SVEUČILIŠTE U ZAGREBU

FILOZOFSKI FAKULTET

ODSJEK ZA FILOZOFIJU

Martin Kirš

PROBABILISTIČKA UZROČNOST I BAYESOVE MREŽE U FILOZOFIJI ZNANOSTI

Diplomski rad

Mentor: doc. dr. sc. Davor Lauc

Zagreb, prosinac 2014.

SADRŽAJ

Probabilistička uzročnost i Bayesove mreža u filozofiji znanosti

Sažetak

Tema rada je probabilistički pristup uzročnosti i modeli Bayesovih mreža za reprezentaciju uzročnih veza. Središnja tvrdnja probabilističke uzročnosti je da uzroci povećavaju vjerojatnost svojim posljedicama. U prvom dijelu rada izložit ću glavne elemente probabilističkog pristupa uzročnosti te kako putem toga pristupa riješiti probleme Humeove teorije uzročnosti. Posebnu pažnju posvetit ću problemu lažnih korelacija, odnosno neuzročnih korelacija u kojoj nijedna korelirana varijabla nije uzrok drugoj. Predloženo rješenje problema lažnih korelacija je Reichenbachovo načelo zajedničkog uzroka putem kojeg korelirane varijable postaju uvjetno nezavisne s obzirom na zajednički uzrok. Raspravit ću također Simpsonov paradoks te predložiti dva rješenja koristeći probabilistički pristup uzročnosti. U drugom dijelu rada prikazat ću osnove Bayesovih mreža i skicirati moguće načine grafičke reprezentacije probabilističke uzročnosti. Dva osnovna dijela Bayesovih mreža su usmjereni aciklički graf i distribucija vjerojatnosti. Prikazat ću glavne uvjete koji moraju biti zadovoljeni pri konstruiranju Bayesovih mreža: Markovljev uvjet, uvjet minimalnosti i uvjet vjernosti. U trećem dijelu rada razmotrit ću argumente za i protiv Bayesovih mreža kao adekvatnih modela za prikazivanje uzročnih veza.

Ključne riječi: uzročnost, vjerojatnosti, Bayesove mreže, Markovljev uvjet, zajednički uzrok, uvjetne vjerojatnosti

Probabilistic Causality and Bayesian Networks in Philosophy of Science

Abstract

The topic of the thesis is the probabilistic approach to causation and the Bayesian networks model for representation of causal relations. Central claim of probabilistic causation is that causes increase the probabilities of their effects. In the first part of the thesis I will lay out the main elements of the probabilistic approach to causation and ways to deal with the problems of Humean theory of causation using that approach. I will give special attention to the problem of spurious correlations, non-causal correlations in which neither variable is a cause of the other. The suggested solution to the problem of spurious correlations is Reichenbach’s principle of common cause which makes correlated variables conditionally independent given their common cause. I will also discuss Simpson's paradox and suggest two solutions using the probabilistic approach to causation. In the second part of the thesis I will present the basics of Bayesian networks and sketch out possible ways of graphically representing probabilistic causation. Two main parts of Bayesian networks are a directed acyclical graph and a probability distribution. I will present main conditions that need to be satisfied in order to build Bayesian networks: Markov condition, minimality condition and faithfulness condition. In the third part of the thesis I will consider the pros and cons of Bayesian networks as appropriate models for the representation of causal relations.

Keywords: causation, probability, Bayesian networks, Markov condition, common cause, conditional probabilities

1. UVOD

Ako kažemo da pušenje uzrokuje rak pluća, na što zapravo mislimo? Tvrdnja je sama po sebi jasna, ako netko puši, znači da će oboljeti od raka pluća. Međutim, ne kaže nam se koji će konkretno pušači oboljeti, kada će točno nastupiti bolest niti što je to što povezuje jedan događaj s drugim. Ako događaj C uzrokuje događaj E, čak i da znamo se sve o C i E, kakva je konkretno narav te veze koja spaja C i E?[1] Je li ona nešto što je inherentno i uzroku i posljedici ili je ona nešto njima izvanjsko što ih drži čvrsto skupa? Ako znamo da postoji uzročna veza između C i E, hoćemo li svaki puta kada opazimo C sa sigurnošću moći reći kako slijedi E? Prema determinističkom shvaćanju uzročnosti, nakon C nužno i bez iznimke slijedi E. Jedan od glavnih zastupnika takvog shvaćanja uzročnosti bio je David Hume. Iako je Aristotel prvi u povijesti filozofije detaljno elaborirao pojam uzročnosti, sve suvremene rasprave počinju s Humeovim empirističkim pristupom uzročnosti. Aristotel je klasificirao uzroke na četiri vrste: djelatni, materijalni, formalni i svršni uzrok, dok su Humeu bili važni samo djelatni uzroci. Oba su uzimala zdravo za gotovo da su uzročne veze deterministički uređene, da uzroci nužno proizvode svoje posljedice. No, je li determinizam održiv?

Determinizam igra na kartu sve ili ništa. Ako uzrok C nužno i bez iznimke dovodi do posljedice E, onda je dovoljno naći jedan primjer u kojem se dogodio C, a nije se dogodio E kako bi determinizam pao u vodu. Pokušajmo s početnim primjerom. Je li istina da svi pušači kad-tad obole od raka pluća? Teško. Neki cjeloživotni pušači umiru prirodnom smrću bez obolijevanja od raka pluća, neki pušači redovito vježbaju pa razvijaju imunost na rak pluća, dok neki pušači imaju izvrsne gene i ne bi oboljeli od raka čak da puše duplo više nego sad. Dakle, iznimaka ima mnogo. S druge strane, pušenje nije jedini uzrok raka pluća. Velika izloženost azbestu također može dovesti do raka pluća. Ako je deterministička koncepcija uzročnosti neistinita, kako adekvatno analizirati uzročne veze?

Opcija koja se prirodno nameće je putem jezika vjerojatnosti. Središnja teza probabilističkog pristupa uzročnosti je kako uzrok povećava vjerojatnost svoje posljedice. Probabilistički pristup nije obavezan igrati na kartu sve ili ništa poput determinizma. Probabilistički pojam uzročnost kaže samo kako je vjerojatnije da će se dogoditi posljedica ako se već dogodio njezin uzrok, nego da se dogodi posljedica bez njenog uzroka. Tvrdnju možemo i preciznije kvalificirati tako da kvantitativno izrazimo kolika je točno vjerojatnost

[1] Kroz cijeli rad koristit ću kratice C (cause) za uzrok, a E (effect) za posljedicu. Englesku riječ „effect“ prevodim kao „posljedica“, a ne kao „učinak“, jer smatram kako je adekvatnija u ovom kontekstu i bolje zahvaća značenje engleske riječi nego doslovan prijevod

da će neki uzrok proizvesti posljedicu. Koristeći matematiku teorije vjerojatnosti, probabilistička uzročnost može se izraziti puno preciznije i elegantnije od grubog determinizma. Ideal probabilističke uzročnosti je reduciranje svih uzročnih tvrdnji na tvrdnje o vjerojatnostima. Drugim riječima, ako uspijemo sve uzročne veze izraziti putem tvrdnji koje se ne pozivaju na uzroke, već samo na jezik vjerojatnosti, redukcija će biti uspješna. Generalno, redukcija bi izgledala ovako:

„C je uzrok E" ako i samo ako C povećava vjerojatnost E

Ideja koja stoji iza reduciranja uzročnih tvrdnji na jezik vjerojatnosti je da se na desnoj strani ekvivalencije ne pojavljuje riječ „uzrok", čime se omogućava govor o uzrocima u terminima jezika vjerojatnosti. Međutim, takva redukcija je samo ideal i to veoma loše formulirani ideal. U zadnjih par poglavlja vidjet ćemo kako postoje snažni argumenti da takva redukcija nije uopće moguća.

Plan rada je sljedeći. U prvom dijelu rada razmotrit ćemo pojam probabilističke uzročnosti te koje su glavne karakteristike takvog pristupa uzročnosti. Krenut ćemo od Humeove teorije uzročnosti i istaknuti glavne nedostatke i protuprimjere koje će probabilistički pristup uspješno riješiti. Veliki dio prvoga dijela bit će posvećen problemu lažnih korelacija, odnosno neuzročnih korelacija koje ne impliciraju uzročne veze. Razmotrit ćemo i Simpsonov paradoks te ga pokušati objasniti koristeći elemente probabilističke uzročnosti kako bi detektirali prave uzroke. U drugom dijelu rada izložit ćemo modele Bayesovih mreži putem kojih možemo na grafički način prikazati uzročne veze i izraziti odnose vjerojatnosti između elemenata Bayesovih mreža. Pokušat ćemo na pristupačan način, potkrijepljen primjerima, definirati glavne uvjete i postupke Bayesovih mreža kako bi putem njih mogli što jednostavnije prikazati uzročne veze. U zadnjem dijelu rada ispitat ćemo jesu li Bayesove mreže adekvatne i dovoljne za reprezentiranje uzročnih veza koje postoje u stvarnosti te izložiti koje su prednosti i nedostaci grafičkog prikazivanja uzročnosti putem Bayesovih mreža.

2. PROBABILISTIČKA UZROČNOST

2.1. Humeova teorija uzročnosti

Suvremene rasprave o uzročnosti počinju s Humeovom teorijom uzročnosti. Hume je razlikovao dvije vrste sudova od kojih se sastoji ljudsko znanje: sudovi o odnosima između ideja i činjenični sudovi. (Morris i Brown, 2014) Sudovi o odnosima između ideja su sudovi *a priori* koji su nezavisni od iskustva te njihova istinosna vrijednost ne ovisi o stanju stvari u svijetu. Primjer takvih sudova su matematički ili geometrijski sudovi poput „2x4=8" ili „zbroj unutarnjih kutova u trokutu je 180º". Činjenični sudovi ovise o stanju stvari u svijetu te se njihova istinosna vrijednost utvrđuje empirijskim putem. Sud poput „Split je južno od Zagreba" je istinit zbog geografskog položaja dvaju gradova, dok je sud „Slovenija je najmnogoljudnija zemlja na svijetu" neistinit zbog postojećih demografskih podataka.

Kojoj vrsti sudova pripadaju sudovi koji izriču uzročno-posljedične veze? Prema Humeu, do uzročnih veza ne možemo doći *apriornim* putem. Glavni razlog tomu je što ne postoji nikakva *nužna* veza između uzroka i posljedice jer uzroci nemaju nikakvu posebnu silu ili energiju u stvaranju određenih posljedica. Čak i ako bismo pretpostavili postojanje *apriornih* nužnih veza između uzroka i posljedice, kako bi znali koji uzroci dovode do kojih posljedica? Ako uzmem u ruke tabletu aspirina i pokušam *apriornim* putem deducirati koje će posljedice uslijediti ako je popijem, neću doći do nikakvih saznanja. (ibid) Dakle, do uzročnih veza možemo doći jedino zaključivanjem iz činjeničnih sudova na temelju iskustva. Do posljedica uzimanja tablete aspirina najbolje ću doći ukoliko popijem nekoliko aspirina i opažam koje posljedice aspirin ima na moje tijelo. Hume definira uzročnost na sljedeći način:

> Slični predmeti uvijek su združeni sa sličnima. O tome nam govori iskustvo. Prema tom iskustvu možemo dakle definirati uzrok kao: *predmet za kojim slijedi drugi, pri čemu za svim predmetima koji su slični prvom slijede predmeti slični drugom.* Ili drugim riječima: *da prvog predmeta nije bilo, ne bi se nikad pojavio ni drugi*. Pojava uzroka običajnim prijelazom uvijek dovodi duh do ideje posljedice. Prema tom iskustvu možemo zato stvoriti drugu definiciju uzroka i nazvati ga *predmetom za kojim slijedi drugi i kojega pojava uvijek vodi misao tom drugom predmetu*. (Hume, 1988: 127; kurziv autorov)

Humeov primjer za ilustraciju uzročnosti bio je udaranje jedne biljarske kugle u drugu. Udaranje prve kugle u drugu je događaj koji uzrokuje drugi događaj, ili posljedicu, a to je kretanje druge kugle. Ako ponovimo udaranje dviju kugli uočit ćemo neke karakteristike. Prvo, uzrok, u ovom slučaju udar prve kugle u drugu, vremenski prethodi posljedici, tj. kretanju druge kugle. Stoga uzročnost podrazumijeva vremenski slijed u kojem uzrok vremenski prethodi svojoj posljedici. Drugo, opažamo kako se u trenutku kada prva kugla udara u drugu i uzrokuje njezino kretanje dviju kugle dodiruju. Međutim, to je sve što možemo osjetilno opaziti ukoliko promatramo interakciju dviju kugli. (Loux, 2010: 215) U našim osjetilima se ne javlja nikakva uzročna veza ili metafizičko ljepilo koje povezuje ta dva događaja. Svaki zaključak koji bi iz dva fizikalna događaja doveo do neopažljive uzročne veze bi, prema Humeu, bio neopravdan. Odnos između uzroka i posljedice, udarca prve biljarske kugle u drugu i kretanja druge kugle, odnos je između dva različita i odvojena događaja.

Humeov empiristički pristup uzročnosti danas je poznat kao teorija pravilnosti (*regularity theory*). Humeovu teoriju uzročnosti možemo definirati na sljedeći način:

c je uzrok e akko:

1) c prostorno-vremenski graniči sa e

2) e vremenski slijedi nakon c

3) svi događaji iz skupa E (događaji koji su slični e) redovito slijede nakon događaja iz skupa C (događaji koji su slični c)[2] (Psillos, 2009)

Prva dva uvjeta su intuitivno jasna, dok treći uvjet kaže da nakon opaženih pojedinačnih događaja c i e možemo generalizirati na skupove događaja C i E koji su slični opaženim događajima c i e. Događaj koji nalikuje udaru prve kugle u drugu kuglu redovito će slijediti događaj koji je sličan događaju u kojemu se druga kugla kreće. Iako Hume negira da između skupova C i E postoji uzročna veza koja bi ih nužno povezivala, ipak smatra kako u „svakom nizu događaj iz prvog skupa stoji u relevantnim vremenskim i prostornim relacijama prema događaju iz drugog skupa." (Loux, 2010: 217) Ako se dogodi događaj iz skupa C, za njim će prostorno slijediti događaj iz skupa E. Uzročne veze se, prema Humeu, mogu reducirati na vrstu stalne konjunkcije koja se odvija prema obrascu koji zadovoljava gornja tri uvjeta.

Međutim, Humeova teorija uzročnosti podložna je mnogim kritikama i propustima. Istaknut ćemo tri najpoznatija prigovora Humeovoj teoriji koje će probabilistički pristup uzročnosti pokušati riješiti.

1) *Problem lažnih korelacija (spurious correlations) ili neuzročnih asocijacija.* Činjenica da su dva događaja korelirana, odnosno da nakon prvog događaja redovito slijedi drugi događaj, ne implicira nužno da je prvi događaj uzrok, a drugi događaj njegova posljedica. Uzmimo trivijalni primjer izmjene dana i noći. Dolazak noći redovito slijedi nakon završetka

[2] Velika slova označuju skupove sličnih događaja, dok mala slova pojedinačne događaje

dana i taj događaj zadovoljava tri Humeova uvjeta uzročnosti. Međutim, krivo bi bilo reći kako dan uzrokuje noć. (Loux 2010: 218) Primjer koji su zastupnici probabilističkog pristupa uzročnosti isticali protiv Humeove teorije je visoka korelacija između oluje i pada žive u barometru. Svaki puta kada žive padne u barometru dogodi se oluja. Međutim, pad žive u barometru nije uzrok oluje, već je njihova korelacija rezultat zajedničkog uzroka koji djeluje na oba događaja, u ovom slučaju pada atmosferskog tlaka zraka. (Hitchcock: 2012) Humeova teorija ne uočava u potpunosti kako je uzročni odnos nužno *asimetričan* odnos između uzroka i posljedice, tj. ako je C uzrok E, onda E ne može biti uzrok C. Kod primjera oluje i pada žive u barometru taj nedostatak nije vidljiv, ali u primjeru izmjene dana i noći možemo proizvoljno mijenjati uzrok i posljedicu.

2) *Problem nesavršenih pravilnosti*. Humeova teorija pretpostavlja determinizam, nakon uzroka invarijantno slijedi posljedica. Međutim, takav uvjet je prestrog jer većina uzročno-posljedičnih odnosa ima iznimke. Nesavršene pravilnosti nastaju zbog dva razloga. Prvi je *heterogenost* okolnosti. Iako je pušenje uzrok raka pluća, ne obolijevaju svi pušači od raka pluća, neki pušači mogu redovito vježbati i postati imuni na rak pluća, dok neki nepušači mogu oboljeti od raka pluća jer su izloženi azbestozi. Drugi razlog nesavršenih pravilnosti je znanstvena upitnost fizikalnog determinizma u kojem je svijet deterministički strukturiran i svaki uzrok invarijantno dovodi do svoje posljedice. (ibid) Razvoj kvantne mehanike poljuljao je nade u determinizam u prirodi, dok je u društvenim znanostima vrlo teško pronaći uzroke koji redovito dovode do istih posljedica.[3]

3) *Problem irelevantnosti*. Dva događaja mogu biti prostorno-vremenski povezana i korelirana, ali intuitivno možemo procijeniti kako je jedan događaj irelevantan za drugi. Vrač može „začarati“ sol riječima „hokus pokus!“ prije nego što stavi sol u vodu, ali vračeva izjava ne doprinosi tomu da se sol otopi u vodi. Sol bi se otopila u vodi i da vrač nije izjavio „hokus pokus!“ jer njegova izjava *ne čini razliku* za posljedicu, otapanje soli u vodi, iako joj vremenski prethodi. (ibid)

[3] U društvenim znanostima problem nesavršenih pravilnosti dolazi do izražaja prilikom generalizacije pojedinačnih događaja na opće događaje. Prema Humeovoj teoriji, generalizacijom na temelju pojedinačnih uzroka dolazi se do općih uzročno-posljedičnih veza. Iako je istina da je ubojstvo Franje Ferdinanda uzrok Prvog svjetskog rata, generalizacija da su atentati visokih dužnosnika uzrok svjetskih ratova nije istinita zbog prevelikog broja protuprimjera. (Loux, 2010: 219)

2.2. Zašto probabilistička uzročnost?

Potrebu za probabilističkom formulacijom uzročnosti najlakše možemo dočarati primjerima iz svakodnevnog života. Ukoliko želimo reći kako je brza i neoprezna vožnja uzrok prometnih nesreća, ne bi nam puno pomoglo ako kažemo da svaka brza i neoprezna vožnja dovodi do prometnih nesreća. Točnije bi bilo reći kako je vjerojatnost prometnih nesreća puno veća ukoliko netko brzo i neoprezno vozi, nego ako netko vozi dopuštenom brzinom. Ukoliko nam neoprezan vozač koji vozi znatno brže od dozvoljene brzine kaže kako do sada nije imao niti jednu prometnu nesreću, reći ćemo mu kako ima sreće te da, ukoliko nastavi brzo i neoprezno voziti, će mu se kad-tad dogoditi prometna nesreća. Teško će nam biti povjerovati kako brz i neoprezan vozač još nije imao prometnu nesreću jer nam intuicija govori da brzi i neoprezni vozači u prosjeku imaju više nesreća nego oprezni vozači te, u skladu s tim, njihova vjerojatnost da dožive nesreću je veća. (Suppes, 1970: 7) Ukoliko ne učimo za ispit vjerojatnije je da ga nećemo položiti, iako pad na ispitu ne mora biti nužna posljedica neučenja, možemo biti prirodno talentirani za taj predmet ili ispit može sadržavati lagana pitanja. U oba slučaja uzrok čini posljedicu vjerojatnijom u onoj mjeri u kojoj brzi i neoprezni vozači češće dožive prometnu nesreću od opreznih te u onoj mjeri u kojoj studenti koji uče za ispit češće polože ispit od onih koji ne uče.

Središnja tvrdnja probabilističkog pristupa uzročnosti je kako uzroci povećavaju vjerojatnost posljedicama.[4] Iako još nismo uveli potrebnu notaciju, probabilističku definiciju uzročnosti možemo izraziti na sljedeći način:

(1) $P(E \mid C) > P(E)$

(2) $P(E \mid C) > P(E \mid \neg C)$

Prva nejednakost kaže kako je vjerojatnost da će se posljedica E dogoditi veća ukoliko se uzrok C već dogodio, od vjerojatnosti da se dogodi samo posljedica E. Druga nejednakost kaže kako je vjerojatnost da će se dogoditi E, ako se već dogodio C, veća od vjerojatnosti da će se dogoditi E ukoliko se C nije dogodio.

Nakon što smo definirali probabilističku uzročnost, barem na općenitoj razini, pokušajmo riješiti probleme s kojim se susrela Humeova teorija uzročnosti. Rješenje problema lažnih korelacija se, ukratko, sastoji u tome da se postulira kako za neke korelacije postoji zajednički

[4] Postoje dvije razine probabilističke analize uzročnosti: analiza općih (*type*) uzročnih veza i pojedinačnih (*token*) uzročnih veza. Prema Eellsu (1991: 5) opće uzročne veze se analiziraju *komparativnom uvjetnom vjerojatnošću*, dok se pojedinačne uzročne veze analiziraju *temporalnom promjenom vjerojatnosti*. U ovom radu bavit ću se uglavnom općom razinom uzročnosti

uzrok putem kojeg dvije korelirane varijable postaju nezavisne jedna o drugoj. Međutim, o načelu zajedničkog uzroka bit će riječi kasnije. Kako probabilističkim pristupom uzročnost riješiti ostala dva problema?

Problem nesavršenih pravilnosti ne predstavlja problem za probabilistički pristup jer je on uklonjen već u samoj definiciji probabilističke uzročnosti. Teorija vjerojatnosti polazi od pretpostavke kako svijet nije deterministički strukturiran i kako uzroci ne dovode invarijantno do svojih posljedica. U primjerima neoprezne vožnje i prometnih nesreća te neučenja i pada na ispitu očito je kako nijedan uzrok ne mora invarijantno dovesti do svoje posljedice te da su iznimke potpuno kompatibilne s probabilističkim pristupom uzročnosti. Štoviše, probabilističkim pristupom možemo kvantitativno izraziti kolika je vjerojatnost da će dogoditi određene iznimke koje odstupaju od pravila. Čak i ukoliko bismo željeli izraziti invarijantnu uzročno-posljedičnu vezu mogli bismo uzročnoj vezi dodijeliti vjerojatnost 1, $P(E \mid C) = 1$, čime bismo rekli kako se posljedica E javlja svaki puta kada se dogodio uzrok C.

Drugi problem, problem irelevantnosti, probabilistički pristup također uspješno rješava. Problem irelevantnosti unutar Humeove teorije nije dozvoljavao razlikovanje između događaja koji ne čine razliku za javljanje posljedice od pravih uzroka koji čine razliku. Neznatnom modifikacijom druge definicije probabilističke uzročnosti možemo skicirati rješenje tog problema. Već nam i sama nejednakost kaže kako C čini razliku za E, odnosno povećava vjerojatnost da će se E dogoditi, tj. $P(E \mid C) > P(E \mid \neg C)$. Ukoliko želimo razlikovati događaje koji ne čine razliku za javljanje posljedice E, možemo staviti pravi uzrok C u konjunkciju s irelevantnim događajima. Dobili bismo sljedeći izraz: $P(E \mid C \;\&\; A) = P(E \mid C \;\&\; \neg A)$. Drugim riječima, činjenica da se A dogodio ili nije ne mijenja vjerojatnost da će se dogoditi posljedica E. A je u ovom slučaju irelevantan događaj koji ne čini razliku za posljedicu E, dok vjerojatnost E varira ovisno o tome je li se dogodio pravi uzrok C ili nije.

2.3. Aksiomi teorije vjerojatnosti i uvjetna vjerojatnost

Aksiome teorije vjerojatnosti prvi je formalizirao Kolmogorov 1933. godine. Probabilistički model ili sustav sastoji se od funkcije vjerojatnosti P koja dodjeljuje skupu mogućih ishoda A numeričku vrijednost, te prostora uzorka Ω koji je skup svih mogućih ishoda. (Bertsekas i Tsitsiklis, 2008: 6) Funkcija vjerojatnosti dodjeljuje se događajima unutar prostora uzorka, iako se može pripisivati i sudovima. (Howson i Urbach, 2005) Događaji unutar prostora uzorka moraju biti međusobno isključivi te kolektivno iscrpni.[5] (Bertsekas i Tsitsiklis, 2008: 7) Funkcija vjerojatnosti mora zadovoljiti sljedeće aksiome:

(1) *Ne-negativnost.* $P(A) \geq 0$

(2) *Aditivnost.* Ako su A i B dva razdvojena događaja (tj. $A \cap B = \emptyset$), onda je vjerojatnost njihove unije $P(A \cup B) = P(A) + P(B)$. Aksiom aditivnosti možemo generalizirati na niz razdvojenih događaja unutar Ω: $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$

(3) *Normalizacija.* Vjerojatnost cijelog prostora uzorka je 1, tj. $P(\Omega) = 1$ (ibid: 9)

Primjenom aksioma možemo dokazati negaciju ili komplement od događaja A:

(4) $P(\neg A) = 1 - P(A)$

$P(A \cup \neg A) = 1$	prema 3. aksiomu, A i ¬A iscrpljuju prostor uzorka
$P(A \cup \neg A) = P(A) + P(\neg A)$	prema 2. aksiomu, A i ¬A su odvojeni događaji
$1 = P(A) + P(\neg A)$	
$P(\neg A) = 1 - P(A)$	

Vjerojatnost praznog skupa je 0:

(5) $P(\emptyset) = 0$

$P(\emptyset) = 1 - P(\Omega)$	prema (4), Ω je komplement od Ø
$P(\emptyset) = 1 - 1$	prema 3. aksiomu

[5] Ukoliko jednom bacamo kocku prostor uzorka ne može istovremeno sadržavati događaje „kocka padne na brojeve 4 ili 6“ te „kocka padne na brojeve 4 ili 5“ jer dva događaja nisu međusobno isključiva. Događaji unutar prostora uzorka moraju biti kolektivno iscrpni u smisli da su svi mogući ishodi koje možemo dobiti bacanjem kocke zastupljeni unutar prostora uzorka. Primjerice, prostor uzorka ne može sadržavati događaje „kocka padne na brojeve 2 ili 3 ili 4“ i „kocka padne na brojeve 5 ili 6“ jer ne bismo pokrili događaj da kocka padne na broj 1

$P(\emptyset) = 0$

Ukoliko su A i B logički ekvivalentni:

(6) $P(A) = P(B)$

$P(A \cup \neg B) = 1$	A i $\neg$B su međusobno isključivi, iscrpljuju prostor uzorka
$P(A \cup \neg B) = P(A) + P(\neg B)$	prema 2. aksiomu
$1 = P(A) + P(\neg B)$	
$1 = P(A) + 1 - P(B)$	prema (4)
$P(A) = P(B)$	

Ukoliko znamo da se neki događaj unutar prostora uzorka već dogodio tada ograničavamo prostor uzorka i usklađujemo distribuciju vjerojatnosti s obzirom na taj događaj. U tom slučaju koristimo *uvjetnu vjerojatnost*, tj. $P(A \mid B)$. Ako znamo da se dogodio događaj B, želimo izračunati kolika je vjerojatnost događaja A *s obzirom na* B. Uvjetnu vjerojatnost definiramo kao:

(7) $P(A \mid B) = P(A \cap B) / P(B)$, ako je $P(B) > 0$

Iz definicije uvjetne vjerojatnosti možemo izvesti popularan Bayesov teorem. Ako znamo da se dogodio događaj B te imamo više mogućih scenarija $A_1, \ldots A_n$ putem kojih se mogao dogoditi događaj B, želimo izračunati vjerojatnost svakog mogućeg scenarija. $P(A_i \mid B)$ je *posteriorna vjerojatnost* A_i s obzirom na B, $P(A_i)$ je *prethodna vjerojatnost* pojedinog mogućeg scenarija, dok je $P(B \mid A_i)$ vjerojatnost događaja B s obzirom na određeni scenarij. Bayesov teorem možemo izraziti na dva načina:

(8a) $P(A \mid B) = P(A)\, P(B \mid A) / P(B)$

$P(A \mid B) = P(A \cap B) / P(B)$	uvjetna vjerojatnost iz (7)
$P(A \mid B) = P(A)\, P(B \mid A) / P(B)$	$P(A \cap B) = P(B \mid A)\, P(A)$, prema (7)

(8b) $P(A_i \mid B) = P(A_i)\, P(B \mid A_i) / P(A_1)\, P(B \mid A_1) + \ldots + P(A_n)\, P(B \mid A_n)$

Ako se dogodio događaj B, i to nam ništa ne govori o događaju A, kažemo kako su A i B *nezavisni.* U tom slučaju P(A | B) = P(A), odnosno uvjetna vjerojatnost A s obzirom na B je jednaka kao i bezuvjetna vjerojatnost A. Ukoliko su dva događaja nezavisna važit će sljedeća jednakost:

(9) $P(A \cap B) = P(A)\, P(B)$

Nezavisnost dvaju događaja možemo izraziti i putem uvjetne vjerojatnosti. Događaji A i B bit će *uvjetno nezavisni* s obzirom na događaj C ukoliko vrijedi:

(10) $P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$, ako $P(C) > 0$

Nezavisnost dva događaja A i B ima svojstvo simetrije, ako je A nezavisan sa B, onda je i B nezavisan sa A. Isto svojstvo vrijedi i za uvjetnu nezavisnost. Ako su A i B nezavisni događaji možemo jedan događaj prebaciti i na stanu uvjetnog događaja C. $P(A \mid B \cap C)$ je jednako kao i $P(A \mid C)$. Isto vrijedi i za B, $P(B \mid A \cap C) = P(B \mid A)$. Međutim, ako su A i B bezuvjetno nezavisni, ne slijedi da su oni nužno i uvjetno nezavisni s obzirom na treći događaj C. Vrijedi i obratno, uvjetna nezavisnost dvaju događaja s obzirom na treći ne implicira bezuvjetnu nezavisnost tih dvaju događaja.

Ako imamo sekvencu događaja $A_1, A_2, \ldots, A_n$, množimo uvjetne vjerojatnosti svakog događaja s obzirom na sve prethodne događaje, osim prvog događaja u nizu. U tom slučaju koristimo *pravilo multiplikacije*:

(11) $P(\bigcap_{i=1}^{n} A_i) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_2 \cap A_1) \ldots\ P(A_n \mid \bigcap_{i=1}^{n-1} A_i)$

2.4. Načelo zajedničkog uzroka i zaslonjavanje lažnih korelacija

Nakon što smo definirali aksiome teorije vjerojatnosti i iz njih deducirali najvažnije teoreme, možemo detaljnije analizirati problem lažnih korelacija. Za Humeovu teoriju uzročnosti korelacije su jednake uzročnim vezama, odnosno, uzročne veze se mogu svesti na stalne konjunkcije dvaju događaja (ili varijabli). Međutim, korelacija dviju varijabli ne implicira da su one u uzročnoj vezi te da jednu varijablu možemo smatrati kao uzrok drugoj. Uzročnost je asimetrični odnos, dok je korelacija simetričan odnos. Ako su X i Y u pozitivnoj korelaciji onda X povećava vjerojatnost Y ako i samo ako Y povećava vjerojatnost X. S druge strane, ako je X uzrok Y, onda Y nije uzrok X. Ako su X i Y korelirane varijable, onda su X i Y međusobno zavisni:

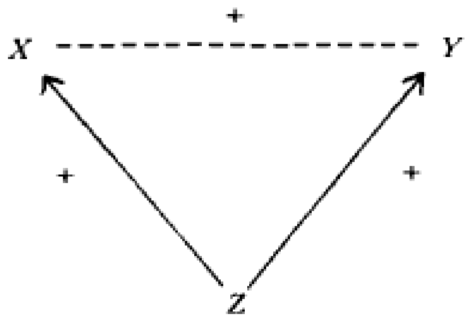
$$P(X \mid Y) > P(X)$$

$$P(Y \mid X) > P(Y)$$

Vjerojatnost da se dogodi X je veća ako se već dogodio Y ili ako imamo neku informaciju o Y, od toga da se dogodi samo X. Nejednakost vrijedi i ako zamijenimo X i Y. Ranije je spomenut primjer pozitivne korelacije pada žive u barometru i oluje u kojem nema uzročne veze između dva događaja, niti pad žive u barometru ne uzrokuje oluju niti oluja uzrokuje pad žive u barometru. Sličan primjer daje Sober (1987). Zamislimo da u nekoj kazališnoj grupi svakih sto dana jedan glumac oboli od želučane infekcije. Pretpostavimo kako je frekvencija obolijevanja glumaca dugoročno stabilna. Slučajno odabrani glumac ima 1/100 vjerojatnost od obolijevanja od želučane infekcije, tj. 0.01. Ako su obolijevanja glumaca međusobno nezavisna, vjerojatnost da će dva glumca oboljeti je 1/10,000 ili jednom u 10,000 dana. Pretpostavimo kako smo opazili da ukoliko jedan glumac oboli od želučane infekcije, od iste bolesti obole i ostali glumci iz kazališne grupe. U tom slučaju, vjerojatnost da dva glumca obole od želučane infekcije (1/100) veća je od umnoška vjerojatnosti da svaki glumac oboli (1/100 x 1/100). Tada korelaciju možemo izraziti na sljedeći način:

$$P(X_1 \cap X_2) > P(X_1)\, P(X_2)$$

Jedno od rješenja problema lažnih korelacija je postuliranje zajedničkog uzroka koji uzrokuje korelaciju dvije varijable. Ako je korelacija X i Y lažna, odnosno X ne uzrokuje Y niti Y uzrokuje X, kontroliranjem zajedničkog uzroka Z objašnjavamo korelaciju između X i Y na uzročan način. Graf prikazuje korelaciju isprekidanom linijom između X i Y, dok je uzročna veza između Z i X te Z i Y prikazana punom linijom.



Slika 1. Struktura zajedničkog uzroka (Eells, 1991: 59)

Načelo zajedničkog uzroka prvi je formulirao Reichenbach (1956: 157-160). Načelom zajedničkog uzroka Reichenbach je želio učiniti korelirane varijable X i Y *uvjetno nezavisnima* s obzirom na zajednički uzrok Z. Ako postoji zajednički uzrok korelacije $P(X \cap Y) > P(X) P(Y)$ sljedeći uvjeti moraju biti ispunjeni:

(1) $P(X \mid Z) > P(X \mid \neg Z)$

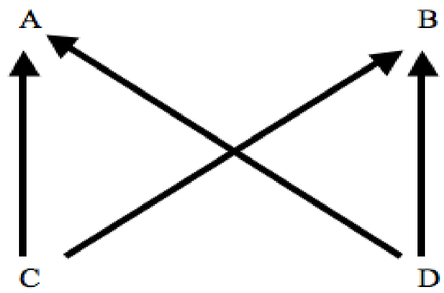
(2) $P(Y \mid Z) > P(Y \mid \neg Z)$

(3) $P(X \cap Y \mid Z) = P(X \mid Z) P(Y \mid Z)$

(4) $P(X \cap Y \mid \neg Z) = P(X \mid \neg Z) P(Y \mid \neg Z)$

Uvjeti (1) i (2) govore kako je Z uzrok X, odnosno Y. Vjerojatnost da se dogodi posljedica X, ili Y, veća je kad se već dogodio Z nego kad se nije. Uvjeti (3) i (4) pokazuju kako korelacija između X i Y nestaje ukoliko postavimo kao uvjet Z. X i Y su tada uvjetno nezavisni s obzirom na Z i $\neg Z$, što jednakosti iz (3) i (4) pokazuju. Drugim riječima, kažemo da zajednički uzrok Z *zaslanja (screens off)* X od Y i/ili Y od X. Početna nejednakost u kojoj je $P(X \cap Y)$ bio veći od umnoška $P(X) P(Y)$, pretvorila se u jednakost $P(X \cap Y \mid Z) = P(X \mid Z) P(Y \mid Z)$, sukladno formuli uvjetne nezavisnosti, kada smo uvjetovali korelirane varijable X i Y na zajednički uzrok Z.

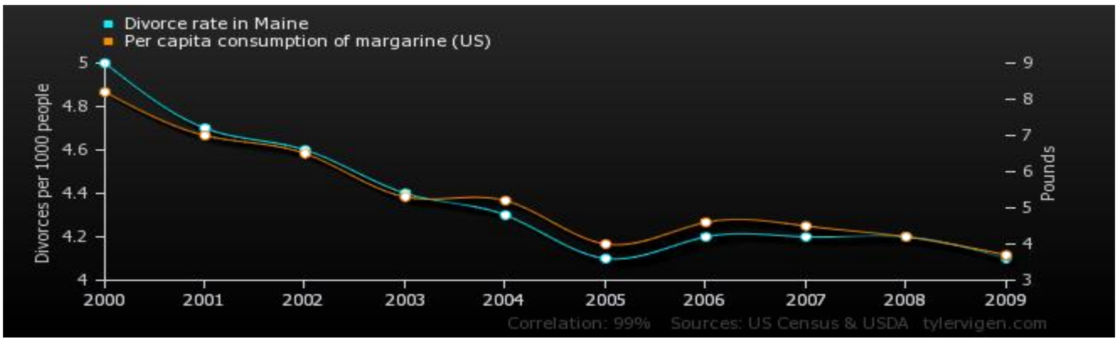
Načelo zajedničkog uzroka ne može uzročno objasniti sve lažne korelacije. Možemo uzeti dva primjera korelacija koje ne mogu postati uvjetno nezavisne kad ih uvjetujemo na zajednički uzrok. Prvi primjer su korelacije koje mogu imati više od jednog zajedničkog uzroka te Reichenbachova četiri uvjeta ne bi bila dovoljna da korelirane varijable postanu uvjetno nezavisne. Fenomen možemo ilustrirati sljedećim grafom



Slika 2. Lažna korelacija sa dva zajednička uzroka (Hitchcock, 2012)

Ako korelaciju između A i B uvjetujemo na zajednički uzrok C, C još uvijek neće zaslanjati A od B jer postoji dodatni uzrok D koji uzrokuje korelaciju A i B. A i B neće biti uvjetno nezavisne s obzirom na C. Dakle, Reichenbachov uvjet (3) je narušen jer $P(A \cap B \mid C) \neq P(A \mid C) P(B \mid C)$ zbog dodatnog uzroka D. (Hitchcock, 2012: Sober, 1987)

Drugi primjer kada zajednički uzrok nije dovoljan za uzročno objašnjenje korelacije je kada korelirane varijable imaju slične trendove unutar vremenskih nizova. Pretpostavimo kako cijene kruha u Velikoj Britaniji i razina mora u Veneciji rastu identičnim trendom. (Sober, 2001) U onoj mjeri u kojoj raste cijena kruha u Velikoj Britaniji u jednom periodu, u istoj mjeri raste i razina mora u Veneciji. Iako su dvije varijable visoko pozitivno korelirane intuicija nam govori kako ne možemo naći zajednički uzrok jer su varijable same po sebi nezavisne. Obje varijable su rezultat izoliranih endogenih uzroka koji operiraju u lokalnim okolnostima. Možemo navesti još neke primjere u kojima nam intuicija govori da su varijable već same po sebi nezavisne i da nije potreban zajednički uzrok kako bi postale uvjetno nezavisne. Iznimno je visoka korelacija ($\rho = 0.9925$) između potrošnje margarina u SAD (po glavi stanovnika) i stope razvoda u saveznoj državi Maine u razdoblju od 2000. do 2009. godine.



Slika 3. Korelacija varijabli sa sličnim trendovima (http://www.tylervigen.com)

2.5. Simpsonov paradoks

Načelo zajedničkog uzroka pokazalo se kao učinkoviti način uzročnog objašnjenja nekih lažnih korelacija. Međutim, zaključak o uzročnim vezama na temelju korelacija nije posve riješen. Statistički fenomen poznat kao Simpsonov paradoks sredinom 70-tih godina prošlog stoljeća postao je popularan zbog problema rodne diskriminacije pri upisu na diplomske studije na Berkeleyju u SAD-u. Radi se o sljedećem: zamislimo da u cijeloj populaciji uočimo kako je faktor X pozitivno koreliran s ishodom Y, tj. vjerojatnost da će se Y dogoditi je veća ukoliko je prisutno svojstvo X. Međutim, ako podijelimo populaciju po nekom kriteriju uočit ćemo kako je pozitivna korelacija X na Y *obratna* u svakoj podijeljenoj grupi unutar populacije. X više nije pozitivno koreliran sa Y već negativno, a svojstvo X smanjuje vjerojatnost da će doći do ishoda Y. (Malinas i Bigelow, 2012) Paradoks možemo ilustrirati na primjeru rodne zastupljenosti na dva odsjeka unutar fakulteta:[6]

	Odsjek A	Odsjek B	Ukupno
Muškarci	63/90	2/10	65/100
Žene	8/10	27/90	35/100

Tablica 1. Simpsonov paradoks na primjeru 2 odsjeka (Eells, 1991)

Ako gledamo ukupnu populaciju, od 100 slobodnih mjesta na dva odsjeka primljeno je 65 muškaraca te 35 žena, dok je vjerojatnost da ćete biti upisani na fakultet skoro dvostruko veća ukoliko ste muškarac. Nameće se zaključak kako svojstvo „biti muškarac“ povećava vjerojatnost upisa na fakultet. Međutim, ukoliko podijelimo fakultet na dva odsjeka i pogledamo koji su omjeri muškaraca i žena na svakom odsjeku pojedinačno, uočit ćemo suprotan utjecaj roda na upis. Na odsjeku A, vjerojatnost da ćete biti upisani ako ste žena je 0.8, dok je za muškarce nešto manja, 0.7. Odsjek B također favorizira žene, 0.3 spram 0.2. Paradoks možemo generalizirati na sljedeći način, ako su a, b, c, d, A, B, C i D cijeli brojevi (ibid):

$$a / b < A / B$$

$$c / d < C / D$$

$$(a + c) / (b + d) > (A + C) / (B + D)$$

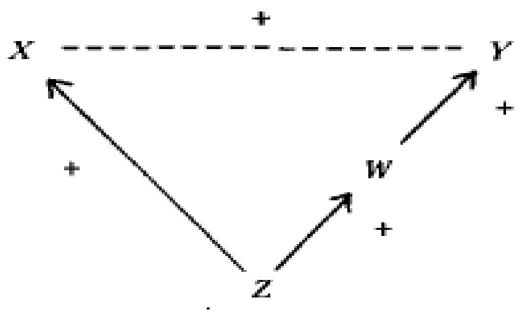
Kako interpretirati Simpsonov paradoks? Jesu li žene diskriminirane pri upisu na fakultet ili nisu? Gledajući tablicu, možemo zaključiti kako je odsjek B puno teže upisati nego odsjek

[6]Izvorna studija o rodnoj diskriminaciji na Berkeleyju podijelila je populaciju fakulteta na 85 odsjeka (Cartwright, 1979), idealizirani primjer sa 2 odsjeka uzet je, uz neznatne modifikacije podataka, iz Eells (1991: 63)

A. Odsjek A je upisalo 71 od 100 studenata, dok je odsjek B upisalo tek 29 od 100 studenata. Odsjek B ima devet puta više prijava žena nego odsjek A (90:10), dok je kod muškaraca taj omjer obratan. Objašnjenje koje možemo izvesti iz podataka u tablici je kako su se žene u puno većoj mjeri (9x) prijavljivale na odsjek koji je teže upisati, dok je kod muškaraca slučaj obratan, oni su u istoj mjeri (9x) prijavljivali lakši odsjek. Muškarci su zastupljeniji u ukupnoj populaciji fakulteta jer su se prijavljivali na lakši odsjek stoga ih je veći broj u konačnici upisan.

Simpsonov paradoks možemo riješiti na dva načina probabilističkim pristupom uzročnosti. Osnovno je pitanje uzrokuje li rod upis na fakultet, odnosno, povećava li rod (u slučaju muškaraca) vjerojatnost upisa na fakultet ili rod (u slučaju žena) smanjuje vjerojatnost da će prijave rezultirati upisom. Paradoks smo već djelomično objasnili interpretacijom same tablice te utvrdili kako je veći broj upisanih muškaraca na fakultet posljedica prijavljivanja žena na teži odsjek. Međutim, postoji li zaista uzročna veza između roda i upisa te kako se nositi s kontraintuitivnom činjenicom da svaki pojedinačni odsjek ipak favorizira žene?

Prvo rješenje paradoksa iz perspektive probabilističkog pristupa uzročnosti je primjena načela zajedničkog uzroka (Eells, 1991). Ako je X svojstvo „biti muškarac“, a Y upis na fakultet te je Y pozitivno koreliran sa X, postoji li zajednički uzrok Z putem kojeg X i Y postaju uvjetno nezavisni? Eells smatra kako prvo trebamo razlučiti postoji li veza između svojstva „biti muškarac“ i prijavljivanja na lakše odsjeke. S obzirom da prijava na lakše odsjeke ne uzrokuje nečiji rod, Eells tvrdi kako svojstvo „biti muškarac“ uzrokuje prijavu na lakše odsjeke. Prijava na lakše odsjeke je posrednički faktor između svojstva „biti muškarac“ kao početnog uzroka i upisa na fakultet kao konačne posljedice. Radi se o svojstvu tranzitivnosti kod uzročnih veza, ako C uzrokuje P, a P uzrokuje E, onda možemo zaključiti da C uzrokuje E *posredstvom* P.



Slika 4. Rješenje Simpsonovog paradoksa načelom zajedničkog uzroka (Eells, 1991: 66)

Z na slici označuje svojstvo „biti muškarac", W prijavu na lakše odsjeke, Y upis na fakultet, dok je X vjerovanje fakulteta da je prijavljeni student muškarac. Ako fakultet namjerno diskriminira žene pri upisu, smatra Eells, tada je opravdano razlikovati varijable „biti muškarac" od vjerovanja fakulteta da upisuje više muškaraca nauštrb žena.[7] Z pozitivno utječe na W, dok W pozitivno utječe na Y, uzročna veza je Z→W→Y te sukladno tranzitivnosti Z uzrokuje Y. X i Y su u lažnoj korelaciji koja ne sadržava uzročne veze od X prema Y. Svojstvo „biti muškarac" je pravi uzrok upisa na fakultet, dok je vjerovanje fakulteta da je netko muškarac uvjetno nezavisno od upisa na fakultet jednom kada znamo pravi uzrok. (ibid: 66)

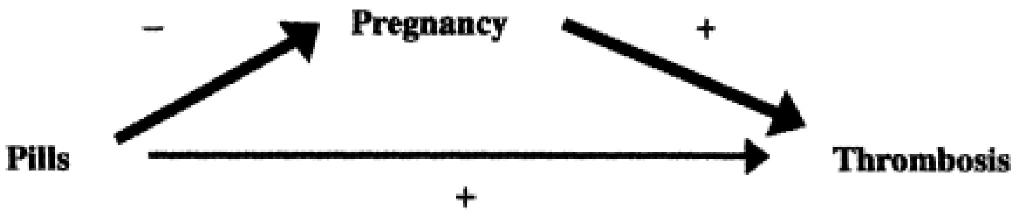
Slično rješenje nudi Cartwright (1979), ali bez pozivanja na načelo zajedničkog uzroka. Prema njezinoj interpretaciji, ako želimo procijeniti uzročnu vezu između svojstva „biti muškarac" i upisa na fakultet moramo posjedovati relevantno pozadinsko znanje o svim mogućim uzrocima koji su mogli utjecati na određenu posljedicu. Kako bi zaključili da C uzrokuje E moramo kontrolirati (*hold fixed*) sve uzročne čimbenike koji su korelirani sa C i uzrokuju E. Situaciju koju želimo izbjeći je ona iz slike 2, u kojoj nepoznati uzroci djeluju na posljedice i onemogućavaju procjenu uzročnog utjecaja C na E. Situacije u kojima uzrok C nije koreliran s drugim relevantnim uzrocima za E su *homogene* situacije s obzirom na sve ostale uzroke. (ibid: 423) C je uzrok E ako i samo ako C povećava vjerojatnost E u svim homogenim situacijama.

U slučaju Simpsonovog paradoksa, kriteriji na temelju kojih dijelimo populaciju moraju biti uzročno relevantni za posljedicu čiji uzrok želimo utvrditi. Uzročne prosudbe temeljimo na već stečenom, pozadinskom znanju koje nam govori da su vjerojatnosti za upis na teži odsjek manje od upisa na lakši odsjek za bilo kojeg prijavljenog studenta. S druge strane, znamo da činjenica je li netko dobar biciklist ne čini nikakvu razliku u upisu na fakultet. Sukladno pozadinskom znanju donosimo uzročne prosudbe i ako veća vjerojatnost upisa za muškarce nestane kad kontroliramo uzročno relevantne varijable, odbacujemo hipotezu kako postoji diskriminacija žena pri upisu na fakultet. (ibid: 433)

[7] Uvođenje nove varijable Eells opravdava misaonim eksperimentom. Zamislimo da fakultet diskirminira žene pri upisu te da su jedne godine svi muškarci na prijavnom listu zaokružili da su žene, dok su žene zaokružile da su muškarci. U tom slučaju, svojstvo „biti muškarac" bi imalo negativan utjecaj na upis. Ipak, namjerna diskriminiacija fakulteta ostaje jer, u tom slučaju, *vjerovanje* fakulteta da je netko muškarac je različito od činjenice da netko posjeduje svojstvo „biti muškarac" (ibid: 64)

2.6. Smjer uzročnosti i kontrastivna uzročnost

Eellsovo rješenje Simpsonovog paradoksa uključivalo je svojstvo tranzitivnosti uzročnih veza. Ako C uzrokuje P, a P uzrokuje E, onda C uzrokuje E posredstvom P. Pretpostavili smo kako je uzročan utjecaj pozitivan, C povećava vjerojatnost da se dogodi P, a P povećava vjerojatnost da se dogodi E. Međutim što ako uzročna veza između C i P nije pozitivna, odnosno, C smanjuje vjerojatnost da se P dogodi, umjesto da je povećava. Pretpostavimo dodatno kako C može imati i direktan pozitivni utjecaj na E. Sličan protuprimjer konstruirao je Hesslow (1976, iz Hitchcock, 2001: 364) Uzimanjem kontracepcijskih pilula povećava se rizik od obolijevanja od tromboze zbog stvaranja krvnih ugrušaka u arterijama. Tromboza je nuspojava uzimanja kontracepcijskih pilula. Uz to, kontracepcijske pilule također imaju izravan negativan utjecaj na trudnoću, odnosno smanjuju vjerojatnost da će žena zatrudnjeti ukoliko uzima kontracepcijske pilule. Međutim, trudnoća također ima pozitivan utjecaj na trombozu i povećava vjerojatnost oboljenja od tromboze. Protuprimjer je ilustriran na slici 5.



Slika 5. Smjer uzročnosti (Hitchcock, 2001)

Kako bi odgovorili na pitanje uzrokuju li kontracepcijske pilule trombozu trebamo razlikovati uzročne veze s obzirom na pojedine *smjerove uzročnosti***.** Ako uzmemo *ukupni* utjecaj kontracepcijskih pilula na trombozu, onda je dovoljno usporediti vjerojatnosti svih mogućih načina na koje žene mogu oboljeti od tromboze. S obzirom da je veća vjerojatnost da kontracepcijske pilule sprječavaju trudnoću nego da uzrokuju trombozu kao nuspojavu, ukupni utjecaj pilula na trombozu je negativan. (Hitchcock, 2001) S druge strane, ako razmotrimo smjer uzročnosti kontracepcijskih pilula na trombozu tako da kontroliramo trudnoću, zaključak je kako pilule povećavaju vjerojatnost obolijevanja od tromboze. Žene imaju veću vjerojatnost da će oboljeti od tromboze ako uzimaju kontracepcijske pilule i zatrudne, nego da ne uzimaju kontracepcijske pilule i zatrudne. Isto tako će vjerojatnost obolijevanja od tromboze biti veća za one žene koje uzimaju kontracepcijske pilule i ne zatrudne od onih žena koje ne uzimaju kontracepcijske pilule i ne zatrudne. (ibid: 375)

Možemo zaključiti kako uzimanje kontracepcijskih pilula ima *ukupni* negativan utjecaj na trombozu i smanjuje vjerojatnost oboljenja, ali da uzimanje kontracepcijskih pilula ima pozitivan utjecaj na trombozu s obzirom na *smjer uzročnosti* koji ne uključuje trudnoću.

Probabilistički pristup možemo dodatno osnažiti ako u analizu uzročnih veza uključimo nekoliko kontekstualnih okolnosti. Uzmimo sljedeći primjer. Pretpostavimo kako se Sherlock Holmes nalazi na dnu litice, dok su Watson i Moriarty na vrhu. Moriarty se sprema gurnuti veliki kamen kako bi ubio Holmesa, no Watson na vrijeme uoči Moriartyjeve namjere i gurne kamen na način da kamen padne što dalje od Holmesa. Međutim, Watson ne uspije dovoljnom jačinom gurnuti kamen i kamen ipak usmrti Holmesa. Iako je Watsonovo guranje kamena uzrokovalo Holmesovu smrt, ono je ipak smanjilo vjerojatnosti da će Holmes poginuti jer bi Moriarty gurnuo kamen puno preciznije da je bio brži od Watsona.

Kako bi riješio problem uzroka koji smanjuju vjerojatnosti posljedica, Hitchcock (1996) predlaže analizu uzročnih veza putem *kontrastivne uzročnosti*. Ideja kontrastivne uzročnosti je da se uzrok i posljedica usporede s određenim alternativama koje su kontekstualno određene. U ovom slučaju Watsonovo guranje kamena je pozitivan uzrok Holmesove smrti u usporedbi sa situacijom u kojoj nitko ne gura kamen. S druge strane, ono je i negativni uzrok koji smanjuje vjerojatnost Holmesove smrti u usporedbi s Moriartyjevim preciznijim guranjem kamena. Ako je W Watsonovo guranje kamena, M Moriartyjevo guranje kamena, N događaj u kojem nitko ne gurne kamen, a H Holmesova smrt, dobivamo sljedeće izraze: (ibid: 274)

(1) $P(H \mid W) > P(H \mid N)$

(2) $P(H \mid W) < P(H \mid M)$

Svrha kontrastivne uzročnosti je da u analizu uzročnih veza uključi, osim uzroka i posljedice, alternativne događaje koji su se mogli dogoditi. Dakle, ako uzrok povećava vjerojatnost posljedici onda, prema kontrastivnoj uzročnosti, tvrdnje o probabilističkim uzročnim vezama moraju biti razmatrane u usporedbi s alternativama koje su se mogle dogoditi u danom kontekstu.

3. BAYESOVE MREŽE

3.1. Osnove Bayesovih mreža: Usmjereni aciklički graf i distribucija vjerojatnosti

Bayesove mreže su grafički modeli za reprezentaciju distribucije vjerojatnosti i zaključaka koji se mogu izvesti na temelju dostupne distribucije. Termin Bayesove mreže prvi puta je skovao Judea Pearl krajem 1980-tih. Bayesove mreže imaju tri uloge u probabilističkom i statističkom modeliranju:

1) Pružiti prikladne načine izražavanja supstancijalnih pretpostavki
2) Olakšati ekonomičnu reprezentaciju zajedničkih distribucija vjerojatnosti
3) Olakšati učinkovito zaključivanje na temelju opažanja. (Pearl, 2009)

Ovdje nas primarno interesira način na koji Bayesovim mrežama možemo reprezentirati uzročne veze te kako na temelju distribucije vjerojatnosti izvesti zaključke o uzročnom utjecaju između varijabli. Bayesove mreže zadovoljavaju polazišni definiciju probabilističke uzročnosti koju smo usvojili u prvom dijelu. Definicija glasi da C je uzrok E ako C povećava vjerojatnost da se dogodi E. Distribucija vjerojatnosti u Bayesovim mrežama sukladna je aksiomima teorije vjerojatnosti i svi teoremi izvedeni u odjeljku 2.3. vrijedit će i ovdje.

Bayesove mreže sastoje se od dva osnovna dijela: *usmjerenog acikličkog grafa* i *distribucije vjerojatnosti*. Oba dijela se primjenjuju na polazišni skup varijabli **V** koje smo dobili empirijskih mjerenjem. Varijable u **V** mogu biti diskretne – poput spola ili stupnja obrazovanja – ili kontinuirane – poput dohotka ili mase -, zbog jednostavnosti koristit ćemo samo diskretne varijable u primjerima.

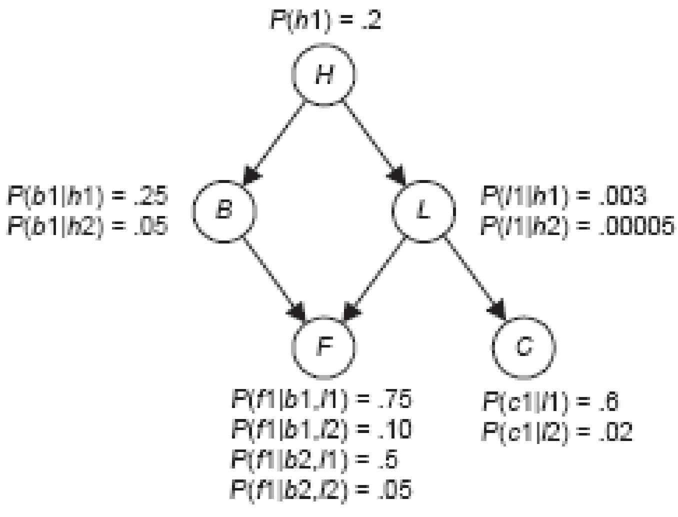
(1) *Usmjereni aciklički graf (directed acyclic graph - DAG)*. Usmjereni aciklički graf G služi za (I) reprezentaciju varijabli i (II) modeliranje njihovog međuodnosa. Graf se sastoji od skupa *čvorova (nodes)* koji označuju varijable u **V**, te skupa *veza* E *(links/edges)* koje povezuju jednu varijablu s drugom i označuju uzročni utjecaj jedne varijable na drugu. Sve veze u grafu ne moraju nužno biti usmjerene prema nekoj varijabli, neke veze mogu ostati neusmjerene kad nemamo dovoljno informacija o uzročnoj vezi, dok neke veze mogu biti dvostruko usmjerene. Ako su sve veze usmjerene, tada je i G usmjeren, ako nijedna veza nije usmjerena onda je graf skelet G-a. [8] G je aciklički u smislu da jedna varijabla ne može uzročno utjecati na samu sebe (X→X). Usmjereni grafovi mogu imati usmjerene cikluse (X→Y, Y→X) koji označuju uzajamno uzrokovanje X i Y, ali onda nisu više aciklički grafovi.

[8] Grafovi bez usmjerenih veza između čvorova nazivaju se još i Markovljevim mrežama (Ben-Gal, 2007)

(2) *Distribucija vjerojatnosti*. Distribucija vjerojatnosti se primjenjuje na varijable iz skupa **V** te se izražava u obliku P(X = x). X je slučajna varijabla u skupu **V**, dok je x numerička vrijednost varijable X. Funkcijom vjerojatnosti P(X = x) označavamo kolika je vjerojatnost da slučajna varijabla X iz **V** ima određenu vrijednost x. Distribuciju vjerojatnosti za Bayesove mreže izražavamo na dva načina: graničnom vjerojatnosti (*marginal probability*) – vjerojatnost jedne varijable, P(X = x) - i zajedničkom distribucijom vjerojatnosti (*joint probability distribution*) – uvjetne vjerojatnosti svih varijabli u G.

Odnosi između čvorova u Bayesovim mrežama izražavaju se koristeći terminologiju rodbinskih odnosa. Ako imamo vezu X→Y onda je X *roditelj* Y, a Y *dijete* X. Jedan čvor može imati više roditelja ako više čvorova ima uzročni utjecaj na njega, primjerice u strukturi *zajedničke posljedice* (*common effect*) X→Y←Z, gdje su X i Z roditelji Y. Jedan roditelj može također imati više djece, kao što je bio primjer sa strukturom zajedničkog uzroka. *Obitelj* nekog čvora je taj čvor i skup svih roditelja tog čvora. Čvorovi bez roditelja nazivaju se *korijeni (root)* i obično se nalaze na vrhu Bayesove mreže. Skup *potomaka* čvora X je skup svih čvorova do kojih se može doći slijedeći usmjerene veze iz X. Ako imamo strukturu X→Y→Z←W, skup potomaka od X je {Y, Z}, ali nije W jer ne postoje usmjerene veze od X prema W. Skup *predaka* čvora X je skup svih čvorova putem kojih se usmjerenim vezama može doći do X. U prethodnom primjeru, skup predaka od Z je skup {X, Y, W} jer se od svakog čvora može doći do Z usmjerenim vezama. Usmjereni aciklički graf u kojem svaki čvor ima najviše jednog roditelja je *stablo*, dok se stablo u kojem svaki čvor ima najviše jedno dijete naziva *lanac*. (Ben-Gal, 2007; Pearl, 2009)

Nakon što smo definirali osnovne dijelove i pojmove Bayesovih mreža možemo na primjeru jedne Bayesove mreže ilustrirati dosad definirane pojmove.



Slika 6. Bayesova mreža (Neapolitan, 2004: 4)

Čvorovi ove Bayesove mreže su slučajne varijable H, B, L, F i C te zajedno čine skup **V**. Čvorovi su povezani vezama, u su ovom slučaju sve usmjerene, koje označavaju uzročne veze ili utjecaje jedna varijable na drugu. Primjerice, čvor H – povijest pušenja – ima izravan uzročni utjecaj na varijable B i L – bronhitis (B) i rak pluća (L) -, a neizravan uzročan utjecaj na F - umor – posredstvom B i L. Uzročne veze izražene su distribucijom uvjetne vjerojatnosti. Sve varijable u ovoj Bayesovoj mreži su binarne, netko može imati povijest pušenja, H = h1, ili nemati povijest pušenja, H = h2. Uvjetne vjerojatnosti izražavaju vjerojatnost da određena varijabla ima neku vrijednost ako je roditelj od te varijable imao neku drugu vrijednost. Tako P(b1 | h2) = 0.05 kaže kako je vjerojatnost da će doći do bronhitisa ukoliko ne postoji povijest pušenja 0.05 ili 5%. Uvjetna vjerojatnost može biti izražena i kada postoje dvije varijable na temelju kojih procjenjujemo vjerojatnost treće. Primjerice, umor je ovisan o tome ima li netko bronhitis i o tome ima li rak pluća. Vjerojatnost da će doći do umora s obzirom da je prisutan rak pluća, a nema bronhitisa, izraženo je kao P(f1 | b2, l1) i iznosi 0.5 ili 50%.

Možemo primijeniti rodbinsku terminologiju kako bi imenovali odnose između čvorova. H je roditelj od L i B, dok čvor F ima dva roditelja, L i B. H je korijen grafa jer nema niti jednog roditelja. Skup predaka od čvora F je skup {L, B, H} jer se od svakog čvora u skupu usmjerenim vezama može doći do F . C ima samo jednog roditelja, L, ali ima dva pretka, L i H. Potomci od L su ujedno i njegova djeca, C i F, jer C i F nemaju djece. Kad bi ih imali tada bi djeca od C i F ušli u skup potomaka od L.

3.2. Markovljev uvjet

Veze između čvorova u Bayesovim mrežama označili smo kao uzročne veze ili uzročne utjecaje čvora roditelja na čvor djeteta. Međutim, sama struktura Bayesovih mreža i pripadajuća distribucija vjerojatnosti nije dovoljna kako bi usmjerene veze između čvorova bile ujedno i uzročne veze. Kako bi došli te pretpostavke, moramo najprije definirati uvjete koji opravdavaju unošenje uzročnih veza u Bayesove mreže.

Potreban uvjet je tzv. Markovljev uvjet i odnosi se na distribuciju vjerojatnosti P primijenjenu na skup varijabli **V**. Kažemo da distribucija vjerojatnosti P na skupu varijabli **V** zadovoljava Markovljev uvjet ukoliko se odnosi prema grafu G sukladno sljedećoj definiciji.

Markovljev uvjet. *Pretpostavimo da imamo distribuciju vjerojatnosti P slučajnih varijabli u skupu* ***V*** *i da je DAG G = (****V****, E). Kažemo da (G, P) zadovoljava Markovljev uvjet ako za svaku varijablu X ϵ* ***V,*** *{X} je uvjetno nezavisan sa skupom svih svojih ne-potomaka s obzirom na skup svih svojih roditelja.* (Neapolitan, 2004: 31)

Markovljevim uvjetom želimo reći kako roditelji od X *zaslanjaju* X od svih ostalih varijabli osim od potomaka od X. Drugim riječima, X je uvjetno nezavisan od svih varijabli koji nisu potomci od X s obzirom na skup roditelja od X. Markovljev uvjet možemo izraziti na sljedeći način:

$$\text{(MU) } P(X_1, X_2, \ldots, X_n) = \prod i \; P(X_i \mid \mathbf{PA}\,(X_i))$$

Na lijevoj strani imamo zajedničku distribuciju vjerojatnost svih varijabli u skupu **V**. Markovljevim uvjetom definirali smo kako je vjerojatnost svake varijable nezavisna o varijablama ne-potomcima s obzirom na varijable roditelje. Zajedničku distribuciju vjerojatnosti računamo umnoškom uvjetne vjerojatnost svake varijable s obzirom na njene roditelje.

Markovljev uvjet bitno pojednostavljuje pravilo multiplikacije u izračunavanju zajedničkih distribucija vjerojatnosti. Pretpostavimo da imamo 4 varijable i želimo izračunati zajedničku distribuciju vjerojatnosti, prema pravilu multiplikacije dobili bi sljedeće:

$$P(X_1 \cap X_2 \cap X_3 \cap X_4) = P(X_1)\, P(X_2 \mid X_1)\, P(X_3 \mid X_2 \cap X_1)\, P(X_4 \mid X_3 \cap X_2 \cap X_1)$$

Ovisno o tome kakvi su odnosi između varijabli u grafu, vjerojatnost varijable možemo izraziti kao uvjetnu vjerojatnost s obzirom na njene roditelje, zanemarujući sve pretke varijable, što nije slučaj kod pravila multiplikacije.

Markovljevim uvjetom izrazili smo uvjetnu nezavisnost neke varijable sa skupom njezinih ne-potomaka s obzirom na njene roditelje. Pogledajmo ponovno primjer Bayesove mreže na slici 6, na kojoj uvjetne vjerojatnosti već zadovoljavaju Markovljev uvjet. Varijabla H je korijen grafa te ona nema roditelje. Varijable B i L možemo izraziti kao P(B = b | H = h), odnosno P(L = l | H = h). Na varijablama F i C vidimo kako možemo reducirati pravilo multiplikacije koristeći Markovljev uvjet. Umjesto pisanja svih predaka i ne-potomaka, uvjetne vjerojatnosti F i C izražavamo kao P(F = f | B = b ∩ L = l), odnosno P(C = c | L = l). Tablica uvjetnih nezavisnosti varijabli s obzirom na njihove roditelje izgleda ovako:

Čvor; X	Roditelj; **PA** (X)	Uvjetna nezavisnost; ⊥
C	{L}	{C} ⊥ {H, B, F} | {L}
B	{H}	{B} ⊥ {L, C} | {H}
F	{B ∩ L}	{F} ⊥ {H, C} | {B, L}
L	{H}	{L} ⊥ {B} | {H}

Tablica 2. Uvjetne nezavisnost prema Markovljevom uvjetu (Neapolitan, 2004: 33)

Zajednička distribucija vjerojatnosti koja zadovoljava Markovljev uvjet izgledala bi ovako:

$$P(f, c, b, l, h) = P(f \mid b, l)\, P(c \mid l)\, P(b \mid h)\, P(l \mid h)\, P(h)$$

Kako bi dobili uzročne Bayesove moramo nadopuniti Markovljev uvjet. Markovljev uvjet govori samo o uvjetnim nezavisnostima unutar grafova, ali ne kaže eksplicitno ništa o uzročnim vezama. Dobivanje uzročnih veza između varijabli oslanja se na ekspertno znanje stručnjaka u određenom području. Uzročne veze ne se mogu deducirati ili iščitati samo na temelju distribucije vjerojatnosti i uvjetnih nezavisnosti između varijabli, već moraju biti dodatno unesene temeljem prethodnog znanja o uzročnoj strukturi svijeta. Konstruiranje uzročnih Bayesovih mreža zahtijeva tri pretpostavke o uzročnosti: (1) pojam izravne uzročnosti je odnos između dvije ili više varijable, (2) uzročni graf na skupu **V** bit će aciklički te (3) uzročni graf će zadovoljavati Markovljev uvjet s obzirom na distribuciju vjerojatnosti koju imamo. (Williamson, 2005: 49-50)

Pomoću posljednjeg uvjeta možemo formulirati uzročni Markovljev uvjet. Ukoliko distribucija vjerojatnosti varijabli u **V** zadovoljava Markovljev uvjet u grafu G možemo pretpostaviti kako su veze između varijabli uzročne. Putem Markovljevog uvjeta varijabla je uvjetno nezavisna od svih svojih ne-potomaka s obzirom na svoje roditelje. Uzročnim Markovljevim uvjetom zahtijevamo da varijable ne budu samo u odnosu uvjetne nezavisnosti, već da budu u uzročno-posljedičnom odnosu. Analogno Markovljevom uvjetu, uzročni Markovljev uvjet kaže kako je svaka varijabla uvjetno nezavisna sa svim svojim ne-posljedicama s obzirom na svoje izravne uzroke. (ibid)

Uzročni Markovljev uvjet opravdan je ako su ispunjena tri uvjeta: (1) ne smije biti skrivenih uzroka, (2) ne smije biti selekcijske pristranosti i (3) nijedna varijabla ne smije biti uzrok samoj sebi (*no feedback loops*). (Neapolitan, 2004: 55) Uzmimo prvi uvjet koji je identičan načelu zajedničkog uzroka. Uvjet kaže kako ne smije postojati korelacija bez uzročnosti, odnosno da svaka zavisnost između dvije varijable mora biti uzročno objašnjena. Uzročno objašnjenje korelacije znači ili da jedna varijabla uzrokuje drugu ili da postoji zajednički uzrok koji uzrokuje obje varijable.

3.3. d-Odvojenost

d-Odvojenost (*d-separation*) je direktna posljedica Markovljevog uvjeta i tablice uvjetnih nezavisnosti skiciranih u prethodnom poglavlju. Pretpostavimo da imamo varijable X, Y i Z i njihove distribucije vjerojatnosti koje odgovaraju čvorovima u grafu G. Ako želimo provjeriti jesu li X i Y uvjetno nezavisni s obzirom na Z u bilo kojoj distribuciji kompatibilnoj sa G, moramo osigurati da uvjetna nezavisnost X i Y bude vidljiva na grafu G. Drugim riječima, čvor Z u grafu G mora „blokirati" sve putove koji vode iz čvora X prema čvoru Y. Čvor Z mora spriječiti tijek informacija od čvora X prema čvoru Y. Ako je put od X prema Y blokiran, kažemo da su X i Y d-odvojeni.

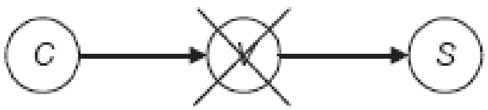
Kao što je uvjetna vjerojatnost svojstvo dobiveno na temelju distribucije vjerojatnosti, d-odvojivost je analogno svojstvo usmjerenih acikličkih grafova. Ako graf G i distribucija vjerojatnosti P zadovoljavaju Markovljev uvjet, tada je i svaka d-odvojenost u G ujedno i uvjetna nezavisnost u P. (ibid: 70) d-Odvojenost možemo definirati na sljedeći način:

d-Odvojenost. *Put p je d-odvojen (ili blokiran) skupom čvorova Z akko:*

(1) p sadrži lanac i → m → j ili i ← m→ j u kojem je srednji termin m u skupu Z

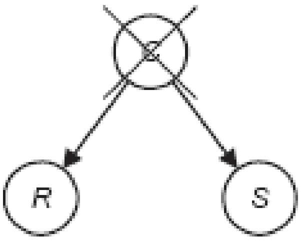
(2)p sadrži i → m ← j u kojoj srednji čvor m nije u skupu Z te nijedan potomak od m nije u skupu Z. (Pearl, 2009: 16-7)

Prvi uvjet sadrži relacije koje su poznate iz rasprave o probabilističkom pristupu uzročnosti. Lanac *i → m → j* je primjer tranzitivnosti uzročnosti, tj. ako *i* uzrokuje *m*, *m* uzrokuje *j*, onda *i* uzrokuje *j*. Ono što definicija d-odvojivosti kaže je da ukoliko opazimo *m* ili znamo vrijednost *m*, tada *i* i *j* postaju uvjetno nezavisni s obzirom na *m*. Uvjetovanje na *m* blokira tijek informacija od *i* prema *j* što ih čini nezavisnima. Dobivamo ovakav put:



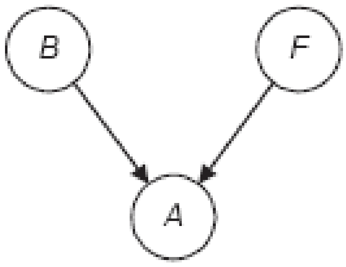
Slika 7. d-Odvojenost (Neapolitan, 2004: 56)

Jednaki je slučaj s drugim članom u prvom uvjetu. U ovom slučaju imamo strukturu zajedničkog uzroka koja je već elaborirana u prvom dijelu rada. Ako znamo koja je vrijednost *m* i uvjetujemo *i* i *j* s obzirom na *m*, *i* i *j* postaju uvjetno nezavisni. Uvjetovanjem na zajednički uzrok blokiramo tijek informacija iz posljedica prema drugim posljedicama i tako sve posljedice postaju uvjetno nezavisne s obzirom na zajednički uzrok.



Slika 8. d-Odvojenost zajedničkog uzroka (ibid: 57)

Drugi uvjet d-odvojenosti je nešto drugačiji. U ovom slučaju imamo strukturu zajedničke posljedice u kojoj *i* i *j* zajedno uzrokuju posljedicu *m*. Ako pogledamo veze koje označavaju uzročni utjecaj i tijek informacija kroz čvorove, vidjet ćemo kako nema interakcije između dva uzroka. Svaki uzrok vodi do zajedničke posljedice, ali ne do drugog uzroka. Tek ako opazimo zajedničku posljedicu ili znamo koju ona vrijednost ima, dva uzroka će postati uvjetno zavisna. Ako znamo da se dogodio sudar na cesti (posljedica) čiji su uzroci jedino mogli biti gužva na cesti i olujno nevrijeme, te dvije alternative su nam potencijalna objašnjenja sudara na cesti. Ako procijenimo kako je mala vjerojatnost da je bilo olujno nevrijeme, vjerojatnost da je bila gužva na cesti raste.[9] Dakle, *i* i *j* su d-odvojeni ako nemamo informaciju o *m* i put od *i* prema *j* je već blokiran.



Slika 9. d-Odvojenost zajedničke posljedice (ibid: 58)

[9] Ovaj obrazac poznat je kao Berksonov paradoks. Primjerice, ako je za upis na fakultet potrebno imati vrlo visoke ocjene ili jedinstveni glazbeni talent, onda će ta dva svojstva biti negativno korelirana u studentskoj populaciji, dok će u cijeloj populaciji biti nezavisni i nekorelirani. Studenti s lošim ocjenama će posjedovati jedinstveni glazbeni talent, dok će glazbeno netalentirani studenti imati visoke ocjene. Obje alternative objašnjavaju upis studenata na fakultet. (Pearl, 2009: 17)

3.4. Ockhamova britva i uvjet minimalnosti

Ockhamova britva ili načelo parsimonije poznato je normativno načelo u izboru teorija iz filozofije znanosti. Ono kaže da ako imamo dvije teorije koje na jednako konzistentan i adekvatan način opisuju i objašnjavaju skup empirijskih podataka, uvijek trebamo izabrati teoriju koja je sintaktički jednostavnija i postulira manje teorijskih entiteta. S obzirom da ne možemo odabrati teoriju samo na temelju empirijskih podataka – jer su dvije teorije empirijski ekvivalentne – favorizirat ćemo onu teoriju koja ima manje teorijskih pretpostavki.

Slično načelo možemo postaviti i pri izboru adekvatne Bayesove mreže koja grafički modelira empirijski dobivenu distribuciju vjerojatnosti. Pretpostavimo da imamo skup **V** = {X, Y} i graf G u kojem X uzrokuje Y, tj. X→Y. Dostupna distribucija vjerojatnosti P nam kaže kako su X i Y nezavisni jedno o drugome. Graf G bi zadovoljavao Markovljev uvjet jer nemamo restrikcija koje nameće Markovljev uvjet za nezavisnost varijabli djece od ostalih ne-potomaka s obzirom na roditelje. Međutim, pretpostavimo kako imamo graf G' koji se sastoji od istih varijabli kao i G, ali nema vezu od X prema Y, tj. X ne uzrokuje Y u G'. (Hitchcock, 2012) Koji graf od ta dva odabrati s obzirom na dostupnu distribuciju vjerojatnosti?

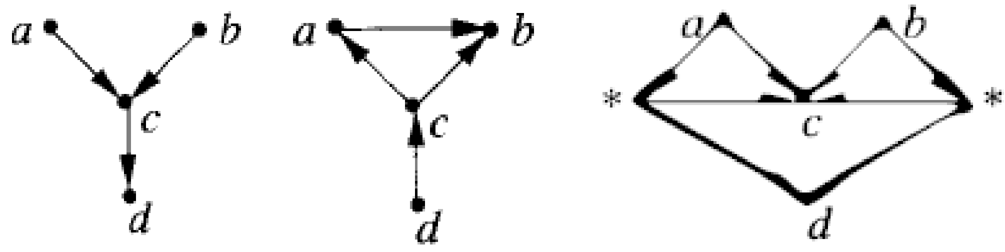
Uvjet minimalnosti ograničava gomilanje grafova koji bi bili konzistentni s distribucijom vjerojatnosti, ali međusobno različiti. Kao i kod znanstvenih teorija, preferiraju se oni grafovi koji su jednostavniji i zahtijevaju manje neopravdanih pretpostavki. Graf je *minimalan* onda kada se ne može naći jednostavniji graf koji bi bio jednako konzistentan s distribucijom vjerojatnosti. Graf G je minimalan ako ne postoji graf G' koji je podskup grafa G i koji zadovoljava Markovljev uvjet.

Uvjet možemo definirati na sljedeći način:

Uvjet minimalnosti. *Latentna struktura L = (D, O), gdje je D uzročna struktura na skupu* ***V****, a O skup opaženih varijabli O ⊆* ***V****, je* minimalna *s obzirom na skup* ***L*** *latentnih struktura akko ne postoji član u* ***L*** *koji se strogo preferira od L – tj., akko za svaki L' ϵ* ***L*** *imamo L = L' kadgod se L' preferira nad L.* (Pearl, 2009: 46)

Nužan uvjet u zaključivanju uzročnog utjecaja varijable C na drugu varijablu E, s obzirom na distribuciju vjerojatnosti P, je da postoji usmjerena veza iz C u E u svakoj minimalnoj latentnoj strukturi konzistentnoj s distribucijom P. Uvjet minimalnosti sprječava pretpostavljanje „skrivenih“ uzroka koji nisu prisutni na distribuciji P te putem kojih bi se uzročna struktura između opaženih varijabli mogla arbitrarno razmještati.

Uzmimo za primjer sljedeća tri grafa s četiri opažene varijable {a, b, c, d} i skrivenom varijablom *. Pretpostavimo kako na temelju opažanja imamo distribuciju vjerojatnosti koja otkriva samo dvije nezavisnosti: (1) a ⊥ b i (2) d ⊥ {a, b} | c.



Slika 10. Uvjet minimalnosti (Pearl, 2009: 47)

Od triju grafova jedino prvi graf zadovoljava uvjet minimalnosti jer jedini modelira opažene nezavisnosti i ne pretpostavlja nikakve druge nezavisnosti i uzročne strukture koje nisu vidljive u distribuciji vjerojatnosti. Drugi graf neopravdano postavlja uzročne veze od a prema b koje nisu zadane distribucijom vjerojatnosti. Uzročna veza od d prema c je također redundantna s obzirom na prvi graf jer je potrebna samo uzročna veza od c prema d kako bi graf bio konzistentan s distribucijom. Obje uzročne veze nisu nužne s obzirom na zadanu distribuciju. Treći graf ne zadovoljava drugi uvjet jer d nije nezavisan sa {a, b} s obzirom na c te je konzistentan s distribucijama u kojima su d i {a, b} zavisni s obzirom na c. Uz to, treći graf uvodi najviše skrivenih uzroka koji nisu opažljivi već moraju biti pretpostavljeni. (ibid)

3.5. Uvjet vjernosti / stabilnosti

Uvjet stabilnosti ili vjernosti zahtijeva da sve bezuvjetne i uvjetne nezavisnosti u distribuciji vjerojatnosti P nad **V** budu prisutne u grafu G koji zadovoljava Markovljev uvjet. S druge strane, Markovljev uvjet u G mora implicirati nezavisnosti iz dostupne distribucije P. Uvjet izražava dvostranu vezu između distribucije P i Markovljevog uvjeta u grafu G. Iz grafa G se ne smiju moći deducirati nikakve nezavisnosti koje nisu prisutne u distribuciji P, dok Markovljev uvjet mora biti dovoljno precizan da preslika sve nezavisnosti iz distribucije P u graf G. Zato kažemo da graf G *vjerno* preslikava distribuciju vjerojatnosti te ujedno ne implicira nikakve nove nezavisnosti.

> **Uvjet vjernosti.** *Pretpostavimo kako imamo zajedničku distribuciju vjerojatnosti P slučajnih varijabli u* ***V*** *i graf G =* ***(V,*** *E). Kažemo da (G, P) zadovoljava uvjet vjernosti ako, temeljem Markovljeva uvjeta, G povlači samo one uvjetne nezavisnosti u P. To jest, sljedeća dva uvjeta moraju vrijediti:*
>
> *(1) (G, P) zadovoljava Markovljev uvjet, tj. G povlači samo uvjetne nezavisnosti u P*
>
> *(2) Sve uvjetne nezavisnosti u P su implicirane putem G temeljem Markovljeva uvjeta.*

(Neapolitan, 2004: 95)

Pearl (2009: 48) koristi termin uvjet stabilnosti. Prema njemu, uvjet kaže kako su sve nezavisnosti u distribuciji vjerojatnosti P stabilne, što znači da graf G i uzročna struktura u G moraju implicirati sve obrasce uvjetne nezavisnosti u P, neovisno o varijacijama parametara. Uzmimo za primjer binarnu varijablu C koja ima vrijednost 1 ako je ishod bacanja dva novčića (A, B) jednak, a vrijednost 0 ako je različit. Jedina uzročna struktura u grafu G koja je stabilna je struktura zajedničke posljedice, A → C ← B. Struktura zajedničke posljedice je jedina koja vjerno održava obrazac uvjetne nezavisnost u primjeru bacanju novčića jer je neosjetljiva na bilo koje promjene parametara kao što su pristrani novčići s vjerojatnostima različitim od 0.5 za svaki ishod.

Neovisno o tome koliko mi mijenjali parametre u grafu G, niti jedna uvjetna nezavisnost u P ne smije biti narušena. Dakle, P je stabilna distribucija G-a ako vjerno preslikava uzročnu strukturu u G, tj. ako

$$(X \perp Y \mid Z)_P \leftrightarrow (X \perp Y \mid Z)_G$$

Odnos između uvjeta minimalnosti i uvjeta stabilnosti možemo ilustrirati na Pearlovom primjeru slike koja prikazuje stolicu. Imamo dvije teorije koje opisuju sliku i trebamo se odlučiti za jednu.

T_1: Predmet na slici je stolica

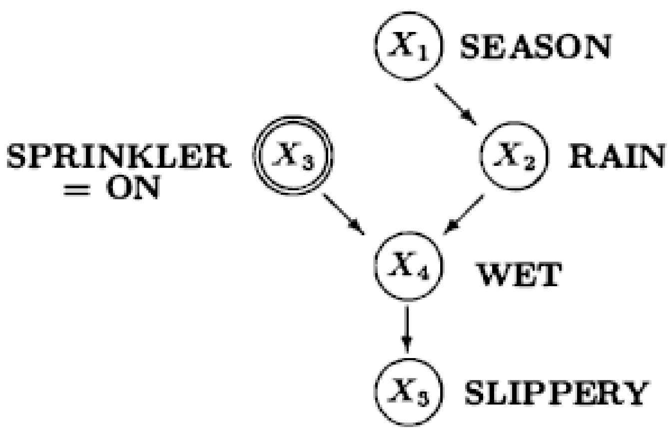
T_2: Predmet na slici je stolica ili dvije stolice posložene jedna iza druge tako prednja stolica zakriva stražnju

Očigledno je kako se trebamo odlučiti za T_1. Prema uvjetu minimalnosti, T_1 je bolje rješenje jer je skup scena koje uključuju samo jedan predmet pravi podskup scena koje uključuju dva ili manje predmeta. Ukoliko nemamo razloga vjerovati da se na slici ne nalazi ništa osim jedne stolice, trebamo preferirati T_1. Prema uvjetu stabilnosti (ili vjernosti), T_1 je također bolje rješenje jer je vjerojatnost da su dva predmeta posložena tako savršeno da prednji zaklanja drugi skoro nikakva. Takav raspored stolica bio bi također vrlo nestabilan s obzirom na razne promjene u okolnostima. Primjerice, varijacija u kutu gledanja potpuno bi opovrgnula T_2 čak i ako bismo zanemarili vrlo male početne vjerojatnosti.

3.6. *Intervencije u uzročnim Bayesovim mrežama*

Kako bi konstruirali uzročnu Bayesovu mrežu nije dovoljno pretpostaviti kako ona zadovoljava Markovljev uvjet. Uzročni mehanizmi koju su reprezentirani putem Bayesove mreže moraju biti stabilni s obzirom na intervencije i manipulacije koje radimo na uzrocima. Ukoliko želimo procijeniti je li uključivanje prskalice uzrok mokre trave, moramo uključiti prskalicu dovoljan broj puta i opažati rezultira li uzrok pretpostavljenom posljedicom. Primjer za takav oblik testiranja uzročnih mehanizama su klinička ispitivanja ili istraživanja provedena u eksperimentalnim uvjetima. Svi uvjeti koji mogu utjecati na uzročno-posljedičnu vezu moraju biti kontrolirani kako bi osigurali da nije bilo intervenirajućih faktora. Najčešće se to postiže nasumičnim odabirom uzorka kako ne bi bilo selekcijske pristranosti te kako bi se uzročni mehanizam mogao ispitati u izolaciji od drugih faktora.

Sličnu situaciju možemo ilustrirati i putem Bayesovih mreža. Ako postoji stvarna uzročna veza između prskalice i mokre trave, X_3 i X_4 na slici 11, onda svaki puta kada interveniramo na X_3 uključivši prskalicu, dobit ćemo X_4, mokru travu. Jedini je uvjet da imamo pod kontrolom ostale uzroke od X_4, u ovom slučaju X_2, kišu. Ako pada kiša, neovisno o tome uključili li mi prskalicu ili ne, trava će biti mokra.



Slika 11. Uzročna Bayesova mreža (Pearl, 2009: 23)

Cilj je konstruirati Bayesovu mrežu koja će vjerno reprezentirati sve uzročne veze koje su relevantne za reprezentirane posljedice. Samo na temelju distribucije vjerojatnosti ne možemo zaključiti koje uzročne veze djeluju između varijabli. Distribucija vjerojatnosti nam kaže kolika je vjerojatnost da će se neki događaja dogoditi ili kolika je vjerojatnost da će se neki događaj dogoditi ako se već dogodio neki drugi događaj. Vjerojatnosti nam također daju informaciju o varijaciji vjerojatnosti nekog događaja s obzirom na varijacije vjerojatnosti događaja koji su zavisni s tim događajem. Međutim, jedino putem uzročnih veza možemo procijeniti koliko bi se vjerojatnosti nekog događaja promijenile ukoliko bi napravili vanjsku intervenciju negdje unutar uzročne strukture. (Pearl, 2009)

Pretpostavimo da manipuliramo X_3 i uključimo prskalicu kao na slici 11. Manipuliranjem X_3 brišemo vezu $X_1 \rightarrow X_3$ i pripisujemo X_3 pripadajuću vrijednost. Bilo koja uzročna veza koja je postojala između X_1 i X_3 više nije bitna jer znamo koja je vrijednost X_3. Stoga brišemo $P(x_3 \mid x_1)$ iz formule zajedničke distribucije vjerojatnosti jer nas više ne zanima vjerojatnost X_3. S obzirom da je X_3 binarna varijabla, vrijednost varijable je X_3 = On. Imamo sljedeću zajedničku distribuciju vjerojatnosti:

$$P_{X3 = On}(x_1, x_2, x_4, x_5) = P(x_1)\, P(x_2 \mid x_1)\, P(x_4 \mid x_2, X_3 = On)\, P(x_5 \mid x_4)$$

Sad možemo i definirati uzročne Bayesove mreže.

Uzročne Bayesove mreže. *Ako je $P(v)$ distribucija vjerojatnosti na skupu varijabli **V** i $P_X(v)$ distribucija nakon intervencije $do(X = x)$ koja određuje podskup varijabli X na konstante x. $\boldsymbol{P_*}$ je skup svih intervencijskih distribucija $P_X(v)$, $X \subseteq \mathbf{V}$, uključujući $P(v)$ koji reprezentira stanje bez intervencije ($X = \emptyset$). Graf G je* uzročna Bayesova mreža *kompatibilna sa $\boldsymbol{P_*}$ akko sljedeća tri uvjeta vrijede:*

(1) $P_X(v)$ zadovoljava Markovljev uvjet s obzirom na G

(2) $P_X(v) = 1$ za sve $V_i \in X$ kadgod je v_i konzistentan sa $X = x$

(3) $P_X(v_i \mid pa_i) = P(v_i \mid pa_i)$ za sve $V_i \notin X$ kadgod je pa_i konzistentan sa $X = x$, tj. svaki $P(v_i \mid pa_i)$ ostaje invarijantan prema intervencijama koje ne uključuju V_i. (ibid: 24)

Distribucija vjerojatnosti nakon intervencije na varijabli X može se izračunati, sukladno prethodnoj definiciji, koristeći sljedeću formulu:

$$P_X(v) = \prod_{\{i \mid Vi \notin X\}} P(v_i \mid pa_i)$$

Intervencijom dobivamo novu distribuciju koju izračunavamo uzimajući u obzir vrijednost varijable X na kojoj je napravljena intervencija. Ukoliko želimo provjeriti ima li varijabla X uzročni utjecaj na drugu varijablu Y, nužno je da Y bude potomak od X. Novu distribuciju granične vjerojatnosti za Y, nakon intervencije do(X = x), označavamo $P_X(y)$ i izračunavamo za svaku vrijednost x za varijablu X. Ako vjerojatnost Y ostaje ista nakon intervencije nad X, možemo zaključiti kako X nema uzročni utjecaj nad Y. Pearlov je zaključak kako su uzročne veze stabilnije od probabilističkih, one imaju *ontološki* karakter i otkrivaju nešto o postojećim uzročnim mehanizmima u prirodi. Probabilističke veze imaju primarno *epistemički* karakter te reflektiraju naša vjerovanja o procesima u prirodi. (ibid)

Razmotrimo sljedeća dva iskaza sukladna slici 11. S1: „Uključivanje prskalice neće utjecati na kišu" i S2: „Stanje prskalice je nezavisno i nepovezano sa stanjem kiše". S2 će biti istinit ako znamo koja je vrijednost X_1, tj. koje je godišnje doba, ali nemamo nikakvu informaciju o tome koje je godišnje doba, tada je S2 neistinit. S2 će također biti neistinit ako znamo da je trava mokra (X_4) jer ćemo u tom slučaju imati strukturu zajedničke posljedice u kojoj su kiša (X_3) i prskalice (X_2) uvjetno zavisne s obzirom na X_4. S druge strane, S1 će biti istinit neovisno o našim opažanjima i modifikacijama uvjetnih nezavisnosti. Čak i da modificiramo bilo koju uzročnu vezu u grafu na slici 11, S1 će još uvijek biti istinit jer prskalica uzrokuje jedino mokru travu, ali nema uzročni utjecaj na kišu. Uzročne veze su stoga robusnije od čisto probabilističkih veza jer ostaju invarijantne s obzirom na sve moguće promjene u uzročnim vezama u grafu na slici 11. (ibid: 25)

Intervencije na neku varijablu manipuliranjem njezine vrijednosti razlikuju se od opažanja vrijednosti koje ta varijabla ima prema prirodnom tijeku stvari. Ukoliko opazimo kako su prskalice uključene, modificirat ćemo uvjetne vjerojatnosti ostalim varijablama u grafu u skladu s distribucijom vjerojatnosti. Uvjetne vjerojatnosti nekih varijabli će se povećati, nekih će se smanjiti, ali nećemo znati je li promjena isključivo rezultat činjenice da su prskalice uključene ili neke druge promjene unutar sustava. Intervencijom, s druge strane, namjerno uvodimo promjenu u sustav i manipuliramo varijablom kako bi ustanovili koji je njezin uzročni utjecaj na ostale varijable. Varijabla na kojoj smo izvršili intervenciju uzročno će utjecati na varijable potomke koje leže na njezinom uzročnom putu. Sukladno Markovljevom uvjetu, varijable potomci od varijable na kojoj je izvršena intervencija bit će uvjetno nezavisne od ostalih varijabli s obzirom na varijablu na koju se intervenira. Brisanjem veze $X_1 \rightarrow X_3$ u primjeru na slici 11 varijabla na kojoj je izvršena intervencija postaje novi korijen grafa, odnosno ona nema roditelja jer ljudska intervencija manipulira njezinom vrijednošću, a ne zavisnost koju ima o varijablama precima.

4. JESU LI BAYESOVE MREŽE DOVOLJNE ZA UZROČNOST?

4.1. Argumenti za – sveobuhvatnost i objektivnost Bayesovih mreža

Nakon što smo skicirali osnove Bayesovih mreža trebamo pokušati odgovoriti na pitanje pružaju li Bayesove mreže zadovoljavajući formalni model uzročnosti. Nakon što se Humeova teorija uzročnosti pokazala kao neadekvatna teorija uzročnosti, probabilističkim pristupom uzročnosti ukazali smo na moguća rješenja na probleme koji su bili nerješivi za Humeovu teoriju. Bayesove mreže poslužile su kao adekvatan način formalizacije probabilističke koncepcije uzročnosti. Međutim, možemo li putem Bayesovih mreža i probabilističke uzročnosti na kojima se baziraju zahvatiti sve aspekte uzročnih veza koje se pojavljuju u znanosti? Jesu li Bayesove mreže same po sebi dovoljne za analizu uzročnih veza? Ima li možda kakvih nekonzistentnosti u prenošenju probabilističke uzročnosti na modele Bayesovih mreža? Filozofi su na ova pitanja odgovarali dvojako. Najprije ću razmotriti argumente koji idu u prilog Bayesovim mrežama kao potpunim opisima uzročnih veza u prirodi i znanosti.

Spohn (2001) tvrdi kako za uzročnu zavisnost posljedice E o uzroku C trebamo samo konstruirati dovoljno preciznu Bayesovu mrežu i suzdržati se od metafizičkih interpretacija uzročnih veza. Prema njemu, pojam uzročne zavisnosti posljedice o uzroku ovisi o referentnom okviru (*frame relative*) onoga koji promatra uzročnu vezu. Na makroskopskoj razini možemo uočiti uzročnu vezu između dva fenomena, međutim to je samo rezultat referentnog okvira unutar kojega tumačimo uzročnost. (Spohn, 2001: 7) Ako neke ekonomske fenomene reduciramo na kemijske fenomene, teško ćemo naći uzročne veze jer su fenomeni opisani i objašnjeni na drugačiji način.

Drugi problem je problem na koji način interpretirati vjerojatnost. Je li ona objektivno svojstvo svijeta ili subjektivna mjera neizvjesnosti? Ako prihvatimo objektivističku interpretaciju, vjerojatnosti su šanse ili sklonosti pojedinačnog predmeta da se ponaša na određen način. Međutim, takva interpretacija nam ne dozvoljava da adekvatno analiziramo uzročnost jer podrazumijeva da će uzrok „djelomično determinirati" posljedicu, što zahtijeva da smo već upoznati s pojmom „potpune determinacije" koji nije jasan. S druge strane, subjektivna interpretacija vjerojatnosti vodi k daljnjoj relativizaciji uzročnih veza jer ih promatra na način kako su one *spoznate* od strane epistemičkog subjekta. (ibid: 8)

Neovisno o nerješivim interpretacijama vjerojatnosti i mutnom pojmu uzročne zavisnosti, tvrdi dalje Spohn, ipak možemo uspješno modelirati uzročnost putem Bayesovih mreža koje zadovoljavaju Markovljev uvjet. Ako možemo napraviti skicu uzročnog grafa koji još nije Bayesova mreža, to ne znači kako ne postoji rafiniranija verzija grafa koji bi bio Bayesova mreža i zadovoljavao Markovljev uvjet. Razlog zašto uvijek možemo konstruirati rafiniranije uzročne grafove poput Bayesovih mreža je taj što ne postoji *nezavisan* pojam uzročnosti, uzročne veze su samo one veze koje možemo modelirati putem Bayesovih mreža.

> Struktura je adekvatno rafiniranih Bayesovih mreža ona koja odlučuje o tome kakve su uzročne zavisnosti. *Ne možemo smatrati da je B uzročno ovisan o A osim ako ne možemo naći niz veza ili usmjerenih veza koji vode od A prema B u adekvatno rafiniranoj Bayesovoj mreži, te osim ako taj posao ne možemo napraviti narednim rafiniranjem.* (ibid: 10, kurziv autorov)

Ako se pomirimo s činjenicom da je pojam uzročne zavisnosti ovisan o referentnom okviru, to još uvijek ne znači kako je pronalaženje uzročnih veza neostvarivo. Ukoliko uključimo sve uzročno relevantne varijable u Bayesovu mrežu koja će sveobuhvatno opisati stvarnost, zavisnost o referentnom okviru nestaje. U toj idealnoj verziji Bayesove mreže subjektivne vjerojatnosti ne bi bile distribuirane na arbitraran način jer će subjektivne vjerojatnosti biti identične objektivnim vjerojatnostima koje postoje u prirodi. Na taj način eliminirat će se relativizacija uzročnosti ovisna o referentnom okviru ili subjektivnim procjenama. (ibid: 11)

Relativizacija uzročnih veza može se zaobići bez pozivanja na ideal-tipsku Bayesovu mrežu ili neizvjestan proces daljnjeg rafiniranja uzročnih grafova dok ne zadovolje Markovljev uvjet. Možemo prihvatiti objektivističku interpretaciju vjerojatnosti i tvrditi kako se subjektivna mjera vjerojatnosti mora uskladiti sa objektivno mjerljivim vjerojatnostima. Na taj način uzročnost bi bila objektivno svojstvo u prirodi i relativizacija uzročnih veza ne bi bila moguća. Williamson (2005) nudi dva uvjeta objektivnog bayesijanizma koji trebaju dodatno ograničavati stupnjeve subjektivnog vjerovanja.[10]

(1) Empirijski uvjet. Informacije o svijetu moraju ograničavati stupnjeve vjerovanja. Ako kocka pada na broj šest frekvencijom 1/3, onda je i vjerojatnost da će pasti šest 1/3

(2) Logički uvjet. Nedostatak informacija o svijetu mora ograničavati stupnjeve vjerovanja. Ako o eksperimentu jedino znam da ima pet mogućih ishoda, svakom ishodu moram dodijeliti vjerojatnost 1/5. (ibid: 66)

[10] Subjektivni bayesijanizam, s druge strane, pati od problema da dva subjekta mogu pridati potpuno različite mjere vjerovanja istom događaju, bez obzira na empirijske informacije

4.2. Argumenti protiv – protuprimjeri Markovljevom uvjetu i neuzročnost probabilističkih zavisnosti

Većina argumenata protiv Bayesovih mreža kao adekvatnih modela za analizu uzročnih veza fokusira se na načelo zajedničkog uzroka. Reichenbachovo načelo zajedničkog uzroka je osnova uzročnog Markovljeva uvjeta koji kaže da ukoliko postoji zavisnost između dvije varijable u Bayesovoj mreži, onda je jedna uzrok druge ili postoji zajednički uzrok obje varijable. Već smo u raspravi o probabilističkoj uzročnosti istaknuli kako načelo zajedničkog uzroka ima svoja ograničenja. Argumenti protiv uzročnog Markovljeva uvjeta sličnog su karaktera kao i argumenti protiv načela zajedničkog uzroka. Fokus je uglavnom na preusku domenu primjene uzročnog Markovljevog uvjeta.

Uzročni Markovljev uvjet podložan je istim ograničenjima kao i načelo zajedničkog uzroka. Osim već spomenutih ograničenja, možemo navesti još par ograničenja u kojima uzročni Markovljev uvjet nije zadovoljen. Uzmimo za primjer pomiješane populacije. Pretpostavimo kako imamo populaciju žirafa i nosoroga. Žirafe su više od nosoroga, ali nosorozi imaju bolji vid. Stoga ćemo u toj populaciji imati pozitivnu korelaciju između dobrog vida i visine. Međutim, razvojni proces svake vrste dogodio se potpuno neovisno o razvojnom procesu druge vrste. Možemo pretpostaviti kako je pripadnost populaciji žirafa ili nosoroga zajednički uzrok visoke korelacije visine i dobrog vida, ali kontraintuitivno je tvrditi kako je pripadnost populaciji pravi uzrok dobrog vida ili visine. (Hitchcock, 2012)

Još jedan primjer u kojem uzročni Markovljev uvjet nije zadovoljen je kad u slučaju događaja koji su neprecizno opisani. Uzmimo za primjer igru biljara. Ako igrač pukne bijelu kugu (I) u crnu kuglu, crna loptica ima određenu vjerojatnost da će ući u rupu. Pretpostavimo kako su kugle poredane na takav način da crna kugla uđe u rupu (C) ako i samo ako bijela kugla uđe u suprotnu rupu (B) od one u koju je ušla crna kugla. Događaji da će dvije kugle ući u suprotne rupe su pozitivno korelirani. Imamo sljedeće vjerojatnosti: $P(B \leftrightarrow C) = 1$, $P(C \mid I) = 1/2$, ali $1/2 = P(B \mid I) \neq P(B \mid I \cap C) = 1$. Dakle, zajednički uzrok I ne uspijeva učiniti B i C uvjetno nezavisnima. (Williamson, 2005: 56)

Cartwrightova (2007) priznaje kako je probabilistički pristup utjelovljen u Bayesovim mrežama dobar putokaz prema uzročnim vezama, ali nedovoljan za pronalaženje uzročnih veza. Cartwright osporava robusnost uvjeta vjernosti (stabilnosti) tvrdeći kako probabilistička zavisnost C i E ne implicira da je C uzrok E, već samo da je C jedan od mogućih uzroka E. Najbolji primjer je Simpsonov paradoks, u kojem se inicijalna probabilistička zavisnost preokreće ukoliko podijelimo populaciju. Uzročne veze, prema Cartwright, možemo detektirati jedino ukoliko kontroliramo sve ostale relevantne faktore unutar populacije koji povećavaju vjerojatnost posljedici. Drugi primjer u kojem Bayesove mreže ne uspijevaju modelirati uzročne veze su slučajevi u kojima jedan uzrok može imati dva suprotna učinka na posljedicu. Ako su oba učinka jednake jačine, oni se mogu međusobno poništiti te uzrok može imati neutralan učinak na posljedicu, što znači kako nema probabilističke zavisnosti između uzroka i posljedice. (ibid: 64-5) Primjer za to je dvostruki utjecaj kontracepcijskih pilula na trombozu, u jednom smjeru uzročnosti pilule povećavaju vjerojatnost tromboze, dok u drugom smjeru uzročnosti pilule smanjuju vjerojatnost tromboze posredstvom smanjenja vjerojatnosti trudnoće koja opet povećava vjerojatnost tromboze. Uzrok proizvodi posljedicu putem različitih smjerova uzročnosti. Uzmimo za primjer da je uzrok zaraza neke osobe virusom HIV-a, a posljedica smrt te osobe. Posljedica je probabilistički zavisna o uzroku. Ali informacija o zavisnosti ne govori na koji se konkretan način uzročni utjecaj manifestira. Iako zaraza virusom HIV-a najčešće dovodi do smrti, ta činjenica nam ne govori u kojem vremenu nakon zaraze dolazi do smrti. Zaraza virusom HIV-a također dovodi do slabljenja imunološkog sustava, što može rezultirati smrću putem različitih bolesti, poput atrofije mozga ili meningitisa. Virus se također može prenijeti različitim uzročnim putovima poput seksualnog odnosa ili transfuzijom krvi. Sve te okolnosti su dio konkretnog smjera uzročnosti, činjenica da je posljedica probabilistički zavisna o uzroku ne govori ništa o specifičnom uzročnom lancu koji je omogućio da dođe do posljedice.

Cartwrightin zaključak je da niti uzročne veze nužno ne impliciraju probabilističku zavisnost niti probabilistička zavisnost nužno ne implicira uzročne veze. Iako priznaje kako postoji veza između uzročnosti i probabilističke zavisnosti, ta veza nije čvrsta. Uzroci mogu povećati vjerojatnosti svojim posljedicama isto kao što povećanje vjerojatnosti može biti posljedica uzročnih veza. (ibid: 79) Uzroci *mogu* povećati vjerojatnosti svojim posljedicama, ali to ne znači nužno da će uzroci *uvijek* povećati vjerojatnosti svojim posljedicama. Veza između uzročnosti i vjerojatnosti je poput veze između bolesti i simptoma te bolesti. Bolest može dovesti do tog simptoma, ali ne mora, dok isti taj simptom može biti rezultat neke druge bolesti.

5. ZAKLJUČAK

Cilj rada bio je istražiti probabilistički pristup uzročnosti i formalne modele Bayesovih mreža. Probabilistički pristup uzročnosti nameće se kao optimalan pristup analizi uzročnih veza s obzirom na neplauzibilnost determinizma u kojem uzroci nužno proizvode svoje posljedice. U prvom dijelu rada izložili smo koncepciju probabilističke uzročnosti iz gledišta filozofije znanosti i metafizike. Jedna od najutjecajnijih analiza uzročnosti u povijesti filozofije bila je ona Davida Humea. Nakon što su istaknuti glavni problemi Humeove teorije uzročnosti – lažne korelacije, nesavršene pravilnosti i irelevantnost –, vidjeli smo kako se probabilistički pojam uzročnosti uspješno nosi sa svim problemima. Naglasak je bio na rješavanju problema lažnih korelacija, ne samo iz razloga što ga probabilistički pristup uspješno rješava, već i zbog toga što se problem vrlo često javlja u znanosti. Prije nego što smo pristupili analizi lažnih korelacija, izložili smo formalizam teorije vjerojatnosti – tri aksioma i mnoštvo korisnih teorema za rigorozniju analizu uzročnosti. Načelo zajedničkog uzroka, uz neka ograničenja, pokazalo se kao učinkovito sredstvo u dolaženju do uzročnih veza na temelju korelacija. Nešto zamršeniji problem bio je Simpsonov paradoks u kojem je pozitivna korelacija pretvorena u negativnu samo na temelju podjele populacije na manje dijelove. Predložena su dva rješenja, Eellsovo rješenje putem načela zajedničkog uzroka i Cartwrightino rješenje putem kontroliranja svih relevantnih uzroka unutar populacije. Na kraju prvog dijela riješili smo nekoliko protuprimjera probabilističkoj uzročnosti.

U drugom djelu izložen je formalni model Bayesovih mreža kao način reprezentacije uzročnih veza. Nakon što smo definirali osnovne sastavnice Bayesovih mreža – usmjereni aciklički graf i distribuciju vjerojatnosti – izložili smo par uvjeta koji moraju biti zadovoljeni kako bi Bayesove mreže što vjernije reprezentirale odnose u stvarnosti. Markovljev uvjet najvažniji je uvjet koji mora biti zadovoljen kako bi se konstruirale Bayesove mreže i tiče se uvjetnih nezavisnosti između varijabli u grafu Bayesove mreže. Dodatna dva uvjeta su uvjet minimalnosti i uvjet vjernosti (ili stabilnosti). Na kraju drugog dijela ukratko je opisan postupak intervencije, način putem kojeg se najučinkovitije mogu iščitati uzročne veze iz Bayesove mreže.

Posljednji dio rada nastojao je označiti snage i nedostatke Bayesovih mreža kao modela za pronalaženje uzročnih veza. Neki su tvrdili kako su Bayesove mreže sveobuhvatan prikaz uzročnih veza, dok su drugi naglasili kako probabilističke zavisnosti nisu dovoljne za uzročne veze. Glavni argumenti u korist Bayesovih mreža kao adekvatnih modela za pronalaženje uzročnih veza oslanjali su se na njihovu potencijalnu sveobuhvatnost i objektivnost. S jedne strane, metafizičke spekulacije o uzročnoj zavisnosti nisu urodile plodom i pružile jedinstveni pojam uzročne zavisnosti posljedice o uzroku. Poimanje uzročne zavisnosti je ovisno o referentnom okviru s kojega se promatraju uzročne veze. Ako uzmemo da je referentni okvir neka znanstvena paradigma, pogled molekularnog biologa i razvojnog psihologa na uzročne veze neće biti identičan. Uzročna ovisnost je također subjektivne naravi, dva epistemička subjekta mogu pripisati različite subjektivne mjere vjerovanja istoj uzročnoj vezi. Jedini izlaz iz relativizacije uzročnosti je modeliranje putem Bayesovih mreža. Idealne Bayesove mreže trebali bi ukloniti sve subjektivne pristranosti i eliminirati metafizičke spekulacije o uzročnim ovisnostima. Čak i ako trenutačno nemamo pristup idealiziranim Bayesovim mrežama, to ne znači kako ne možemo doći do njih jednom kada dodatno rafiniramo uzročne grafove. S druge strane, možemo prihvatiti objektivističku interpretaciju vjerojatnosti te zahtijevati kako se sve mjere vjerojatnosti moraju uskladiti s objektivno mjerljivim svojstvima svijeta. U tom slučaju, Bayesove mreže bi reflektirale objektivna svojstva svijeta, a ne subjektivne mjere vjerovanja.

Argumenti protiv Bayesovih mreža naglašavali su nedostatke uzročnog Markovljeva uvjeta. Prvi skup prigovora nastojao je naći što više protuprimjera uzročnom Markovljevom uvjetu kako bi ograničio domenu primjene Bayesovih mreža. Svi protuprimjeri koji su istaknuti protiv načela zajedničkog uzroka vrijede i u ovom kontekstu. Drugi skup prigovora, poglavito kritike Nancy Cartwright, nastojao je oslabiti vezu između uzročnosti i probabilističke zavisnosti. Niti uzročne veze nužno ne impliciraju probabilističke zavisnosti, niti probabilističke zavisnosti nužno ne impliciraju uzročne veze. Uzročne veze nužno ne impliciraju probabilističke zavisnosti u primjerima kada uzrok djeluje na posljedicu posredstvom dva različita smjera uzročnosti te se dva smjera u konačnici mogu poništiti. Probabilističke zavisnosti, s druge strane, nužno ne impliciraju uzročne veze u slučajevima lažnih korelacija s dva uzroka. U svakom slučaju, jednostrana veza između probabilističkih zavisnosti i uzročnih veza ne postoji.

BIBLIOGRAFIJA:

- Ben-Gal, Irad (2007), „Bayesian Networks“, u F. Ruggeri et al. (ur.), *Encyclopedia of Statistics in Quality and Reliability*, Wiley & Sons
- Bertsekas, Dimitri i Tsitsiklis, John (2008), *Introduction to Probability*, Athena Scientific, Belmont
- Cartwright, Nancy (1979), "Causal laws and effective strategies", *Nous* 13/4, 419-437
- Cartwright, Nancy (2007), *Hunting Causes and Using Them*, Cambridge University Press, Cambridge
- Eells, Ellery (1991), *Probabilistic Causality*, Cambridge University Press, Cambridge
- Hitchcock, Christopher (1996), "Farewell to Binary Causation", *Canadian Journal of Philosophy* 26, 267-282
- Hitchcock, Christopher (2001), "A Tale of Two Effects", *Philosophical Review* 110, 361-396
- Hitchcock, Christopher (2012), "Probabilistic causation", E. Zalta (ur.), *Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/causation-probabilistic/
- Howson, Colin i Urbach Peter (2005), *Scientific Reasoning: The Bayesian Approach*, 3rd ed., Open Court, Chicago/La Salle
- Hume, David (1988), *Istraživanje o ljudskom razumu*, Naprijed, Zagreb
- Loux, Michael (2010), *Metafizika: Suvremeni uvod*, Hrvatski studiji, Zagreb
- Malinas, Gary i Bigelow, John (2012), "Simpson's paradox", u E. Zalta (ur.), *Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/paradox-simpson/
- Morris, William Edward i Brown, Charlotte (2014), "David Hume", u E. Zalta (ur.), *Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/hume/#PriAss
- Neapolitan, Richard (2004), *Learning Bayesian Networks*, Prentice Hall, New Jersey
- Pearl, Judea (2009), *Causality: Models, Reasoning and Inference*, 2nd ed., Cambridge University Press, Cambridge
- Psillos, Stathis (2009), "Regularity Theories", u H. Beebee et al. (ur.), *Oxford Handbook of Causation*, Oxford University Press, Oxford, 131-157
- Reichenbach, Hans (1956), *Direction of Time*, University of California Press, Berkeley/Los Angeles

- Sober, Elliott (1987), „Principle of common cause“, u J. H. Fetzer (ur.), *Probability and Causality: Essays in Honor of Wesley Salmon,* Reidel, 158-174
- Sober, Elliott (2001), „Venetian Sea Levels, British Bread Prices and the Principle of Common Cause“, *British Journal for the Philosophy of Science* 52/2, 331-346
- Spohn, Wofgang (2001), „Bayesian Nets are all there is to Causal Dependence“, u D. Consantini et al. (ur.), *Stochastic Causality*, CSLI Publications, Stanford
- Suppes, Patrick (1970), *A Probabilistic Theory of Causality*, North-Holland Publishing Company, Amsterdam
- Williamson, Jon (2005), *Bayesian Nets and Causality: Philosophical and Computational Foundations,* Oxford University Press, Oxford