UNIVERSITAT DE BARCELONA
Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

INFERÈNCIA CAUSAL EN ESTADÍSTICA

Autor: Núria Foguet Coll

Directors:	**Dr. Josep Fortiana Gregori (Dep. Matem. Inform. UB) i Dra. Anna Esteve Gómez (Institut Català d'Oncologia, Hospital Germans Trias i Pujol)**
Realitzat a:	**Departament de Matemàtiques i Informàtica**

Barcelona, 24 de gener de 2021

Abstract

Causal inference is a branch of mathematics focused on the study of cause-effect relationships. In this work we introduce the basics and some of the most relevant results in the structural causal models framework with the aim to identify causal effects. As an example of implementing such techniques, we apply them to a study of the causal effect of two different treatments on the survival or progression of prostate cancer patients.

Resum

La inferència causal és una branca de les matemàtiques dedicada a l'estudi de les relacions de causa-efecte. En aquest treball introduïm els fonaments i alguns dels resultats més destacats del marc de models causals estructurals amb l'objectiu d'identificar efectes causals. Com a exemple d'implementació d'aquestes tècniques, les apliquem en un estudi sobre l'efecte causal de dos tractaments diferents en la supervivència o la progressió en pacients de càncer de pròstata.

2010 Mathematics Subject Classification. 62-07, 62-09, 62P10, 62G99

Agraïments

Vull agrair al meu tutor Josep Fortiana per haver-me guiat al llarg de tot el procés, així com a l'Anna Esteve per haver-me proporcionat les dades pel treball i pels consells que m'ha donat. Per acabar, també voldria donar les gràcies a la meva família per tot el seu suport durant tot aquest temps.

Índex

1 Introducció

Una de les grans preguntes que s'han plantejat sempre els éssers humans és "Per què?". Com va dir Virgili, "Felix, qui potuit rerum cognoscere causas"[2], que traduït seria "Aquell que ha estat capaç d'entendre les causes de les coses és afortunat". En el nostre dia a dia, ens trobem constantment amb qüestions que es redueixen a relacions de causa-efecte. El pensament causal és inherent a l'ésser humà. Per exemple, si arribem tard a la feina podem preguntar-nos quina n'és la causa per tal d'evitar que es torni a repetir en un futur. Així, si pensem que la culpa és de l'hora en què estava programat el despertador, podríem arribar a concloure que si haguéssim avançat l'alarma, hauríem arribat a temps.

Més enllà de qüestions quotidianes, les preguntes causals són rellevants en àmbits tan diversos com els legals, els econòmics o els mèdics, entre d'altres. En l'àmbit legal, per demostrar la culpabilitat de l'acusat en un determinat incident, es pot argumentar que sense la implicació de l'acusat, l'incident no hagués ocorregut. En l'àmbit econòmic, qüestions com els efectes de modificar el preu d'un determinat producte en les seves ventes també són causals. En l'àmbit mèdic, és especialment rellevant el debat que va tenir lloc a finals de la dècada del 1950 i principis de la del 1960, en què es va intentar respondre a la pregunta "Fumar causa càncer de pulmó?". Malgrat que avui en dia la resposta sembli evident, en el seu moment va haver-hi una divisió d'opinions considerable, amb estadístics famosos com Ronald Fisher, el qual cobrava com a consultor de les empreses tabaqueres, argumentant que no es podia demostrar que l'associació entre fumar i el càncer de pulmó fos causal (Stolley, 1991 [35]).

Amb la finalitat de respondre aquestes qüestions de manera rigorosa, sorgeix la **inferència causal**. Les preguntes causals que busquem respondre es classifiquen en tres nivells que constitueixen l'**Escala de la Causalitat** o **Jerarquia Causal de Pearl** (Barenboim et al., 2020 [1]):

- **Associació**: El primer nivell involucra l'observació de les dades. En essència, el podem caracteritzar amb la pregunta "Què succeeix si observo...?". Per exemple, "Si observo determinats símptomes, com afecta a la probabilitat de patir una malaltia determinada?". Les preguntes d'aquest primer nivell es poden resoldre amb l'estadística tradicional, compilant i analitzant dades, estudiant-ne per exemple la correlació. Aquestes qüestions no necessàriament tindran implicacions causals (recordem el famós mantra "correlació no implica causalitat"). La intel·ligència artificial actual i la majoria d'animals es considera que actuen únicament en aquest nivell.

- **Intervenció**: El segon nivell involucra actuar o intervenir. A diferència de l'anterior nivell, que consistia en observacions passives del nostre món, aquest nivell implica en certa manera canviar-lo. La seva pregunta característica és "I si faig...?". Així, els exemples anteriors referents a modificar els preus d'un producte i a fumar pertanyen a aquest nivell. Els humans primitius que van començar a emprar eines per modificar el seu entorn es considera que van arribar al segon nivell.

- **Contrafactual**: El tercer i últim nivell va un pas més enllà i involucra l'acció d'imaginar, ja que compara el món real amb un món fictici (anomenat món contrafactual). La pregunta característica és "I si hagués actuat de manera diferent?". D'aquesta manera, l'exemple legal és un cas de qüestió contrafactual. El pensament

[2] Vers 490, Llibre 2 de *Les Geòrgiques*

contrafactual és el que ens diferencia als humans dels altres éssers vius, i ens permet reflexionar sobre les conseqüències de les nostres accions i aprendre dels nostres errors.

La classificació de l'Escala de la Causalitat és rellevant, ja que no és possible respondre preguntes d'un nivell superior amb només informació de nivells inferiors. Així, solament amb observacions passives (primer nivell) en general no podrem respondre preguntes d'intervencions (segon nivell). Anàlogament, exclusivament amb dades d'intervencions o experiments (segon nivell) no podrem contestar qüestions contrafactuals (tercer nivell), perquè no hi ha experiments que permetin comparar mons reals i ficticis. Aquest és, de fet, el **problema fonamental de la inferència causal** (Holland, 1986 [9]).

El projecte

El gran objectiu de la inferència causal és proporcionar unes eines per tal d'escalar els nivells de la Escala de la Causalitat. Amb aquesta finalitat, destaquen dos marcs. Per una banda, hi ha el model de Neyman-Rubin dels resultats potencials, el qual va ser proposat per primer cop per Jerzy Neyman en la seva tesi el 1923 (Neyman, 1923 [17]), i va ser acabat de desenvolupar per Donald B. Rubin el 1974 (Rubin, 1974 [30]). Per altra banda, hi ha el marc de models causals estructurals i diagrames causals, on va ser pioner Sewall Wright (Wright, 1920 [42], 1921 [43] i 1934 [44]) i destaca la figura de Judea Pearl (Pearl, 1982 [20], 1988 [21] i 2000 [23]). Aquest treball en particular es focalitza en aquest darrer marc, amb especial èmfasi en el salt del primer al segon nivell.

Estructura de la Memòria

La memòria està estructurada en els següents punts:

- Contextualitzem breument la inferència causal, centrant-nos en el marc dels models causals.
- Introduïm els models i diagrames causals, així com algunes de les seves propietats rellevants.
- Tractem com fer el salt del primer al segon nivell de l'Escala de la Causalitat amb els diagrames causals.
- Apliquem la teoria en un cas pràctic basat en dades reals.

2 Context històric

En els seus inicis la causalitat era un tema considerat tabú en l'estadística. Ens podem remuntar a Francis Galton, el qual es va plantejar preguntes causals sobre l'heretabilitat. El que Galton va acabar descobrint, però, va ser la llei de regressió i el concepte de correlació entre variables. Karl Pearson, deixeble de Galton, va ser molt influenciat pels resultats de Galton a *Natural Inheritance* (Galton, 1889 [5]) i va ser un dels principals responsables d'escindir la causalitat de l'estadística. Pearson considerava la causalitat un cas especial de correlació, que consistia en la repetició de determinades seqüències d'esdeveniments.

En les seves paraules,

> "Que l'univers és una suma de fenòmens, els quals són més o menys dependents entre si, és una concepció més àmplia que la de causalitat [...]. L'objectiu de la ciència deixa de ser el descobriment de 'causa' i 'efecte'; per tal de predir experiències futures busca els fenòmens que estan més altament correlacionats."(Pearson, 1911 [28])

En contra d'aquest context de rebuig a la causalitat, destaca la figura de Sewall Wright (Provine, 1986 [29]), un genetista. El color de pell dels conillets d'Índies no seguia completament les lleis de la genètica de Mendel, i Wright va postular el 1920 que "factors de desenvolupament" en l'úter eren les causes d'aquestes variacions del color. L'objectiu de Wright era determinar quantitativament l'efecte d'aquests factors en el color blanc de la pell dels conillets d'Índies. Wright va representar totes les relacions entre les variables que determinaven els colors de la pell en un diagrama de camins, on els coeficients dels camins indicaven la intensitat dels efectes causals que volia resoldre. Assumint la linealitat de les relacions entre totes les variables, la seva metodologia per trobar l'efecte causal d'una variable en una altra consistia en traçar els camins entre les variables i en multiplicar els seus coeficients. S'obtenien així unes equacions i resolent-les va trobar que el 58% de la variació en els colors de pell dels conillets d'Índies es devia als factors de desenvolupament (Wright, 1920 [42], pàg. 332).

D'aquesta manera, Wright va ser la primera persona en desenvolupar un mètode matemàtic per respondre qüestions causals. Addicionalment, Wright va postular que no es podien obtenir conclusions causals sense hipòtesis causals. En altres paraules, el primer nivell de l'Escala de la Causalitat no és suficient per respondre qüestions causals, per poder saltar a nivells superiors és necessari entendre el procés que ha generat les dades.

Tanmateix, la metodologia de l'anàlisi de camins de Wright va ser durament criticada en l'àmbit acadèmic degut a la influència de la postura de Pearson. Criticaven la subjectivitat del mètode de Wright necessària en representar els diagrames de camins (per exemple, podria succeir que dues persones suposessin diagrames diferents d'un mateix problema i analitzant les mateixes dades arribessin a resultats ben diferents) i preferien aproximacions objectives, lliures de models.

Per aquest motiu, fora del mateix Wright i d'alguns estudiants de la cria d'animals, l'anàlisi de camins va passar desapercebuda fins la dècada del 1960, quan un grup de sociòlegs com Otis Duncan i Hubert Blalock, entre d'altres, i l'economista Arthur Goldberger van redescobrir l'anàlisi de camins com un procediment per predir els efectes de polítiques educatives i socials.

No obstant, aquest vessant de la inferència causal no va avançar de manera signifi-

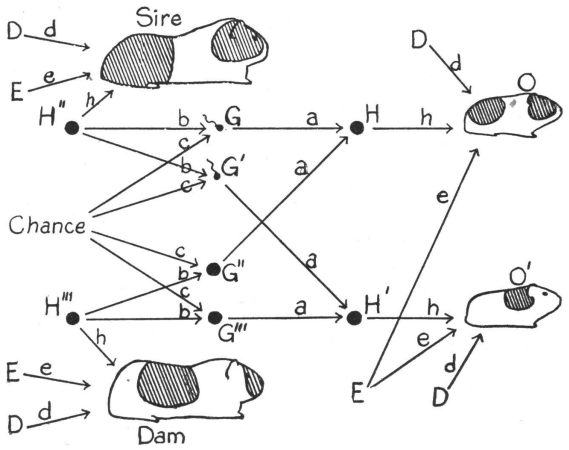

Figura 1: Primer diagrama de camins de Sewall Wright (Wright, 1920 [42], pàg. 328, Figura 5). D correspon als factors de desenvolupament, E als factors de l'entorn, H als factors hereditaris i O, O' representen les cries de conillets d'Índies. Les lletres en minúscules de són els coeficients de camins. L'objectiu era calcular els valors de d, e i h.

cativa fins a la figura de Judea Pearl, un informàtic i filòsof. Entre finals de la dècada de 1970 i principis de la de 1980, en la recerca en la intel·ligència artificial no se sabia com tractar les incerteses. Representar-les mitjançant probabilitats s'havia considerat en el seu moment, però degut a les exigències en espai d'emmagatzematge i temps de processat la idea va ser descartada en un principi. Pearl va adreçar aquestes deficiències computacionals amb el seu article de 1982 (Pearl, 1982 [20]), on va proposar representar les probabilitats amb xarxes Bayesianes, unes xarxes de variables que només interactuessin amb algunes variables dels seus entorns, en comptes d'haver de recórrer a taules de probabilitats enormes.

Pearl va culminar la seva recerca en xarxes Bayesianes amb el llibre *Probabilistic Reasoning in Intelligent Systems* publicat el 1988 (Pearl, 1988 [21]) . Aquesta obra va tenir un gran impacte en el món de la intel·ligència artificial i avui en dia les xarxes Bayesianes tenen moltes aplicacions, com ara prediccions meteorològiques (Cofiño et al., 2002 [3]), filtres d'spam (Jin et al., 2006 [10]) o identificació de víctimes en desastres (Bruijning-van Dongen et al., 2009 [2]), entre d'altres.

Malgrat l'èxit de les xarxes Bayesianes en la recerca de la intel·ligència artificial, no podien replicar la intel·ligència humana en tant que estaven limitades al primer nivell de l'Escala de la Causalitat. Així, partint dels resultats de les xarxes Bayesianes, Pearl va dedicar-se a l'estudi de la causalitat, per tal d'investigar com poder escalar al segon i tercer nivell, desenvolupant d'aquesta manera el formalisme dels models causals. Basant-se en resultats teòrics obtinguts entre el 1987 i el 2000, va establir el nou formalisme en la primera edició del seu llibre *Causality: Models, Reasoning, and Inference* publicada l'any 2000 (Pearl, 2000 [23]).

Aquest formalisme en l'actualitat és bastant utilitzat en diversos àmbits. Destaquen les figures de Sander Greenland (UCLA, 2021 [39]) i James M. Robins (Harvard Catalyst, 2021 [7]) com alguns dels principals responsables de l'adopció de models causals gràfics en l'epidemiologia. Pel que fa a l'àmbit de les ciències socials, destaquen Christopher Winship (Harvard Scholar, 2021 [8]), Stephen L. Morgan (John Hopkins University, 2021 [11]) o Felix Elwert (University of Wisconsin-Madison, 2021 [40]).

3 Models Causals Estructurals i Diagrames causals

3.1 La paradoxa de Simpson

La paradoxa de Simpson, la qual pren el nom per l'article de l'estadístic Edward H. Simpson del 1951 (Simpson, 1951 [33]), és un molt bon exemple de les limitacions de l'estadística tradicional (és a dir, el primer nivell de l'Escala) per adreçar relacions causals i il·lustra la importància de les hipòtesis causals. La paradoxa consisteix en l'existència d'unes dades on en tota la població es dona una associació determinada entre dues variables, mentre que en la població estratificada segons unes altres variables es dona l'associació oposada.

Per exemple, considerem un cas real de l'estudi de dos procediments per l'extracció de càlculs renals (Charig et al., 1986 [4]). El primer procediment consisteix en la cirúrgia oberta clàssica, mentre que el segon és la nefrolitotomia percutània, un procediment menys invasiu en el qual s'extreu el càlcul renal mitjançant un endoscopi introduït per una petita incisió a l'esquena. Les dades recullen les operacions exitoses pels dos procediments, entenent com èxit l'absència de pedres renals tres mesos després de la intervenció, o bé la reducció de les pedres a partícules de menys de 2 mm. Així mateix, també es van classificar les dades segons el diàmetre mig de les pedres en dues categories: petites (de menys de 2 cm) i grans (de més de 2 cm).

	Cirúrgia oberta	Nefrolitotomia percutània
< 2 cm	81/87 (93%)	234/270 (87%)
≥ 2 cm	192/263 (73%)	55/80 (69%)
Total	273/350 (78%)	289/350 (82%)

Taula 1: Taula amb les proporcions de procediments exitosos segons la mida de les pedres.

Si ens fixem en les dades recollides en la Taula 1, observem que la nefrolitotomia percutània té millors resultats en el còmput global: 82% d'èxits contra el 78% de la cirúrgia oberta. Per contra, si mirem les dades estratificades segons la mida de les pedres, en les dues categories la cirúrgia oberta té millors resultats. Per tant, efectivament és un exemple de la paradoxa de Simpson.

En aquesta situació la qüestió és quin tractament és preferible, ja que assignar el procediment en funció de si coneixem o no la mida de la pedra és absurd. Aquesta és una pregunta referent al segon nivell de l'Escala de la Causalitat, mentre que la informació disponible a Taula 1 correspon només al primer. Només amb la taula no en tenim prou per respondre la pregunta, ens cal considerar quin és el procés que ha generat aquestes dades, és a dir, quines són les assumpcions causals de la situació.

Hi ha dos factors rellevants a considerar. Per una banda, les pedres més grans són més severes, i per tant tenen pitjor prognosi. Per altra banda, la mida de les pedres també afecta l'assignació del tractament: s'opta més freqüentment per la cirúrgia oberta quan les pedres són grans, mentre que per les petites els metges prefereixen la nefrolitotomia percutània. En conseqüència, la mida de les pedres afecta tant el procediment quirúrgic com l'èxit de l'operació. En altres paraules, la mida és una causa comuna del tractament i del resultat. Aleshores, per inferir l'efecte del tractament en la recuperació, hem de comparar les dades estratificades segons la mida de les pedres, assegurant-nos així que les diferències entre les proporcions d'èxits entre els dos procediments no són degudes a la severitat dels càlculs renals. Per tant, la cirúrgia oberta és lleugerament millor.

En aquest exemple hem vist com la solució a la paradoxa era estratificar les dades. No obstant, aquesta no sempre és la resposta, el model causal considerat és el que determina com actuar.

Per il·lustrar-ho, considerem el següent exemple fictici. Suposem que una empresa farmacèutica ha desenvolupat un medicament que redueix la pressió arterial i li interessa estudiar si el fàrmac també és útil per la prevenció d'atacs de cor (de nou, és una pregunta referent al segon nivell de l'Escala de Causalitat). Amb aquesta finalitat, l'empresa ha realitzat un estudi on han assignat el medicament a la meitat dels participants. Les dades recollides són la proporció dels participants que no han patit atacs de cor en un període de temps després de l'assignació del tractament, així com la mesura de la pressió arterial mitjana en aquest període.

Les dades estan compilades en la Taula 2. Numèricament són exactament les mateixes que les del primer exemple, solament canvien les variables involucrades. Per tant, de nou observem una paradoxa de Simpson: si mirem el còmput global, els participants als quals s'ha assignat el fàrmac tenen menys atacs de cor (82% amb el medicament contra 78% sense); però si ens fixem en les dades estratificades segons la pressió els resultats són millors en els participants que no han estat medicats.

	No medicament	Medicament
Pressió baixa	81/87 (93%)	234/270 (87%)
Pressió alta	192/263 (73%)	55/80 (69%)
Total	273/350 (78%)	289/350 (82%)

Taula 2: Taula amb la proporció de participants que no han patit atacs de cor segons la pressió arterial.

Les hipòtesis causals, però, són molt diferents en aquest exemple. Per una banda, sabem que el medicament baixa la pressió arterial, fet que queda reflectit en les dades (dels 350 participants tractats amb el medicament, només 80 tenien la pressió alta després de medicar-se). Per altra banda, sabem que la pressió arterial alta és un factor de risc d'atacs cardíacs. Per tant, hi ha dos mecanismes pels quals el tractament pot afectar el risc d'atacs de cor: un mecanisme indirecte mediat per la pressió i un de directe. Com que ens interessa l'efecte global del medicament, no podem desestimar el mecanisme indirecte, ja que podria ser tant o més rellevant que el directe. En conseqüència, hauríem de considerar les dades no estratificades i, per tant, el medicament és beneficiós.

En conclusió, hem vist com solament amb les dades és impossible resoldre la paradoxa de Simpson: és necessari considerar el model causal en cada situació. En els següents apartats veurem com podem representar les hipòtesis causals amb Models Causals Estructurals (MCE) o gràficament amb diagrames causals com els de la Figura 2.

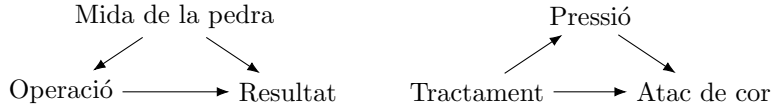

Figura 2: Diagrames causals associats als dos exemples de la paradoxa de Simpson.

3.2 Models Causals Estructurals

Abans de definir formalment els models causals estructurals (MCE), en donem un exemple inventat:

Exemple 3.1. Considerem un model causal hipotètic relacionat amb el pes d'una persona. Les variables del model en les quals estudiem els mecanismes que les causen són el pes W, l'exercici practicat X, la dieta D, l'educació E i un factor genètic G. Anomenem aquest conjunt de variables $V = \{E, G, X, D, W\}$ com el conjunt de **variables endògenes**. A cadascuna d'aquestes variables els hi associem unes variables aleatòries en forma de termes d'error o de soroll, que justifiquen la variació del conjunt V per a cada individu. Les variables per les quals no modelem el mecanisme causal, com és el cas dels termes d'error, les agrupem en un conjunt U de variables anomenades **exògenes**, en tant que es consideren determinades externament al model. Els termes d'error els denotem amb la lletra U amb el subíndex de la variable de V corresponent, de manera que $U = \{U_E, U_G, U_X, U_D, U_W\}$. Malgrat que no és el cas d'aquest exemple, podrien haver-hi variables exògenes que no siguin termes de soroll.

Com a hipòtesis causals en el model, suposem que el pes està causat per la dieta, l'exercici i el factor genètic. Per la seva part, l'exercici i la dieta que la persona realitza considerem que estan fortament influenciats pels valors inculcats en la seva educació. Addicionalment, per totes les variables també cal considerar els seus termes d'error com a causa. Només donem hipòtesis causals per a les variables endògenes, perquè són les determinades a "dins" del model (és a dir, determinades a partir d'altres variables).

Les relacions causals anteriors en els models causals estructurals es representen mitjançant un **conjunt de funcions** F (també anomenades mecanismes causals o equacions estructurals), les quals expressen les variables endògenes en funció de les variables que les causen de manera directa, incloent els termes d'error. En el cas de les variables G i E, com que no tenen altres causes en el model diferents d'aquests termes, com a funció es pren la funció identitat del terme de soroll, que denotem com Id_{U_G} i Id_{U_E} respectivament. Pel que fa a les equacions estructurals de D, X i W, són funcions f_D, f_X i f_W que no explicitarem funcionalment, de manera que el MCE direm que és **no-paramètric**.

Amb totes les variables i les relacions causals definides, podem expressar el MCE en qüestió de la següent manera:

$$U = \{U_E, U_G, U_X, U_D, U_W\}, V = \{E, G, X, D, W\}, F = \{Id_{U_G}, Id_{U_E}, f_X, f_D, f_W\}$$

$$\begin{aligned} G &:= U_G \\ E &:= U_E \\ X &:= f_X(E, U_X) \\ D &:= f_D(E, U_D) \\ W &:= f_W(D, X, G, U_W) \end{aligned}$$

Les variables que apareixen en les funcions ens indiquen les causes directes o immediates de les variables endògenes, que són les que havíem expressat amb les hipòtesis causals. Així, les causes directes del pes són la dieta, l'exercici, la genètica i el terme d'error. A més, els MCE que considerem són transitius, en el sentit que les causes directes de les causes directes d'una variable són causes d'aquesta variable. Aquesta mena de causes les anomenem indirectes, ja que estan mediades per altres variables. Per exemple, l'educació

és una causa directa de la dieta i, per tant, és una causa indirecta del pes. Si repetim aquest procés recursivament per a totes les causes directes del pes, trobem que totes les variables són causes (directes o indirectes) del pes.

Observació 3.2. A diferència de les relacions d'associació, les relacions de causalitat són asimètriques: si X és la causa de Y, no podem dir que Y és la causa de X. Per tal de reflectir aquesta asimetria en les equacions estructurals utilitzem el símbol $\coloneqq$ en comptes d'una igualtat $=$.

Els conceptes introduïts en l'exemple els podem generalitzar amb la definició formal:

Definició 3.3. *(MCE)*

Un ***Model Causal Estructural*** *és una terna* $\{U, V, F\}$ *on:*

1. *U és un conjunt de variables aleatòries determinades externament al model, que anomenem* ***variables exògenes****.*
2. *V és un conjunt de variables aleatòries determinades a partir d'altres variables del model, que anomenem* ***variables endògenes****.*
3. *F és un* ***conjunt de funcions****, una per a cada variable endògena V_i, que li assigna un valor a partir d'altres variables del model. És a dir, per a tot $V_i \in V$, existeix una funció (també anomenada* ***mecanisme causal*** *o* ***equació estructural****) $f_i \in F$ tal que:*

$$V_i \coloneqq f_i(S_i, U_{V_i})$$

on $U_{V_i} \in U$ és el terme d'error associat a V_i i $S_i \subset (V \cup U) \setminus \{V_i, U_{V_i}\}$.

En el cas en què $S_i = \varnothing$, la funció f_i és la funció identitat del terme d'error de V_i, és a dir,

$$V_i \coloneqq U_{V_i}$$

Si la forma funcional dels elements de F no està explicitada, diem que el MCE és ***no-paramètric****. Si tots els termes d'error són independents, diem que el model causal és* ***Markovià****.*

Observació 3.4. En un model paramètric completament especificat, en el qual coneixem totes les funcions de F, cada assignació $U = u$ de les variables exògenes determina unívocament els valors de V. En aquest sentit, diem que cada assignació $U = u$ correspon a una **unitat** de la població.

A partir de la definició de MCE, podem donar una definició de causalitat:

Definició 3.5. *(Causalitat)*

Sigui un MCE $\{U, V, F\}$. Diem que una variable $X \in U \cup V$ és una ***causa directa*** *d'una variable $Y \in V$ si X es troba dins la funció del MCE que assigna el valor a Y. Més generalment, diem que X és una* ***causa*** *de Y si X és una causa directa de Y o és una causa directa d'una causa de Y.*

3.3 Diagrames causals

Tota la informació no-paramètrica d'un MCE es pot representar gràficament mitjançant el que es defineixen com diagrames causals. Apart de la seva claredat visual, en les subseqüents seccions veurem que a partir d'aquests diagrames es pot extreure una gran quantitat d'informació sobre el model. Abans, però, cal introduir la terminologia necessària.

Definició 3.6. *(Graf dirigit)*

*Un **graf** G és una parella de conjunts $\{V, E\}$, on V és un conjunt no buit de vèrtexs o nodes, i E és un conjunt d'arestes que uneixen parelles de nodes. Si dos nodes estan units per una aresta, diem que són **adjacents**. Si les arestes tenen un sentit determinat, diem que es tracta d'un **graf dirigit**. En tal cas, donada una aresta diem que el node inicial és el **pare** i el node final és el **fill**.*

Notació 1. *Donat un node X, denotem per A_X el conjunt dels pares de X i per F_X el conjunt dels seus fills.*

Definició 3.7. *(Camí dirigit)*

*Donat un graf dirigit, un **camí** és una successió de nodes del graf sense cap repetició d'arestes i nodes interns. Diem que es tracta d'un **camí dirigit** si per tot node del camí no hi ha dues arestes entrants o dues arestes sortints.*

*Donat un camí dirigit entre dos nodes X i Y, diem que X és un **ancestre** de Y i que Y és un **descendent** de X. En el cas en què $X = Y$, diem que el camí dirigit que els connecta és un **cicle**.*

Notació 2. *Donat un node X, denotem per D_X els seus descendents i per $Z_X = V \setminus (D_X \cup A_X)$ els **no descendents** (on excloem a més els pares de X).*

Observació 3.8. Per conveni, donat un node X d'un graf dirigit, es considera que X és un ancestre i un descendent de si mateix.

Exemple 3.9. A la Figura 3 trobem un exemple de graf dirigit. Un camí no dirigit és $X \leftarrow Z \rightarrow W$ i un de dirigit és $X \rightarrow Y \rightarrow Z \rightarrow X$, que és un cicle. Pel node Y, els pares són $\{X\}$, els fills $\{Z\}$, els descendents $\{Y, Z, W\}$ i els ancestres $\{Y, X, Z\}$.

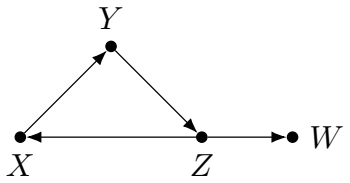

Figura 3: Exemple de graf dirigit amb un cicle.

Definició 3.10. *(Graf Acíclic Dirigit)*

*Un **Graf Acíclic Dirigit (GAD)** és un graf dirigit sense cicles.*

Tots els diagrames amb els quals treballarem són GADs. No obstant, per poder definir els diagrames causals en termes de GADs, encara ens fan falta més conceptes i assumpcions.

Definició 3.11. *(Xarxa Bayesiana)*

*Una **xarxa Bayesiana** és un model probabilístic que representa un conjunt de variables aleatòries i les seves dependències condicionals mitjançant un GAD G on els nodes són les variables aleatòries i P és la distribució de probabilitat conjunta de tots els nodes.*

Observació 3.12. La distribució P ens designa dues coses: la probabilitat pròpiament dita (distribució en el sentit de mesura o llei de probabilitat) i la seva representació computacional com una funció de massa de probabilitats o de densitat. Fem servir la notació de funció de massa, tot i que si les variables aleatòries són contínues seran densitats.

Notació 3. *En general escriurem $P(x, y)$ per expressar de manera concisa $P(X = x, Y = y)$.*

Donada una xarxa Bayesiana de nodes $\{X_i : i \in \{1, ..., n\}\}$, calcular la distribució $P(x_1, ..., x_n)$ és molt complex. En el cas més senzill en què totes les variables són binàries, caldria calcular $2^n - 1$ valors. Per n gran, no és factible perquè computacionalment requeriria massa espai de memòria i temps de càlcul. A més, a nivell experimental caldria una quantitat enorme de dades.

Per tal de simplificar aquest càlcul, podem aprofitar les indepències entre les variables. Aquí és on entra la següent assumpció:

Assumpció 3.13. *(Assumpció local de Màrkov)*

Sigui G el GAD d'una xarxa Bayesiana i sigui X un dels seus nodes. Aleshores, donats els A_X, X és independent de Z_X.

Notació 4. *Fent servir la notació habitual d'independència condicional, l'assumpció local de Màrkov es pot escriure com: $(X \perp\!\!\!\perp Z_X)|A_X$.*

Exemple 3.14. Anem a veure un contraexemple on a priori no se satisfà l'assumpció local de Màrkov i com es pot adreçar aquesta mena de problemes (Gebharter i Retzlaff, 2020 [6]). Considerem una fàbrica química F que produeix una substància S i l'emmagatzema en un recipient. No obstant, només ho aconsegueix exitosament en un 80% dels casos. A més, quan produeix la substància també genera cada vegada un contaminant C emmagatzemat en un recipient diferent. Amb aquesta informació, la xarxa Bayesiana associada al sistema seria $S \leftarrow F \rightarrow C$.

L'assumpció local de Màrkov seria que $(S \perp\!\!\!\perp C)|F$. Considerem $F = 1$ quan la fàbrica és activa, $S = 1$ quan es produeix la substància i $C = 1$ quan es genera el contaminant. Per hipòtesi,

$$P(S = 1|F = 1) = P(C = 1|F = 1) = 0.8$$
$$P(S = 1|F = 1, C = 1) = P(C = 1|F = 1, S = 1) = 1$$

Per tant, $P(S = 1|F = 1, C = 1) \neq P(S = 1|F = 1)$, incomplint l'assumpció local de Màrkov.

Exemples com aquest es poden justificar si assumim que el model no és correcte o les variables escollides no són les més adequades. Així, podem considerar la reacció química R mitjançant la qual la fàbrica genera S i C. Podem representar la nova xarxa Bayesiana incloent R com a la Figura 4.

R no és una variable observada, només n'observem els seus productes. Amb la nova xarxa, podem considerar que el 80% és la probabilitat que s'activi la reacció ($R = 1$) si la fàbrica està en funcionament, i que un cop és activada sempre es generen els productes. La nova xarxa compleix l'assumpció local de Màrkov, ja que $(S \perp\!\!\!\perp C)|R$.

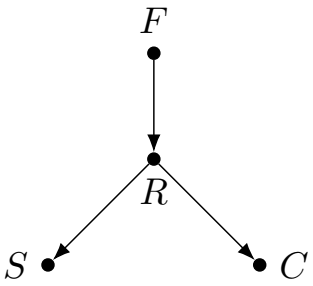

Figura 4: GAD de l'exemple de la fàbrica química modificat.

La importància de l'assumpció local de Màrkov ve donada per resultats com el següent teorema:

Teorema i definició 3.15. *(Regla de la cadena per a les xarxes Bayesianes)(Koller i Friedman, 2009 [12], pàgs. 62-63, Teoremes 3.1 i 3.2)*

Sigui una xarxa Bayesiana de nodes $X_1, ..., X_n$ i GAD G. L'assumpció local de Màrkov és equivalent a la factorització de la distribució conjunta de probabilitat P com:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|a_i) \tag{3.1}$$

on A_i són els pares de X_i.

Aquesta factorització es coneix com la ***regla de la cadena per a les xarxes Bayesianes*** *o* ***compatibilitat de Màrkov*** *(diem que P i G són* ***compatibles*** *si P factoritza d'acord amb la regla de la cadena de G).*

Demostració:

Comencem demostrant que l'assumpció local de Màrkov implica la factorització en qüestió. Suposem sense pèrdua de generalitat que els nodes $X_1, ..., X_n$ estan ordenats topològicament, és a dir, que sempre que tenim una aresta dirigida $X_i \to X_j$ es compleix $i < j$ [3]. Aplicant la regla del producte o de la cadena de les probabilitats (Koller i Friedman, 2009 [12], pàg. 18, equació (2.3)), tenim:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_1, ..., x_{i-1})$$

Hem de demostrar que els termes $P(x_i|x_1, ..., x_{i-1})$ són iguals a $P(x_i|a_i)$. Per l'ordenació dels nodes escollida, tots els pares de X_i estan en el conjunt $\{X_1, ..., X_{i-1}\}$. Similarment, en aquest conjunt no hi poden haver descendents de X_i. Per tant, tenim la igualtat de conjunts:

$$\{X_1, ..., X_{i-1}\} = A_i \cup S_i$$

on S_i és un subconjunt dels no-descendents Z_i de X_i. Finalment, l'assumpció local de Màrkov implica que $P(x_i|a_i, s_i) = P(x_i|a_i)$, com volíem veure.

[3] Existeixen algorismes que permeten trobar sempre una ordenació topològica en un GAD (Koller i Friedman, 2009 [12], pàg. 1146, Algorisme A.1)

Anem a demostrar ara l'afirmació recíproca. Suposem que tenim la descomposició:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|a_i)$$

Sigui X un node de la xarxa qualsevol i reordenem els nodes de manera que $X = X_i$, $A_i = \{X_1, ..., X_k\}$ (amb $k < i$) i els descendents de X_i siguin $\{X_{i+1}, ..., X_n\}$. Hem de veure que $P(x_i|x_1, ..., x_{i-1}) = P(x_i|x_1, ..., x_k)$.

Desenvolupant el terme de l'esquerra:

$$\begin{aligned} P(x_i|x_1, ..., x_{i-1}) &= \frac{P(x_1, ..., x_i)}{P(x_1, ..., x_{i-1})} = \frac{\sum_{x_{i+1},...,x_n} P(x_1, ..., x_n)}{\sum_{x_i,...,x_n} P(x_1, ..., x_n)} = \\ &= \frac{\sum_{x_{i+1},...,x_n} \prod_{j=1}^{n} P(x_j|a_j)}{\sum_{x_i,...,x_n} \prod_{j=1}^{n} P(x_j|a_j)} = \\ &= \frac{\prod_{j=1}^{i} P(x_j|a_j) \cdot \left[\sum_{x_{i+1},...,x_n} \prod_{j=i+1}^{n} P(x_j|a_j)\right]}{\prod_{j=1}^{i-1} P(x_j|a_j) \cdot \left[\sum_{x_i} P(x_i|a_i) \cdot \left(\sum_{x_{i+1},...,x_n} \prod_{j=i+1}^{n} P(x_j|a_j)\right)\right]} \\ &= P(x_i|a_i) \end{aligned}$$

En la segona línia utilitzem la factorització de la hipòtesi. Com que els descendents de X_i no són pares dels $X_1, ..., X_i$ (si ho fossin, hi hauria cicles o bé nodes no-descendents passarien a ser descendents de X_i), els productes corresponents a aquests nodes els podem treure dels sumatoris en la tercera línia. Finalment, simplificant arribem al resultat que volíem demostrar. □

Gràcies al teorema anterior, l'assumpció local de Màrkov ens permet aplicar la regla de la cadena per les xarxes Bayesianes, simplificant enormement el càlcul de la distribució de probabilitat conjunta P.

Així mateix, observem que l'assumpció local de Màrkov ens parla d'independències entre variables, però no ens aporta cap informació de les relacions de dependència. Per aquest motiu, ens cal una versió més potent de l'assumpció que inclogui relacions de dependència, **l'assumpció de minimalitat.**

Assumpció 3.16. *(Assumpció de minimalitat)*

Sigui G un GAD d'una xarxa Bayesiana, es compleix:

1. *L'assumpció local de Màrkov.*
2. *Els nodes adjacents en el GAD són dependents.*

Notació 5. *En el context en què es considerin múltiples GADs, explicitarem la xarxa amb un superíndex. Així, escriurem per exemple A_X^G per referir-nos als pares de X en el GAD G.*

Observació 3.17. El primer punt és l'assumpció local de Màrkov tal com l'hem enunciada anteriorment, mentre que el segon estipula la dependència de dos nodes connectats per una aresta. Diem que és "minimal" perquè implica que no es poden eliminar arestes en el GAD, ja que estaríem creant noves independències i la distribució de probabilitats P canviaria.

Per demostrar-ho, considerem un GAD G, un node X_i i un node $X_j \in A_i^G$. Considerem també el GAD G' fruit d'eliminar l'aresta de G que va de X_j a X_i. En G', X_j és un no-descendent de X_i (si fos un descendent, en G hi hauria un cicle) i per tant tenim la relació d'independència $(X_i \perp\!\!\!\perp X_j)|A_i^{G'}$. Per contra, en G tenim que X_i i X_j són dependents ja que són adjacents i condicionar a $A_i^{G'} = A_i^G \setminus \{X_j\}$ no els pot fer independents perquè no es pot "desactivar" una aresta. Així, G i G' no tenen les mateixes independències, com volíem veure.

Fins aquest punt hem tractat les independències/dependències estocàstiques entre les variables aleatòries, però no hem parlat de causalitat. Amb tots els fonaments establerts, ens fa falta una última assumpció per poder fer el salt de les xarxes Bayesianes als diagrames causals: **l'assumpció de les arestes causals**.

Assumpció 3.18. *(Assumpció de les arestes causals)*

En un GAD, tots els pares són causes directes dels seus fills.

Amb la darrera assumpció tenim tots els ingredients necessaris per definir els diagrames causals.

Definició 3.19. *(Diagrama causal)*

*Un **diagrama causal** és una xarxa Bayesiana que satisfà les assumpcions local de Màrkov i de les arestes causals.*

Observació 3.20. L'assumpció de les arestes causals implica el segon punt de l'assumpció de minimalitat, ja que la relació de causalitat és un cas particular de dependència. Per tant, en la definició de diagrama causal no cal incloure l'assumpció de minimalitat.

Observació 3.21. Donat qualsevol MCE, podem construir el diagrama causal que li correspon. Les variables endògenes i exògenes són els nodes del GAD (tot i que per conveni no se solen representar els termes d'error) i les arestes dirigides estan determinades per la família de funcions i les variables que contenen. En el context d'un diagrama causal, ens referim al mecanisme causal d'una variable X_i com la distribució $P(X_i|A_i)$.

Per altra banda, donat un GAD podem extreure'n el MCE no-paramètric associat. Per cada node, els seus pares en el GAD són les seves causes directes i per tant són les variables que apareixen en les funcions que el generen (apart, s'han d'afegir els termes d'error). Com que la forma funcional la desconeixem, el MCE que obtenim és no-paramètric.

Observació 3.22. En un model Markovià, el fet d'excloure els termes d'errors en el diagrama causal no suposa cap incompatibilitat amb la factorització de les xarxes Bayesianes.

Per veure-ho, considerem un diagrama causal G de nodes $X_1, ..., X_n$. Sigui G' el diagrama causal extès amb els termes d'error $U_1, ..., U_n$, aleshores per tot node X_i de G, els seus pares en G' són $A'_i = A_i \cup U_i$. La factorització de P a G' és:

$$P(x_1, ..., u_n) = \prod_{i=1}^{n} P(x_i|a_i, u_i) \cdot P(u_i) = \prod_{i=1}^{n} \frac{P(x_i, a_i, u_i) \cdot P(u_i)}{P(a_i, u_i)} =$$
$$= \prod_{i=1}^{n} \frac{P(x_i, a_i, u_i)}{P(a_i)} = \prod_{i=1}^{n} P(x_i, u_i|a_i)$$

Per l'assumpció local de Màrkov, A_i i U_i són independents i podem substituir $P(a_i, u_i) = P(a_i)P(u_i)$. Sumant pels termes d'error obtenim la distribució conjunta de G, que en efecte coincideix amb la regla de la cadena aplicada a G:

$$P(x_1, ..., x_n) = \sum_{u_1,...,u_n} P(x_1, ..., u_n) = \prod_{i=1}^{n} P(x_i|a_i)$$

Exemple 3.23. Podem construir el diagrama causal associat al MCE del pes establert en la secció anterior (recordant que hem exclòs els termes d'error per conveni), com es mostra a la Figura 5.

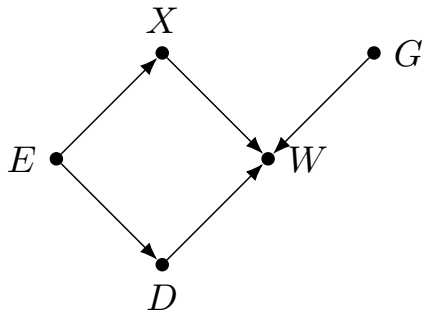

Figura 5: Diagrama causal associat al MCE del pes.

Amb els diagrames causals definits, podem començar a estudiar les relacions d'associació (o dependència) i de causalitat entre les variables a partir dels diagrames. Donades dues variables del diagrama, per cada camí que les uneixin considerem el diagrama constituït únicament pel camí. Per cadascun d'aquests camins, volem determinar si les variables inicials i finals són dependents o són una la causa de l'altra. Si la resposta és afirmativa, diem que la dependència o la causalitat *flueixen* pel camí en qüestió. Si la causalitat/dependència flueixen per almenys un dels camins, diem que les dues variables estan relacionades causalment o són dependents, respectivament. En canvi, si la causalitat/dependència no flueixen per cap dels camins, diem que no estan relacionades causalment o són independents, respectivament. Resoldre aquestes qüestions és l'objectiu amb el qual desenvolupem les següents subseccions.

Exemple 3.24. Per il·lustrar el tipus de problema, considerem el diagrama causal de la Figura 5. Suposem que volem conèixer la relació entre les variables D i X, les quals estan unides per dos camins: $X \leftarrow E \rightarrow D$ i $X \rightarrow W \leftarrow D$. Aleshores, hem d'estudiar aquests camins per separat per veure com estan relacionades X i D en el diagrama causal.

3.3.1 Blocs de construcció elementals dels diagrames causals

En aquesta secció estudiarem les dependències i independències estocàstiques en els diagrames causals més elementals i després ho generalitzarem a la resta de casos. Obviarem en les representacions gràfiques els termes d'error, que considerarem sempre independents. Llevat de les afirmacions causals, la resta és aplicable a totes les xarxes Bayesianes.

Diagrames causals d'un node

Només existeix un diagrama causal d'un sol node, que consisteix en el node aïllat. Com que no hi ha més variables, trivialment és independent de la resta.

Diagrames causals de dos nodes

Existeixen únicament dos diagrames causals consituïts per dos nodes:

- Dos nodes aïllats X i Y. Els nodes són independents, es pot veure aplicant la regla de la cadena de les xarxes Bayesianes, considerant que ni X ni Y tenen pares: $P(X,Y) = P(X)P(Y)$.
- Dos nodes X i Y connectats per una aresta dirigida. Si l'aresta va de X cap a Y, per l'assumpció de les arestes causals X és causa directa de Y. En particular, X i Y estan associades (és a dir, són estocàsticament dependents).

Diagrames causals de tres nodes

Hi ha tres diagrames causals de tres nodes connexes (els casos inconnexes són combinacions de variants d'un o dos nodes):

- **Cadenes**: Diagrames causals de la forma $X \to Y \to Z$ (per exemple: la part $E \to D \to W$ de la Figura 5). La variable del mig Y se sol referir com a **mediador**. Les relacions de dependència/independència són:
  - X, Y i Y, Z són dependents i relacionats causalment pels mateixos arguments que en el cas de dos nodes connectats.
  - Z depèn de Y, que a la seva vegada depèn de X. Per tant, és raonable pensar que Z depengui de X, malgrat que en alguns casos patològics siguin independents (Pearl et al., 2016 [26], pàg. 38). Addicionalment, per la direcció de les arestes, X és una causa potencial de Z (com en la dependència, hi ha alguns casos poc comuns intransitius).
  - $(X \perp\!\!\!\perp Z)|Y$, es pot demostrar aplicant la regla de la cadena de les xarxes Bayesianes:

$$P(x|z,y) = \frac{P(x,y,z)}{P(y,z)} = \frac{P(x) \cdot P(y|x) \cdot P(z|y)}{P(z|y) \cdot P(y)} = P(x|y)$$

- **Bifurcacions**: Diagrames de la forma $X \leftarrow Y \to Z$ (per exemple: la part $X \leftarrow E \to D$ de la Figura 5). La variable del mig Y se sol anomenar la **causa comuna** de X i Z. Les relacions de dependència/independència són les mateixes que les de les cadenes:
  - Y, X i Y, Z són dependents i relacionats causalment segons el sentit de les arestes.
  - X canvia quan Y canvia, i el mateix succeeix per Z i Y. Per tant, un canvi en X ens aporta informació d'un canvi en Y, que implica un canvi en Z. Així, com en les cadenes, usualment podem considerar X i Z dependents malgrat que hi hagi algunes excepcions. No obstant, a diferència de les cadenes, la relació entre X i Z és purament associativa, no és causal pel sentit de les arestes.
  - $(X \perp\!\!\!\perp Z)|Y$, es pot demostrar aplicant la regla de la cadena de les xarxes Bayesianes:

$$P(x|z,y) = \frac{P(x,y,z)}{P(y,z)} = \frac{P(x|y) \cdot P(y) \cdot P(z|y)}{P(z|y) \cdot P(y)} = P(x|y)$$

- **Col·lisionadors:** Diagrames de la forma $X \to Y \leftarrow Z$ (per exemple: la part $X \to W \leftarrow D$ de la Figura 5). Sovint fem servir el terme col·lisionador per referir-nos al node central. Les relacions de dependència/independència són:
  - X, Y i Z, Y són dependents i relacionats causalment segons el sentit de les arestes.
  - X i Z són independents. Ho podem demostrar aplicant la regla de la cadena de les xarxes Bayesianes:

$$P(x,z) = \sum_y P(x,y,z) = \sum_y P(x) \cdot P(y|x,z) \cdot P(z) =$$
$$= P(x) \cdot P(z) \cdot \sum_y P(y|x,z) = P(x) \cdot P(z)$$

  - X i Z són dependents condicionats a Y. Per veure-ho, suposem donat el valor de Y. Aleshores, qualsevol canvi en X ha de ser compensat per un canvi en Z, o el valor de Y canviaria.

Exemple 3.25. Podem resoldre el problema plantejat a l'**Exemple 3.24**. El camí $X \leftarrow E \to D$ és una bifurcació i per tant flueix de X a D la dependència però no la causalitat. En canvi, el camí $X \to W \leftarrow D$ és un col·lisionador i per tant no flueixen ni una cosa ni l'altra. En conseqüència, X i D són dependents degut al primer camí, però no estan relacionades causalment.

3.3.2 Generalització: d-separació i d-connexió

Podem generalitzar els resultats anteriors a un diagrama causal arbitrari. Donats dos nodes X i Y,

- la causalitat només pot fluir de X a Y si existeix un camí dirigit de X a Y (és a dir, una juxtaposició de cadenes en el mateix sentit).
- l'associació entre X i Y pot fluir a través de qualsevol camí (dirigit o no, l'associativitat de les variables és simètrica) entre X i Y que no contingui un col·lisionador.
- si el camí entre X i Y conté un col·isionador, l'associació entre X i Y no pot fluir-hi.

Diem que un camí està **desbloquejat**, quan permet el flux d'associació entre els nodes inicials i finals. Així, qualsevol camí que no contingui un col·lisionador està desbloquejat. De manera anàloga, diem que un camí està **bloquejat** si impedeix el flux d'associació entre els nodes inicials i finals.

Els únics camins que estan bloquejats per defecte són els que contenen col·lisionadors. No obstant, condicionant a determinades variables podem bloquejar més camins. Concretament, condicionar al mediador d'una cadena o la causa comuna d'una bifurcació també bloqueja el camí que els contingui.

Per altra banda, condicionar a un col·lisionador el desbloqueja, és a dir, permet el flux d'associació entre els seus pares. Condicionar a un descendent d'un col·lisionador produeix el mateix efecte. Condicionant a un descendent, com que està associat amb el col·lisionador, obtenim informació del darrer. Aleshores, els canvis en un dels pares del col·lisionador han de ser compensats per l'altre pare, per tal que el valor del col·lisionador

no variï (si variés, també variaria el del seu descendent). Aquest efecte produït per condicionar a un col·lisionador o a un descendent seu es coneix com **biaix de col·lisionador**.

Exemple 3.26. Per veure un exemple del biaix de col·lisionador, considerem solament el fragment $X \to W \leftarrow D$ del diagrama de la Figura 5. Suposem que condicionem al sector de la població amb sobrepès. En aquest sector, trobem una correlació negativa entre l'exercici i la dieta: la gent que faci exercici, seguirà una dieta poc saludable (altrament, sense més variables en el diagrama causal no es podria explicar el sobrepès), i similarment la gent que segueixi una dieta saludable, deurà dur un estil de vida molt sedentari per poder explicar el sobrepès.

Totes aquestes idees poden ser condensades amb els conceptes de d-separació i d-connexió:

Definició 3.27. *(Bloqueig de camins i d-separació)*

*Donat un GAD G, un camí c és **bloquejat** condicionadament a un conjunt de nodes W (que pot ser buit) si, i només si, es dona almenys una de les següents condicions:*

1. *c conté una cadena de tres nodes $X \to Y \to Z$ o una bifurcació de tres nodes $X \leftarrow Y \to Z$ i W conté Y.*
2. *c conté un col·lisionador $X \to Y \leftarrow Z$ i ni Y ni cap dels seus descendents estan dins de W.*

*Si W és tal que bloqueja tots els camins entre dos nodes A i B, diem que A i B estan **d-separats** per W.*

Definició 3.28. *(d-connexió)*

*Donat un GAD G i dos nodes A i B i un conjunt de nodes W que no conté ni A ni B, diem que A i B estan **d-connectats** condicionats a W si existeix almenys un camí c en G de A i B desbloquejat condicionadament a W.*

Exemple 3.29. En el GAD de la Figura 5, les variables E, X i D estan d-separades de G incondicionalment (és a dir, condicionant al conjunt buit) degut al col·lisionador en W, mentre que E i W estan d-separades per $\{X, D\}$ i X i D estan d-separades condicionant a E.

La importància de la d-separació rau en el següent teorema relacionant-la amb les independències en la distribució P:

Teorema 3.30. *(Implicacions probabilístiques de la d-separació)(Pearl, 2009 [24], pàg. 18, Teorema 1.2.4)*

Si dos conjunts de nodes X i Y en un GAD G estan d-separats per un conjunt Z, aleshores X i Y són independents condicionats a Z donada qualsevol distribució de probabilitats compatible amb G. Per contra, si X i Y no estan d-separats per Z en G, aleshores X i Y són dependents condicionats a Z en almenys una distribució P compatible amb G.

Definició 3.31. *(I(P) i I(G))*

Sigui un GAD G amb distribució de probabilitats compatible P. Definim $I(P)$ com les ternes de conjunts de variables (X, Y, Z) on $(X \perp\!\!\!\perp Y)|Z$. Anàlogament, definim $I(G)$ com les ternes de conjunts de variables (X, Y, Z) on X i Y estan d-separades per Z.

Observació 3.32. El teorema anterior ens diu que $I(G) \subset I(P)$. La inclusió contrària $I(P) \subset I(G)$ no se satisfà en alguns casos molt particulars en què hi ha variables independents malgrat estiguin connectades per un camí desbloquejat. Si volem imposar-la l'hem de postular com una assumpció anomenada **fidelitat**.

Assumpció 3.33. *(Assumpció de Fidelitat)*

Un diagrama causal G amb distribució de probabilitat P associada satisfà l'assumpció de fidelitat si $I(P) \subset I(G)$.

Exemple 3.34. Anem a donar un contraexemple on no es compleix la fidelitat. Considerem el següent MCE Markovià:

$$\begin{aligned} X &\coloneqq U_X \\ Z &\coloneqq U_Z + \alpha X \\ W &\coloneqq U_W + \beta X \\ Y &\coloneqq U_Y + \gamma Z + \delta W \end{aligned}$$

Substituint les equacions de Z i W en Y trobem $Y \coloneqq (\alpha\gamma + \delta\beta)X + \gamma U_Z + \delta U_W + U_Y$. Si imposem $\alpha\gamma = -\delta\beta$, aleshores Y és independent de X. No obstant, en el diagrama causal corresponent al MCE considerat hi ha els camins desbloquejats $X \to Z \to Y$ i $X \to W \to Y$, de manera que X i Y no estan d-separades.

Generalment, els incompliments de l'assumpció de fidelitat es donen en casos on els paràmetres del model estan ajustats de manera molt precisa (com en l'exemple, en què calia $\alpha\gamma = -\delta\beta$). A la pràctica és molt improbable que es donin aquestes situacions[4], de manera que en general assumirem fidelitat.

Relacionat amb el concepte de d-separació, podem definir un conjunt amb una propietat interessant que demostrarem seguidament.

Definició 3.35. *(Cobertura de Màrkov)*

*Sigui G un GAD i X un dels seus nodes. La **cobertura de Màrkov** de X és el conjunt constituït pels pares de X, els fills de X i els pares dels fills (excloent X).*

Proposició 3.36. *Donat un GAD G i un node X, la cobertura de Màrkov de X és el conjunt minimal que d-separa X de la resta de nodes.*

Demostració: Sigui A un pare de X, FA un fill d'A diferent de X, AA un pare d'A, F un fill de X, AF un pare de F diferent de X i FF un fill de F. Comencem veient que la cobertura de Màrkov de X el d-separa de la resta de nodes:

- Condicionar als pares de X bloquejarà tots els camins entrants en X, els quals són els que acaben en $AA \to A \to X$ i en $FA \leftarrow A \to X$, ja que A és el node intermig d'una cadena i d'una bifurcació respectivament.
- Condicionar als fills de X bloquejarà els camins sortints de X que comencen de la forma $X \to F \to FF$, ja que F és el node intermig d'una cadena.
- Condicionar als altres pares dels fills de X bloquejarà tots els camins que havíem desbloquejat en condicionar a F en el col·lisionador $X \to F \leftarrow AF$.

[4]Concretament, $I(G) = I(P)$ llevat de per un conjunt de distribucions P amb mesura zero (Koller i Friedman, 2009 [12], pàg. 73, Teorema 3.5).

Per veure que és minimal, hem de veure que si excloem un dels elements del conjunt, X ja no està d-separada de la resta de variables. Si traiem un pare o un fill de X, com que són adjacents a X estarien d-connectades amb X independentment de les variables que condicionem. Per altra banda, si traiem un pare d'un fill AF, tindríem un camí desbloquejat $X \rightarrow F \leftarrow AF$ en condicionar a F, i per tant X i AF no estarien d-separades. □

3.4 Implicacions comprovables

Un cop tenim condensades les assumpcions causals d'un model en un MCE o en el seu diagrama causal associat G, ens pot interessar verificar que el model escollit funcioni adequadament. És a dir, donat un conjunt de dades de les variables implicades en el model, voldríem comprovar que hagin pogut ser generades pel model considerat.

La d-separació és una forma de rebutjar models causals. Assumint fidelitat, en un diagrama causal la d-separació ens indica les independències condicionades que s'han de complir entre les variables. Per tant, si el model causal considerat suggereix que unes variables X, Y són independents condicionades a Z, aleshores aquesta independència ha de ser reflectida en les dades. Si les dades no la reflecteixen, aleshores el model considerat no pot ser correcte. A més, en el model correcte X i Y han d'estar d-connectades donada Z.

Aquest procediment mitjançant la d-separació té l'avantatge que és completament no-paramètric, amb conèixer el diagrama causal és suficient. Addicionalment, permet provar els diferents models localment, en el sentit que ens permet identificar quines parts del model són errònies, en comptes d'haver de buscar un altre model de zero.

Aplicant aquest procés de forma reiterada, podem trobar el conjunt de diagrames causals compatibles amb les independències presents en les dades. Aquest conjunt de diagrames causals és una classe d'equivalència, formada pels diagrames que tenen les mateixes implicacions comprovables. La caracterització de les classes d'equivalència ve donada pel següent teorema:

Teorema 3.37. *(Verma i Pearl, 1990 [41], Teorema 1)*

Dos diagrames causals G_1 i G_2 són equivalents si i només si tenen les mateixes adjacències i les mateixes v-estructures (col·lisionadors amb pares no adjacents).

Es poden representar tots els GADs d'una classe d'equivalència mitjançant un graf acíclic parcialment dirigit, en què les arestes reversibles no tenen direcció. Assumint fidelitat i que totes les variables que indueixen associacions no causals entre qualsevol parell de variables estan observades (**suficiència causal**), es pot identificar la classe d'equivalència seguint els següents passos que constitueixen un algorisme anomenat PC (Spirtes et al., 2001 [34], pàg. 84):

1. Partim d'un graf no dirigit complet de les variables del model (és a dir, on tots els nodes són adjacents).

2. Per cada parella de variables X i Y, eliminem l'aresta no dirigida $X - Y$ si trobem que $(X \perp\!\!\!\perp Y)|S$ per algun conjunt de variables S (on S pot ser el conjunt buit). D'aquesta manera, identifiquem les adjacències del diagrama.

3. Per cada camí $X - Z - Y$ on sabem que no hi ha una aresta $X - Y$ pel pas anterior, si Z no pertany al conjunt S que feia X, Y independents condicionalment, aleshores $X - Z - Y$ forma una v-estructura. Per tant, podem orientar les arestes $X \to Z \leftarrow Y$. Amb aquest pas identifiquem totes les v-estructures del diagrama causal.

4. Per acabar, aprofitant que hem identificat totes les v-estructures, podem orientar algunes arestes addicionals. Per cada camí del tipus $X \to Z - Y$ on no hi ha cap aresta entre X i Y, tenim que $Z \to Y$ (altrament, hauríem trobat una nova v-estructura).

Exemple 3.38. Suposem que volem identificar la classe d'equivalència del diagrama causal de la Figura 5 aplicant l'algorisme PC. Partim del graf complet no dirigit i procedim a identificar les adjacències. Notem que $E \perp\!\!\!\perp G$, $X \perp\!\!\!\perp G$, $D \perp\!\!\!\perp G$, $(X \perp\!\!\!\perp D)|E$ i $(E \perp\!\!\!\perp W)|\{X, D\}$, de manera que podem eliminar aquestes cinc arestes.

Seguidament, identifiquem les v-estructures. $E - X - W$ no n'és una, ja que $X \in \{X, D\}$, el conjunt que feia independents condicionalment E i W. Anàlogament, $E - D - W$ tampoc n'és una. Sí ho són $D - W - G$, $X - W - G$ i $X - W - D$, i per tant en podem orientar les arestes de manera acord.

Finalment, com que no hi ha camins com els descrits pel darrer pas, no podem orientar més arestes i ja hem obtingut la classe d'equivalència. En total hi ha tres diagrames causals equivalents segons si les dues arestes no dirigides estan orientades $X \leftarrow E \to D$ (el cas del model causal considerat), $X \to E \to D$ o $X \leftarrow E \leftarrow D$. La combinació $X \to E \leftarrow D$ no és possible, en tant que hi hauria una nova v-estructura.

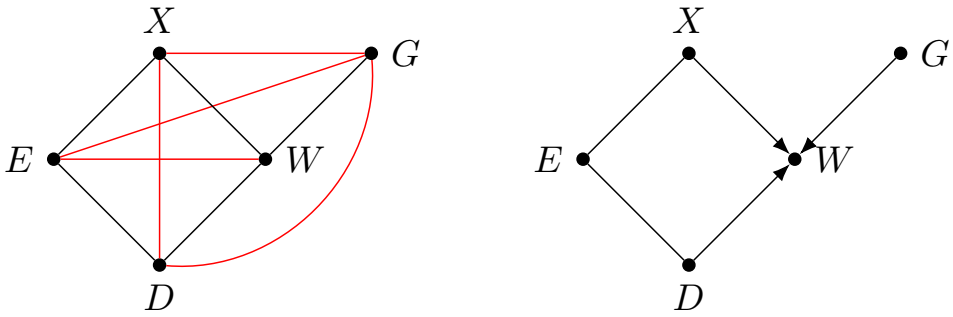

Figura 6: Implementació de PC per identificar la classe d'equivalència del GAD de la Figura 5. A l'esquerra hi ha el graf complet, on les arestes en vermell són les eliminades en el segon pas. A la dreta hi ha el graf amb les arestes orientades després d'identificar totes les v-estructures, que coincideix amb la classe d'equivalència.

Tanmateix, no podem identificar el diagrama causal dins de la classe d'equivalència sense dades d'experiments (és a dir, informació del segon nivell de l'Escala de Causalitat), alguna informació temporal (per exemple, si sabem que hi ha dues variables X i Y adjacents en el GAD i que X succeeix temporalment abans de Y, podem inferir que l'aresta va de X cap a Y), o assumpcions sobre la forma paramètrica del MCE.

Un inconvenient d'algorismes com el PC per identificar diagrames causals a partir de les independències de les dades és que a la pràctica trobar tests d'independències condicionades és un problema complicat. A més, per obtenir bons resultats poden fer falta un nombre elevat de dades.

4 Intervencions en models causals

4.1 Formalisme d'intervencions

Un dels objectius principals de la inferència causal és l'estudi de l'efecte d'intervencions en unes determinades variables **tractament** en unes variables **resultat**. Per exemple, en un estudi sobre l'efecte del tabac en el desenvolupament del càncer de pulmó, voldríem saber el resultat d'*intervenir* en la condició de fumador dels participants de l'estudi i com afecta això a les probabilitats de desenvolupar càncer.

En general, el que volem és comparar el resultat de diferents intervencions en unes mateixes variables tractament. En tal cas, ens trobem amb el problema fonamental de la inferència causal, per la impossibilitat d'observar simultàniament els efectes de dos tractaments diferents en un mateix individu sota les mateixes condicions. En l'exemple del tabac, per un mateix participant de l'estudi no podem observar el desenvolupament de càncer de pulmó quan és un fumador i quan no a la vegada.

Més enllà del problema fonamental, hi ha un obstacle eminentment pràctic. Podem classificar els estudis en dos grups: **estudis experimentals** i **estudis observacionals**. En els primers, és possible escollir com s'assignen els diferents tractaments als participants de l'estudi, de manera que podem dur a terme les intervencions que volem estudiar. En els darrers, no es pot actuar en l'assignació del tractament i l'únic que tenim són les dades observades dels participants, sense poder-hi intervenir. En aquest sentit, els estudis experimentals pertanyen al segon nivell de l'Escala de la Causalitat, mentre que els observacionals pertanyen al primer nivell. Idealment, voldríem disposar sempre d'estudis experimentals, però no sempre és possible. Un dels motius és ètic, com en el cas del tabac en què no seria moralment correcte obligar als participants a començar a fumar. En altres casos, directament és impossible realitzar a la pràctica la intervenció en qüestió (per exemple, no es pot intervenir en variables com el clima o en els gens d'una persona). En aquests casos, ens hem de conformar amb estudis observacionals.

En aquesta secció veurem com gràcies als MCE i als diagrames causals pot ser possible inferir l'efecte de les intervencions en estudis observacionals. En altres paraules, veurem com podem escalar del primer nivell de l'Escala de Causalitat al segon.

Per representar l'acció d'intervenir en el formalisme de models causals, ens cal introduir un operador:

Definició 4.1. *(Operador do)*

Donat un model causal estructural amb una variable X, definim la intervenció que fixa X en un valor x mitjançant l'operador $do(X = x)$.

Definició 4.2. *(Efecte causal)*

*Definim l'**efecte causal** de fixar $X = x$ en Y a la distribució de probabilitat $P(Y|do(X = x))$.*

Notació 6. *Quan el context ho permeti, sovint escriurem $P(y|do(x))$ en lloc de $P(Y = y|do(X = x))$.*

Observació 4.3. És important no confondre intervenir amb condicionar. Condicionar a $X = x$ implica considerar només el conjunt de la població en què X pren el valor x, mentre que intervenir $do(X = x)$ significa que imposem el valor $X = x$ a tota la població. Així, en l'exemple del tabac, si $X = 1$ indica que el participant és fumador i $Y = 1$ indica

desenvolupar càncer de pulmó, $P(Y = 1|X = 1)$ indica la probabilitat que els participants que fumen per voluntat pròpia desenvolupin càncer, mentre que $P(Y = 1|do(X = 1))$ indica les probabilitats que desenvolupin càncer si s'obliga a fumar a tots els participants de l'estudi.

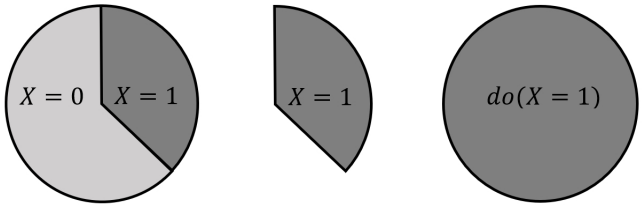

Figura 7: Il·lustració de les diferències entre condicionar i intervenir. A l'esquerra, considerem tota la població, estratificada segons una variable binària X. Al mig es mostra solament la subpoblació en què $X = 1$, la qual és la que veiem condicionant a aquest valor. Finalment, a la dreta hi ha el resultat d'intervenir fixant $X = 1$, que és a tota la població.

A partir de l'efecte causal podem definir l'**Efecte Causal Mig** quan la variable intervinguda és binària:

Definició 4.4. *(Efecte Causal Mig)*

*Donat un model causal, una variable Y i una variable X binària, definim l'**Efecte Causal Mig** (ECM) d'X en Y com:*

$$\tau = E[Y|do(X = 1)] - E[Y|do(X = 0)] \tag{4.1}$$

Observació 4.5. En general, per les diferències entre condicionament i intervenció, $\tau \neq E[Y|X = 1] - E[Y|X = 0]$.

Uns altres efectes a considerar són el d'intervencions condicionades a certs sectors de la població, denotats com **efectes específics**.

Definició 4.6. *(Efecte z-específic)*

*Donades unes variables Y, X i Z, definim com l'**efecte z-específic** de X en Y a la probabilitat*

$$P(y|do(x), z) = \frac{P(y, z|do(x))}{P(z|do(x))} \tag{4.2}$$

que indica la probabilitat de $Y = y$ en el subconjunt de la població en què $Z = z$ després de la intervenció $do(X = x)$.

Calcular els efectes z-específics per diferents valors de Z permet estudiar com afecta Z l'efecte causal de X en Y. Aquests estudis es coneixen com estudis de **moderació** i també permeten determinar els efectes d'intervencions en variables X dependents del valor de variables Z mitjançant funcions $g(Z)$:

$$\begin{aligned} P(y|do(X = g(Z))) &= \sum_z P(y|do(X = g(Z)), Z = z) \cdot P(Z = z|do(X = g(Z))) = \\ &= \sum_z P(y|do(x), z)_{x=g(z)} \cdot P(z) \end{aligned} \tag{4.3}$$

En la derivació de la fórmula, substituïm $P(z|do(X = g(Z))) = P(z)$, ja que com que fixem el valor de X en funció de Z, la probabilitat de $Z = z$ no varia amb la intervenció en X.

Exemple 4.7. Un exemple d'aplicació d'aquestes intervencions condicionades és un estudi clínic on es fixa l'assignació d'un tractament només als pacients amb uns símptomes concrets.

El nostre objectiu en aquesta secció és trobar una manera d'expressar efectes causals $P(y|do(x))$ i efectes z-específics $P(y|do(x), z)$ en termes d'expressions lliures d'operadors *do*, i per tant es puguin determinar a partir de dades observacionals. Aquest procés es coneix com **identificació**. Parlem d'identificació no-paramètrica quan aquesta es pot realitzar únicament amb les assumpcions causals del diagrama.

4.1.1 Assumpció de Modularitat

Una assumpció necessària per poder dur a terme el procés d'identificació és que les intervencions que duem a terme siguin locals. És a dir, que si intervenim una variable X_i fixant-la en un valor x_i, només modifiquem el mecanisme causal de X_i. Formalment, aquesta assumpció es tracta de l'**assumpció de modularitat** o de **mecanismes independents**.

Modularitat en diagrames causals

Assumpció 4.8. *(Assumpció de Modularitat en diagrames causals)*

Sigui G un diagrama causal, P la distribució de probabilitat i S el subconjunt de nodes de G en els quals intervenim. Donat un node qualsevol X_i de G amb pares A_i, la distribució de probabilitats P_m post-intervenció compleix:

1. *Si $X_i \notin S$, $P_m(X_i = x_i|A_i = a_i) = P(X_i = x_i|A_i = a_i)$*
2. *Si $X_i \in S$, $P_m(X_i = x_i|A_i = a_i) = 1$ si $X_i = x_i$ és igual al valor fixat per la intervenció i $P_m(X_i = x_i|A_i = a_i) = 0$ en cas contrari.*

En termes del diagrama causal G, l'assumpció de modularitat implica que una intervenció $do(S = s)$ elimina les arestes entrants als nodes de $X_i \in S$, ja que $P_m(X_i = x_i|A_i = a_i) = 1$ i per tant $X_i = x_i$ independentment dels valors dels seus pares. El graf post-intervenció l'anomenem **graf manipulat** G_m. Recordant la regla de la cadena per les xarxes Bayesianes, podem expressar la distribució conjunta P_m del graf manipulat:

Proposició 4.9. *(Factorització truncada)*

Sigui G un diagrama causal de nodes $X_1, ...X_n$, sigui S el conjunt de nodes sobre el qual intervenim. Aleshores, si els valors dels nodes $\{x_1, ..., x_n\}$ són consistents amb la intervenció (és a dir, si per $X_i \in S$, $X_i = x_i$ és el valor fixat), la distribució de probabilitat conjunta P_m després de la intervenció és:

$$P(x_1, .., x_n|do(S = s)) = P_m(x_1, ..., x_n|S = s) = \prod_{X_i \notin S} P(x_i|a_i) \tag{4.4}$$

En el cas en què els valors dels nodes no són consistents amb la intervenció,

$$P(x_1, .., x_n|do(S = s)) = 0$$

Observació 4.10. La intervenció $do(X = x)$ elimina tots els camins que induirien una associació no causal entre X i qualsevol altra variable Y, en tant que elimina les arestes entrants a X. Aquests tipus de camins s'anomenen **camins per la porta del darrere**. Eliminant-los, en el graf modificat l'única associació possible entre X i Y, si és que n'hi ha alguna, és causal.

Observació 4.11. En el cas en què X no té pares, no hi ha camins per porta del darrere per X i la intervenció $do(X = x)$ sí coincideix amb condicionar a $X = x$, ja que el diagrama manipulat per la intervenció és el mateix que l'inicial. Aquesta situació es dona en els **dissenys experimentals aleatoritzats** (per exemple en assajos clínics), en què s'assigna el tractament als participants de l'experiment de manera aleatòria. En aquests dissenys, l'associació del tractament amb el resultat és causal.

Exemple 4.12. Una intervenció en la dieta en el model causal de la Figura 5 es tradueix en el diagrama causal de la Figura 8.

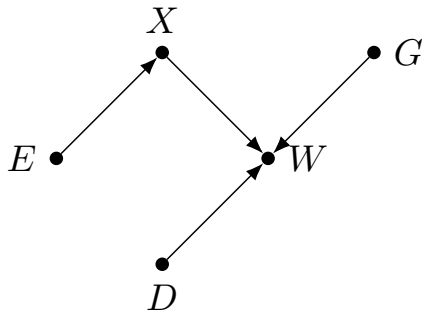

Figura 8: Diagrama causal del model posterior a una intervenció en D

Modularitat en MCEs

En termes dels MCEs, l'assumpció de modularitat l'hem de reescriure adaptada al seu formalisme:

Assumpció 4.13. *(Assumpció de Modularitat en MCEs)*

Donat un MCE $M = \{U, V, F\}$ i un subconjunt de variables endògenes S, el MCE M_s de la intervenció $do(S = s)$ és el mateix que M llevat de les equacions estructurals de les variables de S, les quals estan fixades en els valors s.

Exemple 4.14. En termes del MCE associat al diagrama de la Figura 5, el model modificat després d'intervenir en la dieta fixant-la en una dieta d (per exemple, una dieta vegetariana) és:

$$U = \{U_E, U_G, U_X, U_D, U_W\}, V = \{E, G, X, D, W\}, F = \{Id_{U_G}, Id_{U_E}, f_X, Id_d, f_W\}$$

$$\begin{aligned} G &:= U_G \\ E &:= U_E \\ X &:= f_X(E, U_X) \\ D &:= d \\ W &:= f_W(D, X, G, U_W) \end{aligned}$$

Per l'assumpció de Modularitat en els MCE podem definir el següent terme:

Definició 4.15. *(Resposta potencial)*

*Sigui $M = \{U, V, F\}$ un MCE. Siguin X, Y dos conjunts de variables de V i sigui $U = u$ una unitat. La **resposta potencial** de Y a l'acció $do(X = x)$ de la unitat $U = u$ denotada com $Y_x(u)$ és la solució de les equacions de Y en el model de la intervenció M_x:*

$$Y_x(u) = Y_{M_x}(u) \tag{4.5}$$

Notació 7. *Amb la definició de resposta potencial, es pot expressar de manera més compacta $P(Y = y|do(X = x))$ com $P(Y_x = y)$. Tanmateix, mentre considerem qüestions del segon nivell de l'Escala de Causalitat mantindrem l'operador do i reservarem la nova notació pel tercer nivell.*

Amb la definició anterior es pot demostrar la **Llei de Consistència**:

Teorema 4.16. *(Llei de Consistència)(Pearl, 2009 [24], pàg. 229, Corol·lari 7.3.2)*

Per qualssevol conjunts de variables endògenes X i Y en un model causal,

$$X(u) = x \implies Y_x(u) = Y(u) \tag{4.6}$$

Demostració

Primer justifiquem un resultat més general. Sigui Z un conjunt de variables endògenes del model i considerem l'acció $do(Z = z)$. Aplicant la definició de resposta potencial, determinem la resposta potencial de X com $X_z(u) = X_{M_z}(u) = x$. A continuació, considerem l'acció $do(X = x)$ a més de la intervenció en Z. Com que fixem X en el valor x que hagués pres sota l'acció en Z, la intervenció en X no afecta la resta de variables i la podem obviar. Per tant, $Y_{xz}(u) = Y_{M_{xz}}(u) = Y_{M_z}(u) = Y_z(u)$. Aquesta propietat es coneix com **composició**. La consistència es dedueix de la composició considerant l'acció nul·la amb $Z = \varnothing$. □

La llei de consistència ens permet equiparar observacions amb respostes potencials. És a dir, si observem que un individu rep un determinat tractament, les variables resultat observades coincideixen amb les respostes potencials sota la intervenció en el tractament observat. En el marc dels resultats potencials, la consistència no és un teorema, sinó que és una de les assumpcions que cal imposar.

Així mateix, el problema fonamental de la inferència causal expressat en aquesta terminologia ens diu que si $X(u) = x$ per una unitat $U = u$, coneixem $Y_x(u) = Y$, però desconeixem $Y_{x'}(u)$ per $x \neq x'$.

4.2 Mecanismes d'identificabilitat habituals

Amb el formalisme de les intervencions establert, anem a veure algunes de les eines principals per poder identificar expressions causals del tipus $P(Y = y|do(X = x))$.

Teorema 4.17. *(Ajustament per causes directes) (Pearl, 2009 [24], pàg. 73, Teorema 3.2.2)*

Donat un model causal, sigui X_i una variable, sigui A_i el conjunt de causes directes de X_i i sigui Y un conjunt de variables no inclòs en el conjunt $X_i \cup A_i$. L'efecte de la intervenció $do(X_i = x_i)$ en Y es pot calcular a partir de la fórmula:

$$P(Y = y|do(X_i = x_i)) = \sum_{a_i} P(Y = y|X_i = x_i, A_i = a_i) \cdot P(A_i = a_i) \tag{4.7}$$

Demostració

Siguin $\{X_1, ..., X_n\}$ les variables del model. Aplicant la factorització truncada:

$$P(x_1, ..., x_n|do(x_i)) = \prod_{j \neq i} P(x_j|a_j) = \frac{P(x_1, ..., x_n)}{P(x_i|a_i)} = P(x_1, ..., \widehat{x_i}, \widehat{a_i}, ..., x_n|x_i, a_i) \cdot P(a_i)$$

En la segona igualtat multipliquem i dividim per $P(x_i|a_i)$ i apliquem la regla de la cadena per les xarxes Bayesianes. En la darrera igualtat, la notació $\widehat{x_i}$ indica que la variable en qüestió no hi és. Finalment, sumant l'última expressió en totes les variables llevat de X_i i Y, obtenim el resultat desitjat. □

Observació 4.18. En el cas particular en què X_i no té pares, obtenim que intervenir és el mateix que condicionar, com era d'esperar.

Observació 4.19. Perquè estigui ben definida l'expressió de l'efecte causal, cal que $P(x_i, a_i) > 0$. Equivalentment, cal que $P(x_i|a_i) > 0$ per tots els valors a_i presents a la població (és a dir, amb $P(a_i) > 0$). Aquesta condició es coneix com **positivitat** i es pot interpretar com que el tractament x_i sigui assignat a tots els sectors de la població a_i. A partir d'ara, assumirem positivitat per la variable tractament donat qualsevol conjunt de variables Z en el qual haguem de condicionar.

L'ajustament per causes directes és una manera de calcular l'efecte d'una intervenció, però no sempre és possible aplicar-ho a la pràctica, perquè podria donar-se el cas que no totes les causes directes de la variable considerada siguin observables. Per això, donem un criteri més general:

Definició 4.20. *(Criteri de la porta del darrere)*

*Donat un GAD G, un conjunt de variables Z satisfà el **criteri de la porta del darrere** relatiu a una parella ordenada de variables (X, Y) si cap dels nodes de Z són descendents de X i si bloquegen tots els camins per la porta del darrere de X a Y.*

Teorema 4.21. *(Ajustament per la porta del darrere) (Pearl, 2009 [24], pàg. 79, Teorema 3.3.2)*

Donat un diagrama causal G, si un conjunt de variables Z satisfà el criteri de la porta del darrere relatiu a la parella ordenada (X, Y), aleshores l'efecte causal de X en Y ve determinat per la fórmula:

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z) \cdot P(Z = z) \tag{4.8}$$

Demostració

Siguin A els pares de X. Per l'ajustament per causes directes,

$$P(y|do(x)) = \sum_a P(y|x, a) \cdot P(a)$$

Sigui Z un conjunt de variables satisfent el criteri de la porta del darrere relatiu a

(X, Y). Sumant per tots els valors de z obtenim:

$$P(y|do(x)) = \sum_a \frac{\sum_z P(y,x,a,z)}{P(x,a)} \cdot P(a) = \sum_a \frac{\sum_z P(y|x,a,z) \cdot P(x,a,z)}{P(x,a)} \cdot P(a) =$$
$$= \sum_a \sum_z P(y|x,z,a) \cdot P(z|a,x) \cdot P(a)$$

Se satisfan les següents independències condicionals:

(i) $(X \perp\!\!\!\perp Z)|A$: es dedueix de l'assumpció local de Màrkov, ja que els nodes de Z no són descendents de X.

(ii) $(Y \perp\!\!\!\perp A)|\{X, Z\}$: per absurd, suposem que existeix un camí de Y a A desbloquejat condicionant a $\{X, Z\}$. Aleshores, aquest camí pot ser extès a un camí per la porta del darrere de X a Y desbloquejat condicionant a Z, arribant a una contradicció.

Tenint en compte aquestes independències, la darrera expressió esdevé:

$$P(y|do(x)) = \sum_a \sum_z P(y|x,z) \cdot P(z|a) \cdot P(a) = \sum_a \sum_z P(y|x,z) \cdot P(z,a) = \sum_z P(y|x,z) \cdot P(z)$$

□

Observació 4.22. El criteri de la porta del darrere està relacionat amb el concepte de d-separació, en tant que un conjunt Z satisfà el criteri relatiu a (X, Y) si, i només si, en el graf que elimina les arestes sortint de X, Z d-separa X i Y.

Observació 4.23. Poden existir múltiples conjunts de variables Z satisfent el criteri de la porta del darrere (el conjunt dels pares de X sempre n'és un), però el resultat de la fórmula d'ajustament serà el mateix. Això presenta beneficis, com poder escollir el conjunt Z amb variables més senzilles o barates de mesurar. Un altre avantatge és que en el cas que es disposi de mesures de múltiples conjunts Z que compleixin el criteri, si no s'obtenen els mateixos resultats és un indicador que el model no és correcte.

Exemple 4.24. Com a exemple d'aplicació del criteri de la porta del darrere, considerem els dos exemples de la paradoxa de Simpson amb els diagrames causals de la Figura 2 i comprovem que arribem a les mateixes conclusions.

- Exemple 1: En el primer cas, X indica el procediment ($X = 0$ essent la cirúrgia oberta i $X = 1$ la nefrolitotomia percutània), Z la mida de les pedres ($Z = 0$ les petites i $Z = 1$ les grans) i Y el resultat de l'operació ($Y = 0$ el fracàs i $Y = 1$ l'èxit). Z bloqueja l'únic camí per la porta del darrere $X \leftarrow Z \rightarrow Y$, de manera que hem d'ajustar per Z (és a dir, cal considerar les dades estratificades): $P(y|do(x)) = \sum_z P(y|x,z)P(z)$.

  Substituint els valors de la Taula 1, calculem $P(Z = 0) = \frac{357}{700} = 0.51$, $P(Z = 1) = 0.49$ i $E[Y|do(X = 1)] = P(Y = 1|do(X = 1)) = 0.87 \cdot 0.51 + 0.69 \cdot 0.49 = 0.78$. Similarment, trobem $E[Y|do(X = 0)] = P(Y = 1|do(X = 0)) = 0.93 \cdot 0.51 + 0.73 \cdot 0.49 = 0.83$. Per tant, $\tau = -0.05 < 0$ i veiem que la cirúrgia oberta és lleugerament millor.

- Exemple 2: En el segon cas, X indica el tractament ($X = 1$ si s'administra el medicament i $X = 0$ altrament), Z la pressió ($Z = 0$ si és baixa i $Z = 1$ si és alta) i Y l'atac de cor ($Y = 0$ si el pateixen, $Y = 1$ altrament). Com que no hi ha cap camí per la porta del darrere entre X i Y, $E[Y|do(x)] = E[Y|x]$. És a dir, que hem de considerar les dades sense estratificar.

  Podem calcular l'ECM: $\tau = E[Y|X = 1] - E[Y|X = 0] = P(Y = 1|X = 1) - P(Y = 1|X = 0) = 0.82 - 0.78 = 0.04 > 0$, de manera que en efecte el medicament és beneficiós.

Tanmateix, encara hi ha casos en què el criteri de la porta del darrere no serveix per estimar l'efecte causal considerat. Un exemple és el de la Figura 9. En la figura de l'esquerra, la variable U no és mesurada o observada (diem que és una variable **latent**), fet que impedeix bloquejar el camí per la porta del darrere de X a Y. Aquests camins per la porta del darrere que no es poden bloquejar sovint es representen mitjançant **arcs bidireccionals**, com en la figura de la dreta. En casos com aquests, es pot aplicar el **criteri de la porta frontal**.

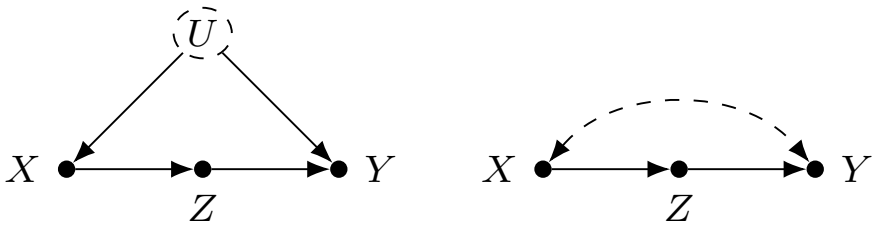

Figura 9: Exemple on no és aplicable el criteri de la porta del darrere si la variable U no és coneguda (esquerra). A la dreta, el mateix exemple amb un arc bidireccional.

Definició 4.25. *(Criteri de la porta frontal)*

*Un conjunt de variables Z satisfà el **criteri de la porta frontal** relatiu a una parella ordenada de variables (X, Y) si:*

1. *Z intercepta tots els camins dirigits de X a Y.*
2. *No hi ha cap camí per la porta del darrere desbloquejat de X a Z.*
3. *X bloqueja tots els camins per la porta del darrere de Z a Y.*

Teorema 4.26. *(Ajustament de la porta frontal) (Pearl, 2009 [24], pàg. 83, Teorema 3.3.4)*

Si Z satisfà el criteri de la porta frontal relatiu a (X, Y) i $P(z, x) > 0 \quad \forall (z, x) \in (Z, X)$, es pot identificar l'efecte causal de X en Y mitjançant la fórmula:

$$P(y|do(x)) = \sum_z P(z|x) \cdot \sum_{x'} P(y|x', z) \cdot P(x') \tag{4.9}$$

Demostració

Donarem una intuïció d'on surt la fórmula i a la següent secció la demostrarem de manera rigorosa seguint un procediment alternatiu.

Per estimar l'efecte causal de X en Y considerem tres passos:

1. Estimem l'efecte causal de X en Z: $P(x|do(z)) = P(x|z)$ ja que, pel segon punt del criteri de la porta frontal, no hi ha cap camí per la porta del darrere de X a Z desbloquejat.

2. Estimem l'efecte causal de Z en Y: per hipòtesi, X satisfà el criteri de la porta del darrere relatiu a (Z, Y), de manera que $P(y|do(z)) = \sum_{x'} P(y|z, x') \cdot P(x')$.

3. Combinem els passos anteriors per estimar l'efecte causal de X en Y:

$$P(y|do(x)) = \sum_z P(z|do(x)) \cdot P(y|do(z)) = \sum_z P(z|x) \cdot \sum_{x'} P(y|x', z) \cdot P(x')$$

□

4.3 Càlcul *do*

Els criteris de la porta frontal i del darrere permeten el procés d'identificació en bastants situacions. No obstant, estan limitats a intervencions en una única variable i en ocasions ens trobem que no podem aplicar-los, però l'efecte causal sí és identificable. En aquests casos, podem aplicar unes regles de càlcul dels operadors *do* que ens permeten identificar *tots* els estimadors causals identificables.

Notació 8. *Donat un GAD G i uns nodes X i Y, ens referim amb $G_{\overline{X}}$ al graf fruit d'eliminar les arestes entrants a X (en altres paraules, el graf manipulat G_m per una intervenció $do(X = x)$) i amb $G_{\underline{X}}$ al graf després d'eliminar les arestes sortints de X. Per referir-nos a aquestes manipulacions amb més d'un node escrivim expressions del tipus $G_{\overline{X}\underline{Y}}$.*

Teorema 4.27. *(Les regles del càlcul do) (Pearl, 1995 [22], Teorema 4.1)*

Donat un diagrama causal G amb distribució de probabilitat P, donats subconjunts disjunts qualssevol de variables X, Y, Z i W se satisfan les següents regles:

***Regla 1** (Inserció/Eliminació d'observacions):*

$$P(y|do(x), z, w) = P(y|do(x), w) \quad si \quad ((Y \perp\!\!\!\perp Z)|\{X, W\})_{G_{\overline{X}}}$$

***Regla 2** (Intercanvi d'intervencions/observacions):*

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad si \quad ((Y \perp\!\!\!\perp Z)|\{X, W\})_{G_{\overline{X}\underline{Z}}}$$

***Regla 3** (Inserció/eliminació d'intervencions):*

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad si \quad ((Y \perp\!\!\!\perp Z)|\{X, W\})_{G_{\overline{X}\overline{Z(W)}}}$$

on $Z(W)$ és el conjunt de nodes dins de Z que no són ancestres de cap node de W en $G_{\overline{X}}$.

Observació 4.28. La primera regla és una generalització de la d-separació en grafs manipulats per intervencions. La segona és una generalització del criteri de la porta del darrere. En el graf $G_{\overline{X}\underline{Z}}$, els únics camins que poden connectar Z i Y són per la porta del darrere (perquè eliminem les arestes sortints de Z). Si aquests camins estan tots bloquejats condicionant a X, W aleshores podem substituir la intervenció en Z per una observació (és a dir, per un condicionament). Finalment, la tercera regla explicita sota quines condicions es pot introduir o eliminar una intervenció sense afectar la probabilitat de $Y = y$.

Teorema 4.29. *(Completesa del càlcul do)(Shpitser i Pearl, 2006b [32], Teorema 7)*

Les regles del càlcul do, juntament amb manipulacions estàndards de probabilitats, són completes per identificar tots els efectes causals identificables.

Com a aplicació d'aquestes regles, demostrem la fórmula d'ajustament de la porta frontal:

Demostració de l'ajustament de la porta frontal:

Seguim els següents passos:

- Marginalitzem en Z: $P(y|do(x)) = \sum_z P(y|z, do(x)) \cdot P(z|do(x))$
- Apliquem la regla 2 al terme $P(z|do(x))$. Tenim $W = \varnothing, Y = Z, X = \varnothing, Z = X$. Se satisfà $Z \perp\!\!\!\perp X$ en $G_{\underline{X}}$, ja que per hipòtesi del criteri de la porta frontal no hi ha camins per la porta del darrere desbloquejats de X a Z, i en eliminar les arestes sortints de X tampoc hi ha camins dirigits de X a Z (camins dirigits de Z a X no n'hi ha perquè els Z són descendents de X). Així, per la regla 2 $P(z|do(x)) = P(z|x)$
- Per la regla 2, $P(y|z, do(x)) = P(y|do(z), do(x))$. Llevat de $W = \varnothing$, la resta de variables estan definides com en la regla 2. Se satisfà $(Y \perp\!\!\!\perp Z)|X$ en $G_{\overline{X}\underline{Z}}$, ja que eliminem les arestes sortints de Z i per tant no hi ha camins dirigits de Z a Y (camins dirigits en el sentit contrari no n'hi ha ja que els elements de Z són ancestres de Y per hipòtesi) i pel criteri de la porta frontal, X bloqueja tots els camins per la porta del darrere de Z a Y.
- Per la regla 3, $P(y|do(z), do(x)) = P(y|do(z))$. En aquest cas, $Y = Y, X = Z, Z = X$ i $W = \varnothing$, de manera que $Z(W) = X$. Es verifica $(Y \perp\!\!\!\perp X)|Z$ en $G_{\overline{ZX}}$, ja que com que eliminem les arestes entrants en X no hi ha camins per la porta del darrere de X a Y, i els camins dirigits de X a Y els eliminem en excloure les arestes entrants a Z (els elements de Z intercepten tots els camins dirigits de X a Y). A més, no es pot donar el cas que un element de Z sigui un col·lisionador en un camí de X a Y que es desbloquegi condicionant-hi, perquè eliminem les arestes entrants a Z.
- Marginalitzem en X: $P(y|do(z)) = \sum_{x'} P(y|x', do(z)) \cdot P(x'|do(z))$
- Apliquem la regla 2: $P(y|x', do(z)) = P(y|x', z)$. En aquest cas $X = \varnothing, Y = Y, Z = Z, W = X$ i es verifica $(Y \perp\!\!\!\perp Z)|X$ en $G_{\underline{Z}}$ ja que eliminem els camins dirigits de Z a Y en eliminar les arestes sortints de Z, i per hipòtesi X bloqueja tots els camins per la porta del darrere de Z a Y.
- Apliquem la regla 3: $P(x'|do(z)) = P(x')$. Tenim $Y = X, X = \varnothing, W = \varnothing, Z = Z = Z(W)$. Es compleix $(X \perp\!\!\!\perp Z)$ en $G_{\overline{Z}}$ ja que en eliminar les arestes entrants a Z no hi ha camins dirigits de X a Z, i per hipòtesi del criteri de la porta frontal, no hi ha camins per la porta del darrere de X a Z desbloquejats.

Encadenant totes aquestes transformacions arribem a la fórmula d'ajustament:

$$\begin{aligned} P(y|do(x)) &= \sum_z P(y|z, do(x)) \cdot P(z|do(x)) = \sum_z P(y|do(z), do(x)) \cdot P(z|x) = \\ &= \sum_z P(y|do(z)) \cdot P(z|x) = \sum_z P(z|x) \cdot \sum_{x'} P(y|x', do(z)) \cdot P(x'|do(z)) = \\ &= \sum_z P(z|x) \cdot \sum_{x'} P(y|x', z) \cdot P(x') \end{aligned}$$

□

4.4 Criteris gràfics d'identificabilitat

Convindria tenir algun criteri gràfic que ens assegurés la identificabilitat abans d'intentar aplicar el càlcul *do*. Els criteris de la porta del darrera i frontal en són exemples, però en podem trobar un de més general.

Teorema 4.30. *(Tian i Pearl, 2002 [37], Teorema 4)*

Sigui G un diagrama causal, sigui X la variable intervinguda i sigui Y un conjunt de variables que no inclou X. $P(y|do(x))$ és identificable si no existeix cap camí compost únicament per arcs bidireccionals de X als fills de X ancestres de Y.

Observació 4.31. El criteri anterior generalitza els criteris de la porta del darrere i frontal, però també inclou casos on no es poden aplicar com a la Figura 10 (esquerra).

Observació 4.32. El criteri és una condició suficient d'identificabilitat, però no és necessària. El graf de la Figura 10 (dreta) no el satisfà, ja que hi ha un camí de X a Y constituït únicament per arcs bidireccionals. Per contra, sí es pot identificar l'efecte causal de X en Y mitjançant el càlcul *do*.

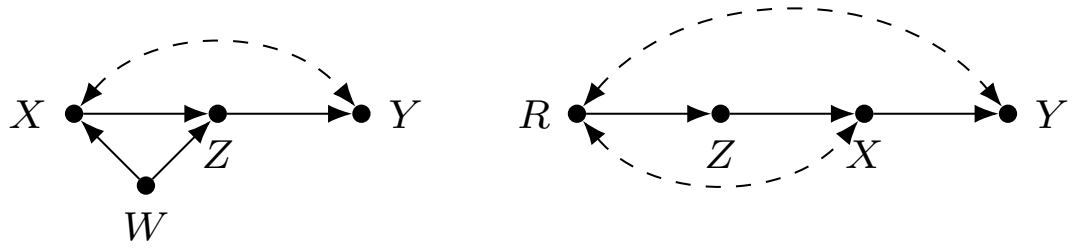

Figura 10: Exemples de diagrames causals. A la figura de l'esquerra és aplicable el criteri gràfic d'identificabilitat donat, mentre que a la de la dreta no es compleix el criteri però sí és identificable.

Una condició necessària (però no suficient) d'identificabilitat és que es puguin bloquejar tots els camins per la porta del darrere de X als seus fills que siguin ancestres de Y (Pearl, 2009 [24], pàg. 92).

4.5 Algorisme per identificar efectes causals

El problema del càlcul *do* és que pot ser complicat trobar l'ordre en el qual aplicar les diferents regles per arribar a un estimador lliure d'operadors *do*, sobretot quan hi ha un nombre elevat de variables i el diagrama causal és complex. Així mateix, no tenim criteris gràfics equivalents a la identificabilitat i per tant podríem estar intentant resoldre problemes sense solució sense saber-ho.

Per aquest motiu, ens interessaria trobar algorismes per identificar efectes causals de manera sistemàtica. En aquesta secció presentarem els algorismes **ID** (Shpitser i Pearl, 2006b [32]) i **IDC** (Shpitser i Pearl, 2006a [31]) elaborats per Shpitser i Pearl. El primer serveix per identificar efectes causals del tipus $P(y|do(x))$, mentre que el segon es

fonamenta en el primer per identificar efectes z-específics $P(y|do(x), z)$. Si els estimadors causals considerats no són identificables, els algorismes fallen. Els dos estan implementats en el paquet de R **causaleffect** (Tikka i Karvanen, 2017 [38]).

Abans d'introduir els algorismes calen un seguit de definicions:

Definició 4.33. *(Subgraf)*

*Sigui $G = \{V, E\}$ un graf. Donat un subconjunt $W \subset V$, definim $G[W]$ com el **subgraf** de G de vèrtexs W i d'arestes les de E que connecten els elements de W.*

Definició 4.34. *(C-component)*

*Sigui $G = \{V, E\}$ un GAD, on V són només les variables observades i les arestes E inclouen els arcs bidireccionals. Si existeix un subconjunt $B \subset E$ que conté només arcs bidireccionals de manera que el graf $\{V, B\}$ sigui connex, aleshores G és una **C-component**.*

Observació 4.35. Encara que un GAD no sigui una C-component, sempre podem trobar subgrafs que ho siguin, ja que per conveni tots els subgrafs induïts per un sol node són C-components.

Definició 4.36. *(C-component maximal)*

*Sigui G un GAD i sigui $C = \{V, E\} \subset G$ una C-component. Diem que C és **maximal** si qualsevol camí d'arcs bidireccionals H de G que conté almenys un node de V satisfà $H \subset C$.*

Observació 4.37. La definició implica que existeix un camí d'arcs bidireccionals entre dos nodes si i només si estan en la mateixa C-component maximal. Per tant, els camins d'arcs bidireccionals defineixen el conjunt de C-components maximals, que constitueix una partició de G. El lema següent és vàlid per totes les particions de C-components, però en particular l'utilitzem per les maximals:

Lema 4.38. *(Tian, 2002 [36], pàg. 56, Corol·lari 1)*

Sigui $G = \{V, E\}$ un GAD, i sigui $C(G) = \{G[S_1], ..., G[S_k]\}$ una partició de C-components de G. Aleshores,

(i) $P(v)$ factoritza com:

$$P(v) = \prod_{i=1}^{k} P(s_i|do(v \setminus s_i)) \tag{4.10}$$

(ii) Sigui $V = \{V_1, ..., V_n\}$ ordenat topològicament, i sigui $V_G^{(i)} = \{V_1, ..., V_i\}$ per $i \in \{1, ..., n\}$ i $V_G^{(0)} = \varnothing$. Aleshores, cadascun dels factors anteriors són identificables com:

$$P(s_j|do(v \setminus s_j)) = \prod_{V_i \in S_j} P(v_i|v_G^{(i-1)}) \tag{4.11}$$

Exemple 4.39. La factorització en C-components maximals dels diagrames de la Figura 10 és $\{G[X, Y], G[Z], G[W]\}$ pel diagrama de l'esquerra i $\{G[X, Y, R], G[Z]\}$ pel de la dreta. Aplicant la factorització de l'equació (4.10), pel segon diagrama:

$$P(x, y, z, r) = P(x, y, r|do(z))P(z|do(x), do(y), do(r))$$

Definició 4.40. *(C-arbre)*

Sigui G una C-component en la qual cada node observat té com a màxim un fill. Si existeix un node Y descendent de tots els altres, diem que G és un ***C-arbre d'arrel*** *Y.*

Amb aquestes definicions tenim un nou criteri de no-identificabilitat en efectes en una sola variable:

Teorema 4.41. *(Shpitser i Pearl, 2006b [32], Teorema 3)*

Sigui G un C-arbre d'arrel Y, aleshores l'efecte de qualsevol conjunt de nodes de G en Y no és identificable.

Generalitzant-ho a més variables, tenim la definició de **C-bosc**:

Definició 4.42. *(C-bosc)*

*Sigui G un GAD i sigui Y el conjunt de nodes sense descendents no trivials (que anomenem conjunt d'****arrels****). Si G és una C-component i tots els nodes observats tenen com a màxim un fill, aleshores G és un* ***C-bosc d'arrels*** *Y.*

L'última definició que ens fa falta és la d'**arbust**, que juga un paper clau en la no-identificabilitat.

Definició 4.43. *(Arbust)*

Sigui $G = \{V, E\}$ un GAD i siguin X, Y dos subconjunts disjunts de V. Si existeixen dos C-boscos $F = \{V_F, E_F\}, F' = \{V_{F'}, E_{F'}\}$ d'arrels R continguts en G, tals que $V_F \cap X \neq \varnothing$, $V_{F'} \cap X = \varnothing$, $F' \subset F$ i R és un subconjunt de $An(Y)_{G_{\overline{X}}}$ (els ancestres de Y en $G_{\overline{X}}$), aleshores F i F' formen un ***arbust*** *per $P(y|do(x))$ en G.*

Teorema 4.44. *(Criteri de l'arbust) (Shpitser i Pearl, 2006a [31], Teorema 3)*

Sigui G un GAD, P la distribució de probabilitat de G i siguin X, Y dos subconjunts disjunts de nodes. $P(y|do(x))$ és identificable si, i només si, no existeix cap arbust per $P(y'|do(x'))$ en G per qualssevol $X' \subset X$ i $Y' \subset Y$.

Observació 4.45. En el cas particular en el qual G és un C-arbre d'arrel Y, prenent com F el C-arbre i com F' el node Y aïllat, tenim que per qualsevol conjunt de nodes X que no contingui Y, F i F' formen un arbust per $P(y|do(x))$ en G. Aleshores, pel criteri de l'arbust $P(y|do(x))$ no és identificable. Així, el criteri de l'arbust generalitza el criteri d'identificabilitat pels C-arbres.

Exemple 4.46. Considerem el diagrama causal G de la Figura 11, on el conjunt $Y = \{Y_1, Y_2\}$ constitueix els nodes sense descendents no trivials. Prenent com F el graf G llevat de l'aresta que uneix X i W_1 i $F' = F \setminus \{X\}$, F i F' formen un arbust per $P(y_1, y_2|do(x))$ en G. Així, $P(y_1, y_2|do(x))$ no és identificable.

Amb totes les definicions establertes, podem donar l'algorisme ID que trobem a la Figura 12. ID retorna una expressió per l'efecte causal $P(y|do(x))$ sempre que sigui identificable, on X i Y són conjunts disjunts de variables. En el cas en què l'algorisme sigui interromput en la línia 5, resulta que $P(y|do(x))$ no és identificable.

A continuació, analitzem les línies de l'algorisme:

1. En el cas en el qual $X = \varnothing$, es retorna la distribució marginal $P(y)$, marginalitzant $P(v)$ en totes les variables que no estan en Y.

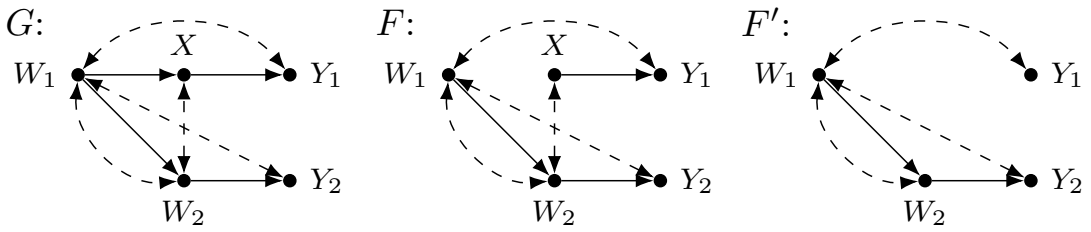

Figura 11: Exemple de diagrama causal amb un arbust.

funció **ID**(y, x, P, G)
ENTRADA: Assignació de valors x, y, una distribució de probabilitat conjunta $P(v)$ i un diagrama causal $G = \{V, E\}$
SORTIDA: Expressió de $P(y|do(x))$ en termes de $P(v)$ o **ERROR**(F, F')

1. si $x = \varnothing$ retorna $\sum_{v \setminus y} P(v)$
2. si $V \neq An(Y)_G$, aleshores
   retorna **ID**(y, $x \cap An(Y)_G$, $\sum_{v \setminus An(Y)_G} P$, $G[An(Y)_G]$)
3. Sigui $W = (V \setminus X) \setminus An(Y)_{G_{\overline{X}}}$
   si $W \neq \varnothing$ retorna **ID**(y, $x \cup w$, P, G)
4. si $C(G[V \setminus X]) = \{G[S_1], ..., G[S_k]\}$, aleshores
   retorna $\sum_{v \setminus (y \cup x)} \prod_i$**ID**($s_i$, $v \setminus s_i$, P, G)
   si $C(G[V \setminus X]) = \{G[S]\}$, aleshores
5. si $C(G) = \{G\}$, aleshores
   retorna **ERROR**(G, G[**S**])
6. si G[S]$\in C(G)$, aleshores
   retorna $\sum_{s \setminus y} \prod_{V_i \in S} P(v_i | v_G^{(i-1)})$
7. si $(\exists S')S \subset S'$ tal que $G[S'] \in C(G)$, aleshores
   retorna **ID**(y, $x \cap s'$, $\prod_{V_i \in S'} P(V_i | V_G^{(i-1)} \cap S', v_G^{(i-1)} \setminus s')$, $G[S']$)

Figura 12: Algorisme per identificar $P(y|do(x))$

2. En la segona línia, s'eliminen totes les variables de X que no són ancestres de Y. Com que no hi ha cap camí dirigit d'aquestes variables cap a Y, són irrellevants en el càlcul d'efectes causals en Y.

3. La tercera línia és una aplicació de la regla 3 del càlcul *do*: $P(y|do(x)) = P(y|do(x), do(w))$ ja que $((Y \perp\!\!\!\perp W)|X)_{G_{\overline{XW}}}$ (W no conté ancestres de Y en $G_{\overline{X}}$ per construcció, de manera que no hi ha camins dirigits de W a Y, i en eliminar les arestes entrants a W en $G_{\overline{XW}}$ tampoc poden haver-hi camins per la porta del darrere de W a Y).

4. A la quarta línia descompon el subgraf de G sense els nodes de X en les seves C-components maximals. En $G_{\overline{X}}$ eliminem els arcs bidireccionals de X, per tant

té tots els nodes de X com a C-components maximals, mentre que la resta de les C-components maximals de la partició són les mateixes que en $G[V \setminus X]$. Suposant que $C(G[V \setminus X]) = \{G[S_1], ..., G[S_k]\}$ amb $k > 1$, aplicant la descomposició donada per l'equació (4.10) al model de la intervenció $do(X = x)$,

$$P(v|do(x)) = \prod_{i=1}^{k} P(s_i|do(v \setminus s_i))$$

on obviem els termes de les C-components dels nodes individuals de $X_j \in X$ ja que $P(x_j|do(v \setminus x_j)) = 1$ suposant x_j consistent amb la intervenció en X. Finalment, sumant aquesta factorització per totes les variables llevat de $X \cup Y$ tenim $P(y|do(x))$.

Per contra, si només hi ha una C-component $G[S]$ a $G[V \setminus X]$, s'executen les línies 5,6 o 7.

5. Si s'executa la línia 5, el graf G (que serà un subgraf del diagrama de l'input inicial) és una C-component. Eliminant unes quantes arestes dirigides de manera que cada node tingui com a molt un fill i sense alterar el conjunt d'arrels R, el graf resultant F és un C-bosc d'arrels R. A més, per construcció $R \subset An(Y)_{G_{\overline{X}}}$ degut a les línies 2 i 3. De manera similar, prenem $F' = F \cap G[S]$, que també és un C-bosc d'arrels R. Com que $F' \subset F$, $V_F \cap X \neq \varnothing$ i $V_{F'} \cap X = \varnothing$, tenim un arbust per X i Y que són subconjunts dels conjunts X i Y de l'input inicial. Per tant, pel criteri de l'arbust $P(y|do(x))$ no és identificable.

6. Si $G[S]$ és una C-component de G (és a dir, el cas en el qual no hi ha arcs bidireccionals entre S i X), s'aplica l'equació (4.11) i se suma per les variables de $S \setminus Y$.

7. L'última línia ocorre quan hi ha arcs bidireccionals entre S i X, aleshores $G[S]$ ha de ser un subgraf d'alguna C-component maximal de G. El resultat que retorna la línia es basa en el següent lema:

   Lema 4.47. *(Shpitser i Pearl, 2006b [32], Lema 6)*

   Siguin X, Y conjunts de variables en un diagrama causal G. Sempre que se satisfan les condicions de la línia 7, $P(y|do(x))$ es pot calcular en G a partir de P si i només si $P(y|do(x \cap S'))$ es pot calcular en $G[S']$ a partir de

$$P' = \prod_{V_i \in S'} P(V_i|V_G^{(i-1)} \cap S', v_G^{(i-1)} \setminus s')$$

A partir de ID es pot construir un segon algorisme IDC per identificar efectes z-específics, és a dir, expressions tipus $P(y|do(x), z)$.

A la primera línia de l'algorisme de la Figura 13 aplica la regla 2 del càldcul *do*, intercanviant observacions de Z per intervencions. Quan arriba un punt en el qual ja no és possible continuar aplicant-la, aleshores s'expressa $P(y|do(x), z)$ com $\frac{P(y,z|do(x))}{P(z|do(x))}$. L'expressió del numerador es pot calcular amb ID, mentre que la del denominador es pot trobar sumant el numerador per totes les variables de Y.

Exemple 4.48. Per il·lustrar el seu funcionament, implementem ID al segon diagrama de la Figura 10.

$V = \{R, Z, X, Y\}$ (ja està ordenat topològicament) i volem trobar $P(y|do(x))$. Seguim els següents passos:

funció **IDC**(y, x, z,P, G)
ENTRADA: Assignació de valors x, y,z, una distribució de probabilitat conjunta $P(v)$ i un diagrama causal $G = \{V, E\}$
SORTIDA: Expressió de $P(y|do(x), z)$ en termes de $P(v)$ o **ERROR**(F, F')

1. si $\exists Z' \in Z$ tal que $(Y \perp\!\!\!\perp Z|X, Z \setminus \{Z'\})_{G_{\overline{X}\underline{Z}}}$, aleshores
   retorna **IDC**(y,$x \cup \{z'\}$,$z \setminus \{z'\}$, P, G)
2. si no, sigui $P' =$**ID**$(y \cup z,\ x,\ P,\ G)$
   retorna $P'/\sum_y P'$

Figura 13: Algorisme per identificar $P(y|do(x), z)$

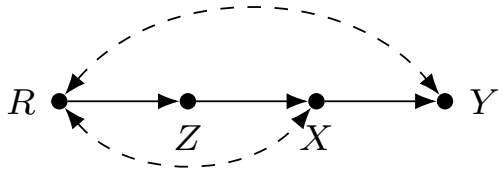

Figura 14: Diagrama causal G on aplicarem l'algorisme ID.

- Com que $X \neq \varnothing$ i $V = An(Y)_G$, saltem directament a la línia 3. Trobem $W = (V \setminus X) \setminus An(Y)_{G_{\overline{X}}} = \{R, Z, Y\} \setminus \{Y\} = \{R, Z\} \neq \varnothing$. Per tant, tornem a cridar la funció ID per trobar $P(y|do(x), do(r), do(z))$.

- Saltem a la línia 4 i trobem la descomposició en C-components maximals de $G[V \setminus \{R, Z, X\}] = G[Y]$. Com que és el node aïllat Y, $C(G[Y]) = \{G[Y]\}$ té una única C-component. Per tant, anirem a les línies 5, 6 o 7 depenent de la descomposició de C-components maximals de G.

- Trobem que $C(G) = \{G[R, X, Y], G[Z]\}$. Com que $G[Y]$ és un subgraf de la C-component $G[R, X, Y]$, estem en la situació de la línia 7, amb $S' = \{R, X, Y\}$. Per tant, hem de cridar ID de nou prenent $Y' = Y$, $X' = \{R, Z, X\} \cap S' = \{R, X\}$, $G' = G[S']$ i $P'(R, X, Y) = P(R)P(X|R, z)P(Y|X, R, z)$.

- En el nou graf G', R no és ancestre de Y. Per tant, implementem la línia 2 i cridem ID per $Y'' = Y$, $X'' = X' \cap An(Y)_{G'} = X$, $G'' = G'[An(Y)_{G'}]$ i $P''(X, Y) = \sum_r P'(r, X, Y)$.

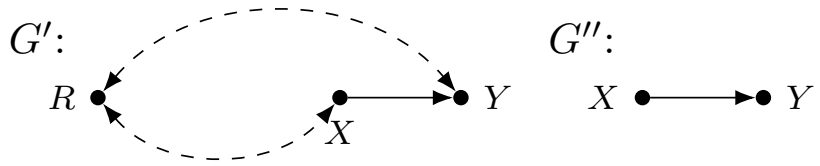

Figura 15: Diagrames G' i G'' dels passos intermitjos en la implementació de ID.

- Saltem a la línia 4 i trobem la descomposició en C-components maximals del subgraf de G'' sense X''. De nou tenim el node Y aïllat, de manera que té una única C-

component $G''[Y]$. En aquest cas, $C(G'') = \{G''[X], G''[Y]\}$ i per tant estem en la situació de la línia 6 amb $S = Y$. Aleshores, l'algorisme retorna $P''(y|x) = \frac{P''(x,y)}{P''(x)} = \frac{P''(x,y)}{\sum_y P''(x,y)}$.

- Per acabar, expressem P'' en termes de P', i P' en termes de la P original i trobem $P(y|do(x))$:

$$P(y|do(x)) = \frac{P''(x,y)}{\sum_y P''(x,y)} = \frac{\sum_r P'(r,x,y)}{\sum_y P'(r,x,y)} = \frac{\sum_r P(r)P(x|r,z)P(y|x,r,z)}{\sum_{r,y} P(r)P(x|r,z)P(y|X,r,z)} = \\ = \frac{\sum_r P(r)P(x|r,z)P(y|x,r,z)}{\sum_r P(r)P(x|r,z)}$$

5 Mediació

Fins ara hem estudiat com calcular l'efecte causal *total* mig τ d'una variable tractament X en un conjunt resultat Y. No obstant, no hem determinat quina fracció d'aquest efecte és directa i quina és indirecta, és a dir, mediada per altres variables. Aquesta mena d'estudis es coneixen com estudis de **mediació**.

Un exemple on és rellevant són els casos de discriminació segons la condició d'una persona (per exemple, segons gènere o raça). En aquests casos, per demostrar la discriminació el que cal trobar és l'efecte directe.

5.1 Efectes directes

Per aïllar l'efecte directe, una primera idea intuïtiva és bloquejar tots els camins indirectes condicionant a les variables mediadores (recordem que els camins indirectes són cadenes del tipus $X \to M \to Y$, de manera que si condicionem a M queda bloquejat). Aquest és un error comú conegut com la **fal·làcia de la mediació**.

Per il·lustrar-ho, recordem el model de la Figura 2 del fàrmac per prevenir atacs de cor. La pressió arterial és la variable mediadora entre el medicament i els atacs cardíacs. Si volem trobar l'efecte directe "Tractament→ Atac de cor", podem condicionar en la pressió arterial i bloquejar el camí indirecte.

Imaginem ara que afegim una variable Edat al model, la qual és una causa directa de la pressió i dels atacs de cor (a més edat, més risc hi ha de tenir ambdues patologies). Aleshores, condicionant a la pressió en aquest segon model es desbloquejaria el camí "Tractament → Pressió ← Edat → Atac de cor" en condicionar a un col·lisionador. Aquest camí indueix una associació no causal entre el tractament i el resultat, de manera que ja no obtenim l'efecte directe.

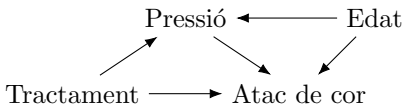

Figura 16: Model de la Figura 2 amb la variable Edat.

Per tant, condicionar al mediador no serveix en els casos en què hi hagi camins per la porta del darrere entre el mediador i el resultat. La solució és intervenir també en el mediador, deixant-lo en un valor fixat. Així, com a una primera mesura de l'efecte directe podem definir l'**Efecte Directe Controlat** com:

Definició 5.1. *(Efecte Directe Controlat)*

Sigui X una variable binària, Y un conjunt de variables resultat i M un conjunt de variables mediadores entre X i Y. Per cada valor de $M = m$, definim l'Efecte Directe Controlat (EDC) com:

$$EDC(m) = E[Y|do(X = 1), do(M = m)] - E[Y|do(X = 0), do(M = m)] \tag{5.1}$$

Observació 5.2. L'EDC es pot calcular amb les tècniques d'identificació d'intervencions descrites en la secció anterior. També notem que depèn del valor en què fixem els mediadors, de manera que per tenir una millor idea de l'efecte directe hauríem de calcular-lo per tots els valors rellevants dels mediadors.

El problema de l'EDC, apart de la dependència en els valors dels mediadors, és que no és realista forçar el mateix valor dels mediadors a totes les unitats. Per això, convé una definició alternativa d'efecte directe que permeti la variació natural dels mediadors en les unitats. Definim així l'**Efecte Directe Natural** (EDN) com la variació esperada en el resultat Y en variar el tractament de $X = 0$ a $X = 1$, mentre els mediadors es fixen en els valors que *haguessin pres* per cada unitat abans del canvi (és a dir, en $X = 0$). A diferència de τ i de l'EDC, la definició de l'EDN és contrafactual, és a dir, del tercer nivell de l'Escala de Causalitat. Formalment, en termes de la notació de respostes potencials es defineix de la següent manera:

Definició 5.3. *(Efecte Directe Natural)*

Sigui X una variable binària, Y un conjunt de variables resultat i M un conjunt de variables mediadores entre X i Y. Definim l'Efecte Directe Natural (EDN) de X en Y com:

$$EDN = E[Y_{M_0}|do(X = 1)] - E[Y_{M_0}|do(X = 0)] = E[Y_{1,M_0} - Y_{0,M_0}] \tag{5.2}$$

Observació 5.4. Recordant la notació de respostes de potencials, M_0 és la resposta potencial dels mediadors sota l'acció $do(X = 0)$. Considerant $X = 0$ com l'absència del tractament, podem interpretar M_0 com el valor "natural" de M. De manera similar, Y_{M_0} és la resposta potencial de Y sota el valor natural de M.

El terme contrafactual és el primer, ja que conté dos mons diferents: un de real on fixem $X = 1$ i un de fictici en contradicció amb el real, on M pren el valor que tindria si fixéssim $X = 0$. Pel que fa al segon terme, per la llei de consistència el podem expressar simplement com $E[Y_{M_0}|do(X = 0)] = E[Y|do(X = 0)] = E[Y_0]$.

5.2 Efectes indirectes

Respecte als efectes indirectes, una primera intuïció seria descompondre l'efecte total com

$$\text{efecte total} = \text{efecte directe} + \text{efecte indirecte} \tag{5.3}$$

i restar-li l'efecte directe. Aquesta aproximació, però, no és correcta en models que involucren termes d'interacció entre el tractament i els mediadors. Per exemple, considerem el cas d'un medicament emprat per tractar una determinada malaltia, que en ser administrat allibera un enzim en l'individu actuant com un catalitzador. Aleshores, l'enzim "activa" el medicament per poder curar la malaltia.

En aquest context, l'efecte total és clarament positiu. Per contra, l'efecte directe és nul, perquè si d'alguna manera impedim externament que s'activi l'enzim, el medicament per si sol no té cap efecte en la malaltia. Similarment, l'efecte indirecte també és nul, ja que si injectem l'enzim als participants externament sense administrar el medicament, l'enzim sol no pot curar la malaltia.

Per tant, la descomposició de (5.3) fora de models sense interacció (per exemple: models amb totes les equacions estructurals lineals) no serveix, de manera que cal donar una definició de l'efecte indirecte independent dels efectes totals i directes. No hi ha cap definició "controlada" en tant que no és possible desactivar els camins directes. No obstant, sí podem donar una definició d'**Efecte Indirecte Natural** (EIN) de manera anàloga a la dels efectes directes. Definim l'EIN com la variació esperada en el resultat Y quan el tractament X es manté constant en $X = 0$, i els mediadors M es fixen als valors que *haguessin pres* en variar X. De nou, és un terme contrafactual que es defineix formalment com:

Definició 5.5. *(Efecte Indirecte Natural)*

Sigui X una variable binària, Y un conjunt de variables resultat i M un conjunt de variables mediadores entre X i Y. Definim l'Efecte Indirecte Natural (EIN) de X en Y com:

$$EIN = E[Y_{M_1}|do(X=0)] - E[Y_{M_0}|do(X=0)] = E[Y_{0,M_1} - Y_{0,M_0}] \tag{5.4}$$

Observació 5.6. Com en la definició de l'EDN, el terme contrafactual és el primer, mentre que el segon terme és simplement $E[Y_0]$.

Amb les definicions dels efectes naturals, podem veure la seva relació amb l'efecte total:

Proposició 5.7. *(Pearl 2014, [25], equació (12))*

L'efecte total de una variable X binària en Y es descompon com:

$$\tau = EDN - EIN_r \tag{5.5}$$

on EIN_r representa l'efecte indirecte natural de la variació inversa del tractament (és a dir, de $X=1$ a $X=0$).

Demostració:

Canviant $X=1$ per $X=0$ i viceversa en la definició de EIN, trobem l'expressió per la transició inversa:

$$EIN_r = E[Y_{1,M_0} - Y_{1,M_1}] = E[Y_{1,M_0} - Y_1]$$

Aleshores, substituint les definicions demostrem la igualtat amb l'efecte total:

$$EDN - EIN_r = E[Y_{1,M_0} - Y_0] - E[Y_{1,M_0} - Y_1] = E[Y_1 - Y_0] = \tau$$

□

Observació 5.8. De l'equació (5.5) podem deduir que si τ i l'EDN són identificables, l'EIN també ho és (concretament, l'EIN de la transició inversa és identificable, però aleshores el de la transició directa també ho serà).

5.3 Identificació de la mediació

Tant τ com l'EDC són estimadors corresponents al segon nivell de l'Escala de Causalitat, de manera que ja sabem com calcular-los sempre que sigui possible (per exemple amb el càlcul *do* o bé realitzant els experiments quan sigui factible). El problema que queda pendent de resoldre és el dels efectes naturals, que contenen termes contrafactuals. El nostre objectiu és convertir aquestes expressions del tercer nivell en expressions del segon nivell que sabem tractar. Amb aquesta finalitat, existeixen les següents condicions gràfiques suficients per identificar els efectes naturals de tractament X en resultat Y mediats per M:

Condicions suficients d'identificació d'efectes naturals

Existeix un conjunt de variables del model W mesurades que satisfan:

1. Cap element de W és descendent del tractament X.

2. W bloqueja tots els camins per la porta del darrere de M a Y que no passen per X.

3. L'efecte w-específic de X en M és identificable.

4. L'efecte w-específic de $\{X, M\}$ en Y és identificable.

Teorema 5.9. *(Identificació de l'EDN)(Pearl, 2014 [25], Teorema 1)*

Quan es compleixen les condicions 1 i 2, l'efecte directe natural és identificable experimentalment i és donat per:

$$EDN = \sum_m \sum_w [E_1 - E_0] \cdot P(m|do(X = 0), w) \cdot P(w) \tag{5.6}$$

on $E_x = E[Y|do(X = x, M = m), W = w]$ *per abreviar. Si a més se satisfan les condicions 3 i 4, totes les expressions amb operadors do de (5.6) són identificables.*

Corołari 5.10. *(Identificació de l'EIN)*

Quan es compleixen les condicions 1 i 2, l'efecte indirecte natural és identificable experimentalment i és donat per:

$$EIN = \sum_m \sum_w [E_0 - E_1] \cdot P(m|do(X = 1), w) \cdot P(w) + \tau \tag{5.7}$$

Si a més se satisfan les condicions 3 i 4, totes les expressions amb operadors do de (5.7) són identificables.

Demostració:

Per (5.5), $EIN = EDN_r - \tau_r$. El terme EDN_r és identificable experimentalment ja que es compleixen les condicions 1 i 2 per hipòtesi. Permutant $X = 0$ i $X = 1$ de l'equació (5.6) s'obté la seva expressió. El terme $-\tau_r = -(E[Y|do(X = 0)] - E[Y|do(X = 1)]) = \tau$, que és identificable experimentalment.

Si a més es compleixen les condicions 3 i 4, les expressions amb operadors *do* de EDN_r són identificables. Queda veure que τ sigui identificable, que es pot deduir demostrant que $P(y|do(x))$ ho és:

$$\begin{aligned} P(y|do(x)) &= \sum_w P(y, w|do(x)) = \sum_w P(y|w, do(x)) \cdot P(w|do(x)) = \\ &= \sum_m \sum_w P(y, m|w, do(x)) \cdot P(w|do(x)) = \\ &= \sum_m \sum_w P(y|m, w, do(x)) \cdot P(m|w, do(x)) \cdot P(w|do(x)) \end{aligned}$$

En la darrera expressió, la condició 1 implica que $P(w|do(x)) = P(w)$, ja que W no conté descendents de X i per tant no és afectat causalment per X. Així mateix, $P(m|w, do(x))$ és identificable per la condició 3. Per acabar, queda el terme $P(y|m, w, do(x))$. Si veiem que $P(y|m, w, do(x)) = P(y|w, do(m), do(x))$, per la condició 4 haurem demostrat que és identificable. Hem de veure que sigui possible aplicar la segona regla del càlcul *do*: cal comprovar que $(Y \perp\!\!\!\perp M|X, W)_{G_{\overline{X}\underline{M}}}$.

En $G_{\overline{X}\underline{M}}$, els únics camins que uneixen Y i M són per la porta del darrere, perquè hem eliminat les arestes sortints de M. Condicionant a W en aquest graf bloquegem tots

els camins de M a Y per la porta del darrere que no passen per X per la condició 2. Els que sí passen per X els bloquegem condicionant a X (com que eliminem les arestes entrants en X, no estarem condicionant a cap col·lisionador). Per tant, podem aplicar la segona regla de càlcul *do* i $P(y|m, w, do(x))$ és identificable. □

Exemple 5.11. En l'exemple de la Figura 16, prenent com X el tractament, Y els atacs de cor, M la pressió i E l'edat, trobem que E satisfà les quatre condicions:

1. E no és descendent de X.
2. E bloqueja tots els camins per la porta del darrere de M a Y que no passen per X.
3. Com que X no té arestes entrants, intervenir en X equival a condicionar i per tant $P(m|do(x), e) = P(m|x, e)$ i l'efecte e-específic de X en M és identificable (ho podem veure també com una aplicació de la segona regla del càlcul *do*).
4. Com que $((Y \perp\!\!\!\perp \{X, M\})|E)_{G_{\underline{XM}}}$, per la segona regla del càlcul *do*, $P(y|do(x), do(m), e) = P(y|x, m, e)$ i l'efecte e-específic de $\{X, M\}$ en Y és identificable.

Corołari 5.12. *(Fórmules de Mediació)(Pearl, 2014 [25], Corol·lari 1)*

En el cas particular en el qual les condicions 1 i 2 es compleixen per un conjunt W que bloqueja els camins per la porta del darrere X a M i de $\{X, M\}$ a Y, els efectes naturals venen donats per les expressions:

$$EDN = \sum_m \sum_w [E[Y|X = 1, m, w] - E[Y|X = 0, m, w]] \cdot P(M = m|X = 0, w) \cdot P(w) \quad (5.8)$$

$$EIN = \sum_m \sum_w [P(m|X = 1, w) - P(m|X = 0, w)] \cdot E[Y|X = 0, m, w] \cdot P(w) \quad (5.9)$$

Demostració:

La nova hipòtesi ens permet aplicar la segona regla del càlcul *do* per convertir totes les intervencions en observacions en l'equació (5.6), obtenint la fórmula de mediació per l'EDN.

Per l'EIN, aplicant els mateixos arguments intercanviem les intervencions per observacions en el primer terme de (5.7). Pel terme τ, recordem l'expressió de l'efecte causal de X en Y donada en l'anterior demostració:

$$P(y|do(x)) = \sum_m \sum_w P(y|w, do(m), do(x)) P(m|w, do(x)) P(w)$$

De nou, podem substituir les intervencions per observacions per hipòtesi. Finalment, substituint i simplificant els termes que es cancel·len, arribem a la fórmula de mediació de l'EIN. □

6 Cas pràctic

Com a aplicació pràctica, implementem les tècniques descrites fins ara en un estudi de cohorts observacional coordinat pel Dr. Albert Font de l'Institut Català d'Oncologia. L'estudi inclou dades de l'evolució de 160 pacients diagnosticats amb càncer de pròstata metastàsic sotmesos a dos tractaments diferents (Notario et al., 2020 [18]).

L'objectiu principal és estimar l'efecte causal del tractament en l'esdeveniment de mort; l'objectiu secundari és estimar l'efecte causal del tractament en l'esdeveniment combinat progressió de la malaltia o mort.

6.1 Variables

Les variables involucrades són les següents:

- **Tractament (T):** Indica el grup de tractament al qual estan assignats els pacients. N'hi ha dos: $T = 0$ per teràpies supressores d'andròsgens i $T = 1$, on a més de la teràpia supressora s'administra un tractament anomenat Docetaxel.
- **Resultat (Y):** Considerem una variable indicador que, segons estudiem l'objectiu principal o secundari, s'interpreta de la manera següent:
  - **Objectiu principal**: pren el valor 1 quan el pacient ha mort (per qualsevol motiu, no necessàriament relacionat amb la malaltia) i 0 en cas contrari.
  - **Objectiu secundari**: pren el valor 1 quan la malaltia ha progressat o el pacient mor i 0 en cas contrari.
- **Covariables**: La resta de variables de les quals tenim dades són les següents:
  - **Edat (A)**: L'edat del pacient. Està dividida en tres rangs: menors de 65 anys, entre 65 i 75 anys i majors de 75.
  - **Grup de risc (V)**: El grup de risc del pacient en funció del volum de la metàstasi. Si el volum és alt, es consideren d'alt risc. Similarment, si és baix es consideren de risc baix.
  - **Gleason (G)**: Escala de l'1 al 10 emprada per mesurar el grau d'agressivitat del càncer de pròstata. S'obté estudiant les mostres obtingudes en una biòpsia en un microscopi. Les dades estan classificades segons si el valor està per sota de 8 (agressivitat intermitja o baixa) o entre 8 i 10 (agressivitat alta)(Medline Plus, 2020 [15]).
  - **ECOG (E)**: Escala del grau d'activitat que pot dur a terme el pacient i la capacitat que té per cuidar-se un mateix, també anomenat com l'estat funcional (Oken et al., 1982 [19]). Tots els nivells d'ECOG documentats dels participants de l'estudi corresponen als tres primers nivells descrits en la Taula 3.
  - **Símptomes (S)**: La presència o no de símptomes en els participants. En el cas que tinguin símptomes, a les dades s'especifica si són lleus, moderats o severs.
  - **Hemoglobina (Hb)**: La concentració d'hemoglobina en la sang, en g/dL. Les dades es divideixen en dos rangs: per sota i per sobre de 13 g/dL.

– **Antigen prostàtic específic (PSA)**: Una proteïna sintetitzada per les cèl·lules de la pròstata. Es mesura en ng/mL i s'utilitza com a marcador del progrés de la malaltia (amb valors més elevats associats amb un major risc)(Koo et al., 2015 [13]). Les dades estan dividides en els rangs de menys de 20 ng/mL, entre 20 i 100 ng/mL i per sobre de 100 ng/mL.

– **Lactat deshidrogenasa (LDH)**: Enzim utilitzat com a marcador de la progressió del càncer de manera similar al PSA (Wulaningsih et al., 2015 [45]). Es mesura en U/L (on U indica "unitats").

Nivell	Descripció
0	Activitat sense restricció.
1	Activitat física limitada, però pot realitzar feines poc exigents físicament.
2	Pot caminar i cuidar-se d'un mateix, però no pot treballar. Pot moure's més del 50% de les hores que està despert.
3	Pot cuidar-se d'un mateix de manera limitada. Ha d'estar estirat o assentat més del 50% de les hores que està despert.
4	Totalment incapacitat. No es pot cuidar d'un mateix i ha d'estar tot el temps estirat o assentat.
5	Mort.

Taula 3: Nivells de l'escala ECOG.

6.2 Diagrama causal

El diagrama causal del model suggerit pels experts el trobem a la Figura 17. Com a consideracions a tenir en compte:

- Totes les variables influencien el resultat Y, indistintament de si es tracta de l'esdeveniment de mort o de progressió/mort.
- Les variables que influeixen l'assignació del tractament són l'edat, els símptomes, el nivell ECOG i el grup de risc.
- L'antigen prostàtic específic i el lactat deshidrogenasa es comporten d'igual manera en el GAD, influenciant el resultat i el grup de risc. Per aquest motiu, per no complicar excessivament el diagrama les agrupem en un sol node H.
- De manera anàloga, els símptomes i el nivell ECOG tenen el mateix comportament (arestes entrants des del grup de risc i arestes sortints cap al tractament i el resultat). No obstant, com que són causes directes del tractament en un principi les mantenim separades en el GAD.

6.3 Tractament de dades

Hem de trobar l'efecte causal total de T en Y, que en aquest cas com que no hi ha camins indirectes coincideix amb l'efecte directe.

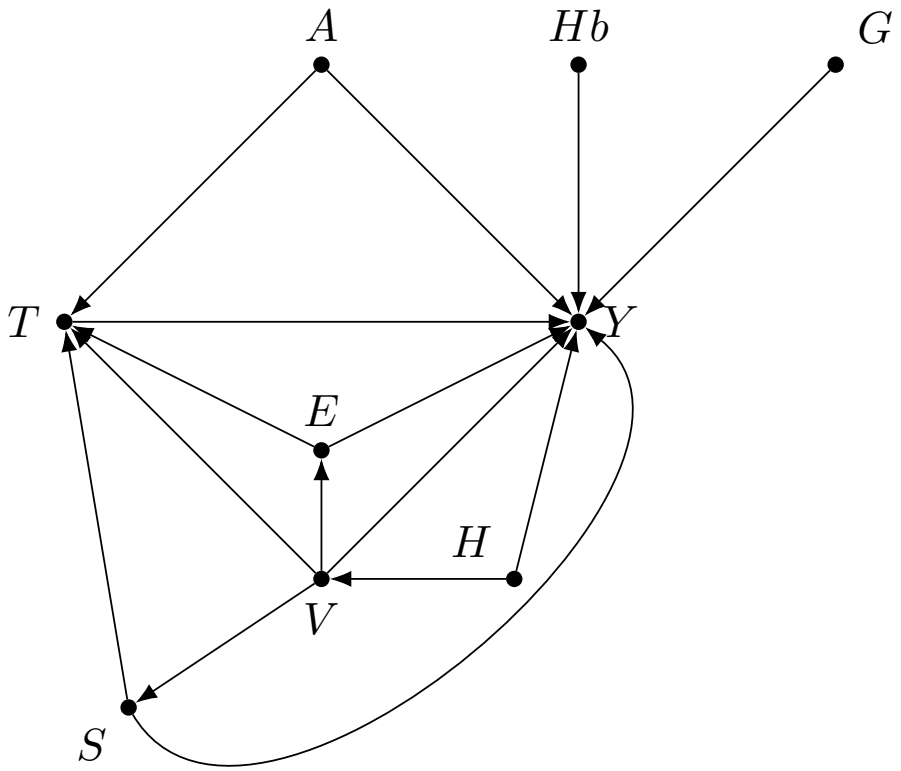

Figura 17: Diagrama causal del model.

Com que no hi ha variables latents, podem bloquejar tots els camins per la porta del darrere de T a Y condicionant al conjunt $\{A, V, S, E\}$. La fórmula d'ajustament ens dona l'expressió de l'efecte causal:

$$P(y|do(T=t)) = \sum_{(a,v,s,e)} P(y|t,a,v,s,e)P(a,v,s,e) \tag{6.1}$$

Recordem que per tal que estigui ben definida l'expressió anterior, s'ha de verificar la positivitat. És a dir, que per totes les combinacions (a, v, s, e) amb $P(a, v, s, e) > 0$ ens hem d'assegurar que hi hagi assignats els dos tractaments diferents per poder contrastar-los. Quan la mida de la mostra és petita i/o el conjunt de variables que ajustem és gran, és més probable que es violi la positivitat.

Recordem que A pren tres valors, S en pren quatre, E en pren tres i V en pren dos. A més, en el cas de S i E trobem alguns casos on els seus valors no estan documentats, de manera que podrien prendre qualsevol valor. Per tant, el primer que cal fer és eliminar aquests casos de la mostra, resultant en 128 pacients dels 160 originals.

A continuació, si ens fixem en la distribució dels símptomes a les dades, trobem que hi ha molts pocs casos de símptomes severs i de moderats en comparació amb les instàncies sense símptomes o amb símptomes lleus (hi ha 4 casos severs, 14 de moderats contra 46 de lleus i 64 casos sense símptomes). Així, havent-hi tants pocs casos severs i moderats, la probabilitat que es violi la positivitat és elevada. En conseqüència, redefinim la variable S com una variable binària que val 1 quan el pacient presenta símptomes (de qualsevol intensitat) i 0 altrament.

Amb la redefinició de S, com a màxim hi ha 36 combinacions amb $P(a, s, e, v) > 0$

on s'ha de verificar la positivitat. En total trobem 27 combinacions amb probabilitat no nul·la, entre les quals es viola la positivitat en 13 instàncies (els resultats els trobem en una taula a l'annex).

Com a solució, definim una nova variable I com un indicador de l'estat del pacient a partir d'E i S. Recordem que E i S tenen el mateix comportament en el diagrama causal i ambdues variables estan relacionades amb l'estat en el qual es troba el pacient. Prenem $I = 0$ quan $S = 0 = E$ (és a dir, quan el pacient no presenta símptomes i pot realitzar qualsevol activitat) i $I = 1$ en la resta de casos (quan el pacient o bé presenta símptomes o bé veu limitada la seva activitat en algun nivell).

Considerem un nou diagrama causal substituint E i S per I, com a la Figura 18. En el nou diagrama, hem d'ajustar pel conjunt $\{A, V, I\}$, que com a màxim pren 12 combinacions de valors diferents. Els resultats estan compilats a la Taula 4.

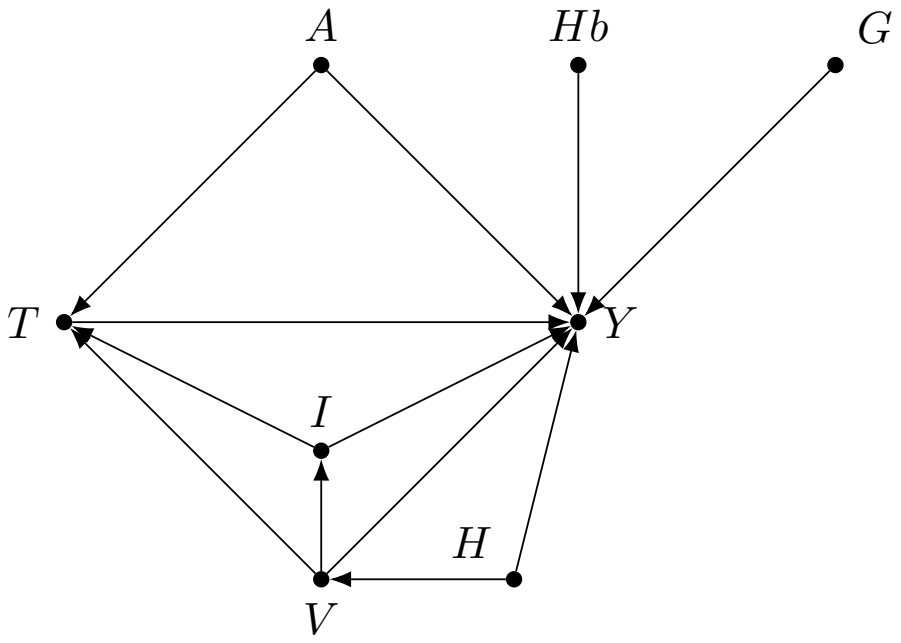

Figura 18: Diagrama causal alternatiu del model amb la variable I.

Observem dues violacions de positivitat pel grup de pacients menors de 65 anys amb $I = 0$, que reben tots el tractament que inclou el Docetaxel. Com que són pacients joves i en bon estat de salut, és raonable que s'optés per un tractament més dur en la seva situació.

Eliminem aquests pacients de la mostra, quedant una mostra de 110 pacients. D'aquesta manera, el que determinem no és τ sinó un estimador $\tau' = E[Y|do(1), (A, I) \neq (< 65, 0)] - E[Y|do(0), (A, I) \neq (< 65, 0)]$. Tanmateix, per no carregar en excés la notació, seguirem escrivint τ.

6.4 Resultats

Donat que les variables resultat Y són binàries amb valors 0 i 1, els valors esperats $E[Y|do(t)]$ que hem de calcular coincideixen amb les probabilitats $P(Y = 1|do(t))$, on

(A, V, I)	$T = 0$	$T = 1$
(<65,Alt risc,0)	0	8
(<65,Alt risc,1)	1	18
(<65,Baix risc,0)	0	10
(<65,Baix risc,1)	3	5
(>75,Alt risc,0)	3	2
(>75,Alt risc,1)	14	1
(>75,Baix risc,0)	5	1
(>75,Baix risc,1)	5	3
(65 - 75,Alt risc,0)	2	7
(65 - 75,Alt risc,1)	3	12
(65 - 75,Baix risc,0)	4	11
(65 - 75,Baix risc,1)	6	4

Taula 4: Distribució del nombre de pacients segons combinacions de (A, V, I) i tipus de tractament.

$t = 0, 1$. Aquests efectes causals els determinem amb la fórmula d'ajustament:

$$P(Y = 1|do(T = t)) = \sum_{(a,v,i)} P(Y = 1|t, a, v, i)P(a, v, i) \tag{6.2}$$

Les probabilitats involucrades en l'equació (6.2) les trobem en les Taules 6-8 de l'Annex, determinades a partir de les freqüències en les dades. Substituint els valors i calculant la diferència entre el tractament que inclou el Docetaxel ($T = 1$) i el que no ($T = 0$), trobem l'efecte causal mig per les dues variables resultat:

- Quan Y és l'esdeveniment de mort, $\tau = P(Y = 1|do(1)) - P(Y = 1|do(0)) = 0.285472 - 0.772403 = -0.486931$.
- Quan Y és l'esdeveniment de progressió/mort, $\tau = 0.559881 - 0.956234 = -0.396353$.

En ambdós casos, l'efecte causal mig és negatiu, amb una diferència al voltant del 50%. Per tant, la probabilitat de mort o de progressió de la malaltia és substancialment major en els pacients sota el tractament que no inclou Docetaxel (aproximadament 2.7 vegades més gran per la mort i 1.7 per la mort/progressió). Aquest resultat no és extrapolable als pacients de < 65 anys que no presenten símptomes ni veuen la seva activitat limitada, independentment del grup de risc al qual pertanyin.

7 Conclusions

La inferència causal és una branca de les matemàtiques que no ha estat gaire explorada fins fa relativament poc. En aquest treball hem introduït les bases del formalisme dels models causals estructurals i dels diagrames causals, desenvolupat principalment en els darrers 30 anys però que segueix creixent en l'actualitat.

Aquest marc teòric permet una representació intuïtiva de les hipòtesis causals del sistema en estudi en termes dels diagrames causals. Com hem vist amb els exemples de la paradoxa de Simpson, aquestes hipòtesis són necessàries per tal de poder respondre qüestions causals. Cal tenir un coneixement previ de l'estructura causal del sistema per tal d'implementar aquestes tècniques.

Partint de models ben definits, hem pogut proporcionar eines com el càlcul *do* o els algorismes ID/IDC que permeten inferir els efectes d'intervencions a partir de dades purament observacionals. Així mateix, també hem vist en l'estudi de la mediació com inferir en quina mesura aquests efectes són directes o deguts a variables mediadores.

Com a exemple d'aplicació d'aquests mètodes, els hem implementat en un estudi observacional, amb dades procedents de la pràctica clínica, que té com objectiu quantificar l'efectivitat de la inclusió del fàrmac *Docetaxel* en el tractament habitual basat en teràpies supressores d'andrògens en pacients diagnosticats de càncer de pròstata metastàsic. Considerant el model causal suggerit pels experts en l'àmbit, hem pogut concloure que la inclusió del Docetaxel en la teràpia supressora d'andrògens redueix substancialment la probabilitat de progressió o mort dels pacients diagnosticats amb càncer de pròstata metastàsic.

8 Annex

8.1 Taules

(A, V, S, E)	$T = 0$	$T = 1$
(<65,Alt risc,0,0)	0	8
(<65,Alt risc,0,1)	0	3
(<65,Alt risc,1,0)	0	3
(<65,Alt risc,1,1)	0	7
(<65,Alt risc,1,2)	1	5
(<65,Baix risc,0,0)	0	10
(<65,Baix risc,0,1)	0	1
(<65,Baix risc,1,0)	1	2
(<65,Baix risc,1,1)	2	2
(>75,Alt risc,0,0)	3	2
(>75,Alt risc,0,1)	2	0
(>75,Alt risc,1,1)	9	1
(>75,Alt risc,1,2)	3	0
(>75,Baix risc,0,0)	5	1
(>75,Baix risc,0,1)	2	0
(>75,Baix risc,1,0)	1	2
(>75,Baix risc,1,1)	2	1
(65 - 75,Alt risc,0,0)	2	7
(65 - 75,Alt risc,0,1)	0	1
(65 - 75,Alt risc,1,0)	0	3
(65 - 75,Alt risc,1,1)	2	7
(65 - 75,Alt risc,1,2)	1	1
(65 - 75,Baix risc,0,0)	4	11
(65 - 75,Baix risc,0,1)	1	1
(65 - 75,Baix risc,1,0)	0	2
(65 - 75,Baix risc,1,1)	4	1
(65 - 75,Baix risc,1,2)	1	0

Taula 5: Combinacions dels valors de (A, V, S, E) amb els dos tractaments.

(A,V,I)	Probabilitat
(<65,Alt risc,1)	0.17272727
(<65,Baix risc,1)	0.07272727
(>75,Alt risc,0)	0.04545455
(>75,Alt risc,1)	0.13636364
(>75,Baix risc,0)	0.05454545
(>75,Baix risc,1)	0.07272727
(65 - 75,Alt risc,0)	0.08181818
(65 - 75,Alt risc,1)	0.13636364
(65 - 75,Baix risc,0)	0.13636364
(65 - 75,Baix risc,1)	0.09090909

Taula 6: Probabilitats de les combinacions de (A, V, I) amb probabilitat positiva.

(A,V,I)	$T = 0$	$T = 1$
(<65,Alt risc,1)	1.0000000	0.3888889
(<65,Baix risc,1)	0.0000000	0.0000000
(>75,Alt risc,0)	0.6666667	0.5000000
(>75,Alt risc,1)	0.7142857	1.0000000
(>75,Baix risc,0)	0.6000000	0.0000000
(>75,Baix risc,1)	0.8000000	0.0000000
(65 - 75,Alt risc,0)	1.0000000	0.1428571
(65 - 75,Alt risc,1)	1.0000000	0.1666667
(65 - 75,Baix risc,0)	0.7500000	0.1818182
(65 - 75,Baix risc,1)	0.6666667	0.0000000

Taula 7: Probabilitats condicionades $P(Y = 1|t, a, v, i)$ per a l'esdeveniment mort.

(A,V,I)	$T = 0$	$T = 1$
(<65,Alt risc,1)	1.0000000	0.7222222
(<65,Baix risc,1)	1.0000000	0.0000000
(>75,Alt risc,0)	1.0000000	1.0000000
(>75,Alt risc,1)	0.7857143	1.0000000
(>75,Baix risc,0)	1.0000000	1.0000000
(>75,Baix risc,1)	0.8000000	0.3333333
(65 - 75,Alt risc,0)	1.0000000	0.4285714
(65 - 75,Alt risc,1)	1.0000000	0.5833333
(65 - 75,Baix risc,0)	1.0000000	0.2727273
(65 - 75,Baix risc,1)	1.0000000	0.2500000

Taula 8: Probabilitats condicionades $P(Y = 1|t, a, v, i)$ per a l'esdeveniment mort/progressió.

Referències

[1] Bareinboim, E., Correa, J., Ibeling, D. i Icard, T. (2020). *On Pearl's Hierarchy and the Foundations of Causal Inference*. Inf. tèc. R-60. A: *Probabilistic and Causal Inference: The Works of Judea Pearl*, ACM Books, per publicar. Causal Artificial Intelligence Lab, Columbia University.

[2] Bruijning-van Dongen, C.J., Slooten, K., Burgers, W. i Wiegerinck, W. (2009). "Bayesian networks for victim identification on the basis of DNA profiles". A: *Forensic Science International: Genetics Supplement Series* 2.1, pàg. 466 - 468. DOI: https://doi.org/10.1016/j.fsigss.2009.08.024.

[3] Cofiño, A., Cano, R., Sordo, C. i Gutiérrez, J. (2002). "Bayesian Networks for Probabilistic Weather Prediction". A: *Proceedings of the 15th European Conference on Artificial Intelligence*. Amsterdam, pàg. 695 - 699.

[4] Charig, C., Webb, D., Payne, S. i Wickham, J. (1986). "Comparison Of Treatment Of Renal Calculi By Open Surgery, Percutaneous Nephrolithotomy, And Extracorporeal Shockwave Lithotripsy". A: *British Medical Journal (Clinical Research Edition)* 292.6524, pàg. 879 - 882. DOI: https://doi.org/10.1136/bmj.292.6524.879.

[5] Galton, F. (1889). *Natural Inheritance*. Londres: Macmillan.

[6] Gebharter, A. i Retzlaff, N. (2020). "A new proposal how to handle counterexamples to Markov causation à la Cartwright, or: fixing the chemical factory". A: *Synthese* 197, pàg. 1467 - 1486. DOI: 10.1007/s11229-018-02014-7.

[7] Harvard Catalyst (2021). *James Robins*. URL: https://connects.catalyst.harvard.edu/Profiles/display/Person/11825 (cons. 10-01-2021).

[8] Harvard Scholar (2021). *Christopher Winship*. URL: https://scholar.harvard.edu/cwinship (cons. 10-01-2021).

[9] Holland, P. (1986). "Statistics and Causal Inference". A: *Journal of the American Statistical Association* 81.396, pàg. 945 - 960. DOI: https://doi.org/10.2307/2289064.

[10] Jin, X., Xu, A., Bie, R., Shen, X. i Yin, M. (2006). "Spam email filtering with bayesian belief network: using relevant words". A: *2006 IEEE International Conference on Granular Computing*. Atlanta, pàg. 238 - 243. DOI: 10.1109/GRC.2006.1635790.

[11] John Hopkins University (2021). *Stephen L. Morgan*. URL: http://socweb.soc.jhu.edu/faculty/morgan/ (cons. 10-01-2021).

[12] Koller, D. i Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Estats Units: The MIT Press.

[13] Koo, K.C., Park, S.U., Kim, K.H., Rha, K., Hong, S., Yang, S. i Chung, B. (2015). "Predictors of survival in prostate cancer patients with bone metastasis and extremely high prostate-specific antigen levels". A: *Prostate International* 3.1, pàg. 10 - 15. DOI: 10.1016/j.prnil.2015.02.006.

[14] McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2a edició. CRC Press. URL: http://xcelab.net/rm/statistical-rethinking/.

[15] Medline Plus (2020). *Gleason grading system*. URL: https://medlineplus.gov/ency/patientinstructions/000920.htm (cons. 22-12-2020).

[16] Neal, B. (2020). *Introduction to Causal Inference from a Machine Learning Perspective*. URL: https://www.bradyneal.com/causal-inference-course (cons. 05-11-2020).

[17] Neyman, J. (1923). *Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes*. Tesi de Màster. Fragments reimpresos en anglès a: *Statistical Science* 5.4, pàg. 463-472.

[18] Notario, L., Piulats, J.M., Sala, N., Ferrandiz, U., González, A., Villa, S., Etxaniz, O., Boladeras, A., Heras, L., del Carpio, L., Rosello, A., Barretina, P., Buisan, O., Fina, C., Pardo, J.C., Suárez, J.M., Comet, J., García del Muro, X., Esteve, A. i Font, A. (2020). *Impact of docetaxel plus androgen-deprivation therapy in patients with metastatic castration-sensitive prostate cancer according to extent of disease*. Pòster presentat a l'European Society for Medical Oncology (ESMO), 19-21 Setembre 2020.

[19] Oken, M., Creech, R., Tormey, D., Horton, J., Davis, T., McFadden, E. i Carbone, P. (1982). "Toxicity and response criteria of the Eastern Cooperative Oncology Group". A: *American Journal of Clinical Oncology* 5.6, pàg. 649 - 655.

[20] Pearl, J. (1982). "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach". A: *Proceedings of the Second AAAI Conference on in Artificial Intelligence*. Pittsburgh, pàg. 133 - 136.

[21] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.

[22] Pearl, J. (1995). "Causal diagrams for empirical research". A: *Biometrika* 82.4, pàg. 669 - 710. DOI: https://doi.org/10.1093/biomet/82.4.669.

[23] Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. 1a edició. Nova York: Cambridge University Press.

[24] Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. 2a edició. Nova York: Cambridge University Press.

[25] Pearl, J. (2014). "Interpretation and Identification of Causal Mediation". A: *Psychological methods* 19, pàg. 459 - 481. DOI: 10.1037/a0036434.

[26] Pearl, J., Glymour, M. i Jewell, N. (2016). *Causal Inference in Statistics: A Primer*. Chichester: John Wiley & Sons.

[27] Pearl, J. i Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Nova York: Basic Books.

[28] Pearson, K. (1911). *The Grammar of Science*. 3a edició. Londres: Adam i Charles Black.

[29] Provine, W.B. (1986). *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.

[30] Rubin, D. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies". A: *Journal of Educational Psychology* 66.5, pàg. 688 - 701. DOI: https://doi.org/10.1037/h0037350.

[31] Shpitser, I. i Pearl, J. (2006a). "Identification of Conditional Interventional Distributions". A: *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*. Boston, pàg. 437 - 444.

[32] Shpitser, I. i Pearl, J. (2006b). "Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models". A: *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Boston, pàg. 1219 - 1226.

[33] Simpson, E. H. (1951). "The Interpretation of Interaction in Contingency Tables". A: *Journal of the Royal Statistical Society: Series B (Methodological)* 13.2, pàg. 238 - 241. DOI: https://doi.org/10.1111/j.2517-6161.1951.tb00088.x.

[34] Spirtes, P., Glymour, C. i Scheines, R. (2001). *Causation, Prediction, and Search*. 2a Edició. Estats Units: The MIT Press.

[35] Stolley, P. (1991). "When Genius Errs: R. A. Fisher and the Lung Cancer Controversy". A: *American Journal of Epidemiology* 133.5, pàg. 416 - 425. DOI: 10.1093/oxfordjournals.aje.a115904.

[36] Tian, J. (2002). "Studies in Causal Reasoning and Learning". Tesi doct. Department of Computer Science, University of California.

[37] Tian, J. i Pearl, J. (2002). "A General Identification Condition for Causal Effects". A: *Proceedings of the Eighteenth National Conference on Artificial Intelligence*. Edmonton, pàg. 567 - 573.

[38] Tikka, S. i Karvanen, J. (2017). "Identifying Causal Effects with the R Package causaleffect". A: *Journal of Statistical Software* 76.12, pàg. 1 - 30. DOI: 10.18637/jss.v076.i12.

[39] UCLA (2021). *Sander Greenland*. URL: https://ph.ucla.edu/faculty/greenland (cons. 10-01-2021).

[40] University of Wisconsin-Madison (2021). *Elwert, Felix*. URL: https://sociology.wisc.edu/staff/elwert-felix-2/ (cons. 10-01-2021).

[41] Verma, T. i Pearl, J. (1990). "Equivalence and synthesis of causal models". A: *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. Cambridge, pàg. 220 - 227.

[42] Wright, S. (1920). "The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs". A: *Proceedings of the National Academy of Sciences* 6.6, pàg. 320 - 332. DOI: 10.1073/pnas.6.6.320.

[43] Wright, S. (1921). "Correlation and causation". A: *Journal of agricultural research* 20.7, pàg. 557 - 585.

[44] Wright, S. (1934). "The Method of Path Coefficients". A: *Annals of Mathematical Statistics* 5.3, pàg. 161 - 215. DOI: https://doi.org/10.1214/aoms/1177732676.

[45] Wulaningsih, W., Holmberg, L., Garmo, H., Malmström, H., Lambe, M., Hammar, N., Walldius, G., Jungner, I., Ng, T. i Van Hemelrijck, M. (2015). "Serum lactate dehydrogenase and survival following cancer diagnosis". A: *British journal of cancer* 113.9. DOI: 10.1038/bjc.2015.361.