els vèrtexs.

Aquest procés sempre té èxit, ja que a cada pas el graf restant és un DAG i, per tant, sempre té un vèrtex font. ■

Aquesta secció prepara la notació gràfica que permet representar intervencions causals. El punt d'arribada és definir les operacions de mutilació del graf, que després apareixen en l'operador $\text{do}()$ i en les regles del do-calculus .

Abans de formalitzar els mecanismes d'intervenció, és necessari establir algunes propietats topològiques fonamentals i introduir les operacions de mutilació sobre els grafs dirigits, les quals ens permetran representar algebraicament l'alteració de les estructures causals.

Definició C.22 (Conjunt terminal, [BM08]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un graf dirigit. Definim el conjunt terminal de $\mathcal{G}$, també conegut a la literatura com a conjunt d'embornals (*sink set*), com la col·lecció de vèrtexs que no tenen cap descendent més enllà d'ells mateixos³¹. Formalment:

$$\text{Sink}(\mathcal{G}) = \{i \in \mathcal{V} \mid \text{De}_i^{\mathcal{G}} \setminus \{i\} = \emptyset\}$$

En el context de la inferència causal, aquests nodes representen variables de resposta purament finals; és a dir, factors que no actuen com a causa de cap altra variable dins del model considerat.

Definició C.23 (Connexió en grafs dirigits, [BM08]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un graf dirigit. Direm que $\mathcal{G}$ és *feblement connex* si, en ignorar la direccionalitat de totes les arestes $\mathcal{E}$, el graf no dirigit resultant conté un camí entre qualsevol parell de vèrtexs $V_i, V_j \in \mathcal{V}$. En l'estudi de DAGs i components causals, aquesta és la noció de connexió que farem servir habitualment.

Com veurem al llarg del capítol 2, l'acció física d'intervenir sobre un sistema anul·la els «mecanismes naturals» (cita literal de Pearl, [Pea09, Sec. 3.2.1]) que determinen el valor de certes variables. Gràficament, aquest aïllament es tradueix en l'eliminació de conjunts d'arestes.

Notació C.24 (Operacions de mutilació, [SP08, Sec. 2]; [Pea09, Sec. 3.2.1]). Sigui $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un graf dirigit o un graf acíclic mixt, i sigui $\mathbf{S} \subset \mathcal{V}$ un subconjunt qualsevol de nodes. Es defineixen les següents operacions de reducció estructural:

- Denotem per $\mathcal{G}_{\bar{\mathbf{S}}}$ el graf obtingut d'eliminar de $\mathcal{G}$ totes les arestes **entrants** als nodes pertanyents a $\mathbf{S}$. En un graf mixt, això inclou qualsevol aresta dirigida cap a un node de $\mathbf{S}$; en particular, una aresta bidirigida incident a $\mathbf{S}$ s'elimina.
- Denotem per $\mathcal{G}_{\underline{\mathbf{S}}}$ el graf obtingut d'eliminar de $\mathcal{G}$ totes les arestes **sortints** dels nodes pertanyents a $\mathbf{S}$. En un graf mixt, aquestes són les arestes dirigides que surten de $\mathbf{S}$.

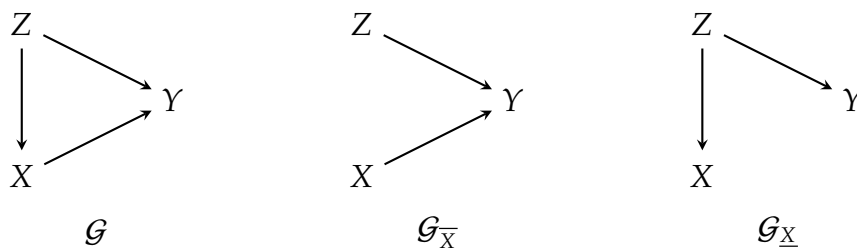


Figura C.1: Exemple de graf i les seves versions mutilades.

³¹ [SP08, Sec. 2, Defs. 6–7] inverteix la nomenclatura clàssica de la teoria de grafs. El que en matemàtiques pures anomenem un *sink*, ells l'anomenen *root*, ja que interpreten que totes les causes conflueixen cap a un efecte final (l'arrel del problema)

Bibliografia

- [Bar25] Elias BAREINBOIM. *Causal Artificial Intelligence: A Roadmap for Building Causally Intelligent Systems*. Causal AI Lab, Columbia University, 2025. URL: <https://causalai-book.net/>.
- [BP16] Elias BAREINBOIM i Judea PEARL. «Causal inference and the data-fusion problem». A: *Proceedings of the National Academy of Sciences* 113.27 (2016), pàg. 7345 - 7352. DOI: 10.1073/pnas.1510507113. URL: https://ftp.cs.ucla.edu/pub/stat_ser/r450-reprint.pdf.
Recents avenços en Causal Data Science amb la combinació causal de dades heterogènies.
- [BZL24] Elias BAREINBOIM, Junzhe ZHANG i Sanghack LEE. *An Introduction to Causal Reinforcement Learning*. Inf. tèc. R-65. Causal Artificial Intelligence Lab, Columbia University, des. de 2024. URL: <https://causalai.net/r65.pdf>.
Introducció a la disciplina de l'aprenentatge per reforç causal.
- [Bar+20] Elias BAREINBOIM et al. *On Pearl's Hierarchy and the Foundations of Causal Inference*. Technical Report R-60. First version: July 2020; last revision: March 2021. New York, NY, USA: Causal AI Lab, Columbia University, jul. de 2020. URL: <https://causalai.net/r60.pdf> (cons. 05-05-2026).
Versió tècnica per jerarquia de Pearl, C-components i resultats contrafactuals. La versió completa amb apèndixos no és pública i ha sigut facilitada pel mateix autor.
- [Bar+22] Elias BAREINBOIM et al. «On Pearl's Hierarchy and the Foundations of Causal Inference». A: *Probabilistic and Causal Inference: The Works of Judea Pearl*. 1a ed. New York, NY, USA: Association for Computing Machinery, febr. de 2022, pàg. 507 - 556. ISBN: 9781450395861. DOI: 10.1145/3501714.3501743. URL: <https://doi.org/10.1145/3501714.3501743>.
Síntesi moderna de la jerarquia causal, CBNs i identificació.
- [BHO75] P. J. BICKEL, E. A. HAMMEL i J. W. O'CONNELL. «Sex Bias in Graduate Admissions: Data from Berkeley». A: *Science* 187.4175 (1975), pàg. 398 - 404. DOI: 10.1126/science.187.4175.398. URL: <https://doi.org/10.1126/science.187.4175.398>.
Dades i anàlisi original del cas d'admissions de Berkeley.
- [BM08] J.A. BONDY i U.S.R MURTY. *Graph Theory*. 1st. Springer Publishing Company, Incorporated, 2008. ISBN: 1846289696. URL: <https://dl.acm.org/doi/10.5555/1481153>.
Definicions bàsiques de grafs, camins, connexió i conjunts terminals.
- [Bon+21] Stephan BONGERS et al. «Foundations of structural causal models with cycles and latent variables». A: *The Annals of Statistics* 49.5 (oct. de 2021). ISSN: 0090-5364. DOI: 10.1214/21-aos2064. URL: <http://dx.doi.org/10.1214/21-AOS2064>.
Intervencions perfectes i extensions dels SCMs amb cicles o latents.
- [BD80] Norman E. BRESLOW i Nicholas E. DAY. *Statistical Methods in Cancer Research. Volume I: The Analysis of Case-Control Studies*. IARC Scientific Publications 32. Lyon: International Agency for Research on Cancer, 1980, pàg. 5 - 338. URL: <https://publications.iarc.who.int/Book-And-Report-Series/Iarc-Scientific-Publications/Statistical-Methods-In-Cancer-Research-Volume-I-The-Analysis-Of-Case-Control-Studies-1980>.

- Test de Breslow–Day per a l'homogeneïtat de les odds ratios estratificades.*
- [Coc54] William G. COCHRAN. «Some Methods for Strengthening the Common χ^2 Tests». A: *Biometrics* 10.4 (1954), pàg. 417 - 451. ISSN: 0006341X, 15410420. URL: <http://www.jstor.org/stable/3001616> (cons. 09-06-2026).
- [07] *Diccionari de la Llengua Catalana*. Institut d'Estudis Catalans (IEC). 2007. URL: dlc.iec.cat (cons. 01-01-2024).
- Consulta terminològica per traduccions al català, especialment estimand.*
- [DH50] Richard DOLL i A. Bradford HILL. «Smoking and carcinoma of the lung». A: *British Medical Journal* 2.4682 (1950), pàg. 739 - 748. DOI: 10.1136/bmj.2.4682.739. URL: <https://doi.org/10.1136/bmj.2.4682.739>.
- Evidència històrica sobre tabac i càncer de pulmó.*
- [Dur19] Rick DURRETT. *Probability: Theory and Examples*. 5th. Cambridge University Press, 2019. URL: https://sites.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.
- Teoria de la mesura, esperança condicionada i esperances iterades.*
- [Fis58] Ronald A. FISHER. «Cigarettes, Cancer and Statistics». A: *The Centennial Review of Arts & Science* 2 (1958), pàg. 151 - 166. ISSN: 0008901X, 23754877. URL: <http://www.jstor.org/stable/23737529>.
- Objecció de Fisher i possible confusió latent en l'exemple del tabac.*
- [Fog21] Núria FOGUET COLL. «Inferència causal en estadística». Treball final de grau. Universitat de Barcelona, 2021.
- Contraexemple de fidelitat i introducció als models causals estructurals.*
- [GP98] D. GALLES i J. PEARL. «An Axiomatic Characterization of Causal Counterfactuals». A: *Foundations of Science* 3 (1998), pàg. 151 - 182. DOI: 10.1023/A:1009602825894.
- Axiomes contrafactuals dels SCMs i exemple del tabac.*
- [GL18] Peter GÄRDENFORS i Marlice LOMBARD. «Causal Cognition, Force Dynamics and Early Hunting Technologies». A: *Frontiers in Psychology* 9 (2018), pàg. 87. DOI: 10.3389/fpsyg.2018.00087. URL: <https://doi.org/10.3389/fpsyg.2018.00087>.
- Perspectiva cognitiva sobre el raonament causal abans de la formalització.*
- [GM87] I. J. GOOD i Y. MITTAL. «The Amalgamation and Geometry of Two-by-Two Contingency Tables». A: *The Annals of Statistics* 15.2 (1987), pàg. 694 - 711. DOI: 10.1214/aos/1176350369. URL: <https://doi.org/10.1214/aos/1176350369>.
- Mesures d'associació i amalgamació en taules 2×2 .*
- [Hal00] Joseph Y. HALPERN. «Axiomatizing Causal Reasoning». A: *Journal of Artificial Intelligence Research* 12 (2000), pàg. 317 - 337. DOI: 10.1613/jair.648.
- Completesa del marc axiomàtic contrafactual, sobretot a les demostracions.*
- [HR20] Miguel A. HERNÁN i James M. ROBINS. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.
- Guia aplicada per consistència, intercanviabilitat, positivitat i disseny d'estudis.*
- [Hol86] Paul W. HOLLAND. «Statistics and Causal Inference». A: *Journal of the American Statistical Association* 81.396 (1986), pàg. 945 - 960.
- Problema fonamental de la inferència causal i taula de resultats potencials.*
- [HT52] Daniel G. HORVITZ i Donovan J. THOMPSON. «A Generalization of Sampling Without Replace-

- ment from a Finite Universe». A: *Journal of the American Statistical Association* 47.260 (1952), pàg. 663 - 685. DOI: 10.1080/01621459.1952.10483446.
Fonament mostral dels pesos inversos via Horvitz–Thompson.
- [Hum75] David HUME. *An Enquiry Concerning Human Understanding*. Ed. de L. A. SELBY-BIGGE i P. H. NIDDITCH. 3rd. Publicat originalment el 1748. Oxford: Clarendon Press, 1975.
Per contrastar conjunció constant, correlació i connexió causal.
- [Hum00] David HUME. *A Treatise of Human Nature*. Ed. de D. F. NORTON i M. J. NORTON. Publicat originalment el 1739-1740. Edició de Norton i Norton. Oxford: Oxford University Press, 2000.
Punt de partida humià: causalitat, regularitat i «ciment de l'univers».
- [Ima13] Kosuke IMAI. *Causal Inference*. <https://imai.fas.harvard.edu/teaching/files/causal.pdf>. Material docent, Harvard University. 2013. (Cons. 10-06-2026).
- [IR15] Guido W. IMBENS i Donald B. RUBIN. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press, 2015. URL: www.cambridge.org/9780521885881.
Referència principal per resultats potencials, mecanismes d'assignació i estimadors.
- [JM11] Olga JULIA DE FERRAN i David MÁRQUEZ-CARRERAS. *Un primer curs d'estadística*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona, 2011. ISBN: 978-84-475-3483-8.
Separació entre l'espai probabilístic subjacent i la família de lleis sobre l'espai observable.
- [KF09] Daphne KOLLER i Nir FRIEDMAN. *Probabilistic Graphical Models: Principles and Techniques*. 1st. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press, 2009, pàg. 1231. ISBN: 978-0262013192.
Resultats de models gràfics probabilístics per a propietats de Markov i equivalència.
- [Lau96] Steffen L. LAURITZEN. *Graphical Models*. Vol. 17. Oxford Statistical Science Series. Oxford University Press, 1996. ISBN: 978-0-19-852219-5.
Base probabilística per independència condicional i propietats de Markov.
- [Lew73] David LEWIS. «Counterfactuals and Comparative Possibility». A: *Journal of Philosophical Logic* 2.4 (1973), pàg. 418 - 446. DOI: 10.1007/bf00262950.
Base filosòfica de la dependència contrafactual de Lewis.
- [Lew81] David LEWIS. «Counterfactuals and Comparative Possibility». A: *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Ed. de William L. HARPER, Robert STALNAKER i Glenn PEARCE. Dordrecht: D. Reidel, 1981, pàg. 57 - 85. ISBN: 978-90-277-1220-2.
Suport complementari per al rerefons contrafactual de Lewis.
- [MH59] Nathan MANTEL i William HAENSZEL. «Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease». A: *Journal of the National Cancer Institute* 22.4 (1959), pàg. 719 - 748. URL: <https://academic.oup.com/jnci/article-abstract/22/4/719/900746>.
Contrast estratificat de Mantel–Haenszel i estimació d'una odds ratio comuna.
- [Men15] Elliott MENDELSON. *Introduction to Mathematical Logic*. Angl. 6a ed. Discrete Mathematics and Its Applications. A Chapman & Hall book. Boca Raton, FL: CRC Press, 2015. 448 pàg. ISBN: 978-1-4822-3772-6. DOI: 10.1201/b18519. URL: <https://doi.org/10.1201/b18519>.
Suport lògic per al lema de Lindenbaum en la demostració de completesa.

- [Mil73] John Stuart MILL. «A System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation». A: *The Collected Works of John Stuart Mill, Volume 7: A System of Logic, Ratiocinative and Inductive, Books I–III*. Ed. de John M. ROBSON. Originally published in 1843; 8th edition 1872. Toronto i London: University of Toronto Press i Routledge & Kegan Paul, 1973. URL: <https://oll.libertyfund.org/title/mill-collected-works-vol-vii-a-system-of-logic-books-i-iii>.
Antecedent històric de la causalitat entesa contrafactualment.
- [MA13] Stephen MUMFORD i Rani Lill ANJUM. *Causation: A Very Short Introduction*. 1a ed. Very Short Introductions 371. Oxford: Oxford University Press, 2013. ISBN: 978-0-19-968443-4.
Context filosòfic breu sobre causalitat, contrafactuals i regularisme.
- [Nag23] Gabriel NAGY. *Real Analysis, Section 4: The Radon-Nikodym Theorem*. 2023. URL: <https://www.math.ksu.edu/~nagy/real-an/4-04-rn.pdf>.
Referència puntual per Radon–Nikodym; últim accés: 16-02-2026.
- [Nea20] Brady NEAL. *Introduction to Causal Inference*. Des. de 2020. URL: https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf (cons. 23-07-2025).
Recurs visual per distingir observació i intervenció.
- [Par23] Álvaro PARAFITA. «Causality without Estimands: from Causal Estimation to Black-Box Introspection». Dirigit per Jordi Vitrià. Tesi doct. Barcelona, Catalunya: Universitat de Barcelona, maig de 2023. URL: <https://hdl.handle.net/2445/202121>.
Notació i exemples propers sobre grafs causals i estimands.
- [PV20] Álvaro PARAFITA i Jordi VITRIÀ. *Causal Inference with Deep Causal Graphs*. 2020. arXiv: 2006.08380 [stat.ML]. URL: <https://arxiv.org/abs/2006.08380>.
Deep Causal Graphs com a línia computacional relacionada.
- [PV22] Álvaro PARAFITA i Jordi VITRIÀ. «Estimand-Agnostic Causal Query Estimation with Deep Causal Graphs». A: *IEEE Access* 10 (2022). Departament de Matemàtica i Informàtica, Universitat de Barcelona, pàg. 71370–71386. DOI: 10.1109/ACCESS.2022.3188395.
Context per als Deep Causal Graphs i l'estimació de consultes causals.
- [Pea88] Judea PEARL. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988. ISBN: 1558604790.
Font fundacional per xarxes bayesianes, d-separació i cobertura de Markov.
- [Pea93] Judea PEARL. «Comment: Graphical Models, Causality and Intervention». A: *Statistical Science* 8.3 (1993), pàg. 266–269. ISSN: 08834237. URL: <http://www.jstor.org/stable/2245965> (cons. 26-03-2026).
Criteris back-door i front-door abans del do-calculus complet.
- [Pea95] Judea PEARL. «Causal diagrams for empirical research». A: *Biometrika* 82.4 (des. de 1995), pàg. 669–688. ISSN: 0006-3444. DOI: 10.1093/biomet/82.4.669. eprint: <https://academic.oup.com/biomet/article-pdf/82/4/669/698263/82-4-669.pdf>. URL: <https://doi.org/10.1093/biomet/82.4.669>.
Article fundacional sobre diagrames causals i do-calculus.
- [Pea09] Judea PEARL. *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press, 2009.
Referència central per SCMs, CBNs, d-separació, do-calculus i contrafactuals.

- [PGJ16] Judea PEARL, M. GLYMOUR i N. JEWELL. *Causal Inference in Statistics: A Primer*. UK: John Wiley i Sons Ltd, 2016.
Introducció compacta a les hipòtesis gràfiques i a la inferència causal.
- [PM18] Judea PEARL i Dana MACKENZIE. *The Book of Why: The New Science of Cause and Effect*. 1st. USA: Basic Books, Inc., 2018. ISBN: 046509760X. URL: <https://dl.acm.org/doi/10.5555/3238230>.
Lectura accessible per la jerarquia causal, els límits de l'observació i exemples com el tabac.
- [PVP21] Martí PEDEMONTE, Jordi VITRIÀ i Álvaro PARAFITA. *Algorithmic Causal Effect Identification with causaleffect*. Technical Report. Universitat de Barcelona, jul. de 2021. arXiv: 2107.04632. URL: <https://arxiv.org/abs/2107.04632>.
Definicions separades de model causal estructural i model causal probabilístic; algorismes ID/IDC.
- [PJS17] Jonas PETERS, Dominik JANZING i Bernhard SCHÖLKOPF. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts; London, England: The MIT Press, 2017. ISBN: 9780262037310.
Suport formal per propietats de Markov, fidelitat, SCMs i intervencions.
- [Rei56] Hans REICHENBACH. *The Direction of Time*. Ed. de Maria REICHENBACH. Berkeley, CA: University of California Press, 1956.
Direcció causal i límits de llegir causalitat des de la correlació.
- [RR83] Paul R. ROSENBAUM i Donald B. RUBIN. «The central role of the propensity score in observational studies for causal effects». A: *Biometrika* 70.1 (abr. de 1983), pàg. 41 - 55. ISSN: 0006-3444. DOI: 10.1093/biomet/70.1.41. eprint: <https://academic.oup.com/biomet/article-pdf/70/1/41/662954/70-1-41.pdf>. URL: <https://doi.org/10.1093/biomet/70.1.41>.
Ignorabilitat forta, valor de propensió i valors d'equilibratge.
- [Rub74] Donald B. RUBIN. «Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies». A: *Journal of Educational Psychology* 66.5 (1974), pàg. 688 - 701. DOI: 10.1037/h0037350.
Formulació moderna dels resultats potencials, SUTVA i efectes causals.
- [Rub05] Donald B. RUBIN. «Causal Inference Using Potential Outcomes: Design, Modeling, Decisions». A: *Journal of the American Statistical Association* 100.469 (2005), pàg. 322 - 331. ISSN: 01621459. URL: <http://www.jstor.org/stable/27590541>.
Font de la figura sobre resultats observats i contrafactuals.
- [Sam93] Myra L. SAMUELS. «Simpson's Paradox and Related Phenomena». A: *Journal of the American Statistical Association* 88.421 (1993), pàg. 81 - 88. DOI: 10.1080/01621459.1993.10594297. eprint: <https://doi.org/10.1080/01621459.1993.10594297>.
Reversió d'associació i condicions en taules estratificades.
- [San99] Marta SANZ I SOLÉ. *Probabilitats*. Vol. 28. Col·lecció UB. Edicions Universitat Barcelona, 1999. ISBN: 9788483380918.
Espais de probabilitat i lleis com a mesures imatge.
- [SP06] Ilya SHPITSER i Judea PEARL. «Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models». A: *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, 2006, pàg. 1219 - 1226. URL: <https://cdn.aaai.org/AAAI/2006/AAAI06-191.pdf>.
Algorisme ID, hedges i identificació de distribucions intervingudes.

- [SP07] Ilya SHPITSER i Judea PEARL. «What Counterfactuals Can Be Tested». A: *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2007, pàg. 352 - 359. *Algorismes ID* i IDC* per a la identificació de consultes contrafactuals.*
- [SP08] Ilya SHPITSER i Judea PEARL. «Complete Identification Methods for the Causal Hierarchy». A: *J. Mach. Learn. Res.* 9 (set. de 2008), pàg. 1941 - 1979. ISSN: 1532-4435. URL: <https://jmlr.org/papers/v9/shpitser08a.html>. *Correctesa i completesa dels mètodes ID/IDC dins la jerarquia causal.*
- [SP12] Ilya SHPITSER i Judea PEARL. «Identification of Conditional Interventional Distributions». A: *CoRR abs/1206.6876* (2012). arXiv: 1206.6876. URL: <http://arxiv.org/abs/1206.6876>. *Algorisme IDC per consultes intervingudes amb condicionament.*
- [Sim51] E. H. SIMPSON. «The Interpretation of Interaction in Contingency Tables». A: *Journal of the Royal Statistical Society: Series B (Methodological)* 13.2 (jul. de 1951), pàg. 238 - 241. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1951.tb00088.x. URL: <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>. *Cas clàssic de reversió en taules de contingència.*
- [Spi18] Benedict de SPINOZA. *Ethics: Proved in Geometrical Order*. Ed. de Matthew J. KISNER. Trad. per Michael SILVERTHORNE. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2018. ISBN: 9781107706972. DOI: 10.1017/9781107706972. URL: <https://www.cambridge.org/core/books/spinoza-ethics/B33D75FB690EA35F270C9CC66B014AD1>.
- [Spl90] Jerzy SPLAWA-NEYMAN. «On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9». A: *Statistical Science* 5.4 (1990), pàg. 465 - 472. DOI: 10.1214/ss/1177012031. *Font històrica dels resultats potencials i de l'aleatorització; traducció de l'original de 1923.*
- [SW21] Jan SPRENGER i Naftali WEINBERGER. «Simpson's Paradox». A: *The Stanford Encyclopedia of Philosophy*. Ed. d'Edward N. ZALTA. Summer 2021. Metaphysics Research Lab, Stanford University, 2021. *Context filosòfic i causal de la paradoxa de Simpson.*
- [TP02] Jin TIAN i Judea PEARL. «A general identification condition for causal effects». A: *Eighteenth National Conference on Artificial Intelligence*. Edmonton, Alberta, Canada: American Association for Artificial Intelligence, 2002, pàg. 567 - 573. ISBN: 0262511290. *C-components i factorització prèvia a l'algorisme ID.*
- [VP90] Thomas VERMA i Judea PEARL. «Equivalence and synthesis of causal models». A: *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. UAI '90. USA: Elsevier Science Inc., 1990, pàg. 255 - 270. ISBN: 0444892648. URL: <https://dl.acm.org/doi/10.5555/1390681.1442797>. *Equivalències de Markov i estructures gràfiques equivalents.*
- [Ver93] Thomas S. VERMA. *Graphical Aspects of Causal Models*. Inf. tèc. R-191. UCLA Cognitive Systems Laboratory, 1993. *Forma canònica dels models semi-Markovians amb variables latents.*
- [Vig15] Tyler VIGEN. *Spurious Correlation #1636: Sales of LP Vinyl Albums Correlates with Costco Wholesale's Stock Price*. Accedit: 2025-07-25. 2015. URL: <https://tylervigen.com/spurious>

[/correlation/1636_sales-of-lpviny1-albums_correlates-with_costco-wholesales-stock-price.](#)

Exemple visual de correlació espúria a la Introducció.

- [Vil23] Mario VILAR. *Càlcul Integral en Diverses Variables: Cinc Cèntims de Càlcul*. Apunts de Càlcul Integral en Diverses Variables. 2023.
Apunts propis sobre mesura de Lebesgue, funcions mesurables, integral de Lebesgue i canvis de variable.
- [Vil24a] Mario VILAR. *Apunts de classe: Estadística (EST)*. Grau en Matemàtiques, curs 2023–2024. 2024.
Apunts de l'assignatura d'Estadística, impartida la primavera de 2024 per Carles Rovira i David Márquez.
- [Vil24b] Mario VILAR. *Apunts de classe: Probabilitats (PROB)*. Grau en Matemàtiques, curs 2023–2024. 2024.
Apunts de l'assignatura de Probabilitats, impartida el curs 2023–2024 per David Márquez i revisats per Alex Cañas.
- [Woo03] James WOODWARD. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press, Inc., 2003.
Lectura intervencionista de la causalitat abans d'introduir do.
- [Wor25] WORLD MEDICAL ASSOCIATION. «World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants». en. A: *JAMA* 333.1 (2025). Published online October 19, 2024; adopted by the WMA General Assembly on October 19, 2024, pàg. 71 - 74. DOI: 10.1001/jama.2024.21972. URL: <https://jamanetwork.com/journals/jama/fullarticle/2825290>.
Marc ètic per explicar per què no sempre podem aleatoritzar.
- [Wri21] Sewall WRIGHT. «Correlation and causation». A: *Journal of Agricultural Research* 20.7 (1921), pàg. 557 - 585.
Antecedent històric dels diagrames causals.
- [WG50] Ernest L. WYNDER i Evarts A. GRAHAM. «Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: a study of six hundred and eighty-four proved cases». A: *Journal of the American Medical Association* 143.4 (1950), pàg. 329 - 336. DOI: 10.1001/jama.1950.02910390001001. URL: <https://doi.org/10.1001/jama.1950.02910390001001>.
Un altre estudi clàssic sobre tabac i càncer, dins l'exemple causal.