# Što je teorija uzročnosti?

Aleksandar Hatzivelkos[1]

## 1. Uvod

Jedan od prvih koncepata koje naučimo razvijajući se kao ljudska bića jest koncept uzroka i posljedice. Već kao djeca znamo da će određene radnje rezultirati specifičnim posljedicama. Taj nas koncept od malih nogu prati kroz život, a posebno kroz školovanje. Ne samo da u školi znamo da će slabo znanje uzrokovati lošu ocjenu, pa potom i reakciju roditelja, nego je i materija koju učimo na svim razinama vezana uz koncept uzroka i posljedice. U fizici učimo kako je pad objekta na zemlju puštenog s neke visine posljedica djelovanja sile teže. U povijesti učimo kako je jedan povijesni događaj posljedica onoga koji mu je prethodio, te uzrok nekom događaju koji slijedi nakon njega. U logici učimo forme zaključivanja kojima na pravilan način zaključujemo kada iz nekog uzroka slijedi neka posljedica, itd.

Iako smo zapravo svakodnevno u kontaktu s uzrocima i posljedicama, te iako bez problema ustvrđujemo kako kiša uzrokuje vlagu pločnika, ili kako gibanje Mjeseca uzrokuje plimu i oseku, zanimljivo je da teorija uzročnosti – što uzrok čini uzrokom, a posljedicu posljedicom, te kako na apstraktnoj razini opisati funkcionalne odnose između ta dva pojma – svoju formalnu definiciju i strukturu pronalazi tek u današnje vrijeme, tj. u posljednjih pedeset godina. Što je, dakle, uzrok, a što posljedica?

### 1.1 Što je uzrok?

Uzročnost je relacija između dvaju događaja, uzroka i posljedice, u kojoj se drugi događaj smatra posljedicom prvog. U široj upotrebi uzročnošću nazivamo i relaciju između skupa faktora (uzroka) i nekog fenomena (posljedice). Sve što utječe na posljedicu je faktor tog događaja. Uzroke dijelimo na neposredne i posredne. Dok neposredni uzroci neposredno djeluju na neki fenomen ili događaj, posredni uzroci to čine kroz neke druge faktore. Iako svi ovi navodi djeluju samorazumljivo, formalno definiranje uzročno-posljedičnih veza kroz stoljeća okupira znanstvenu i filozofsku raspravu.

[1]Aleksandar Hatzivelkos, Zagreb

David Hume prvi je u 18. stoljeću sustavno pokušao objasniti koncept uzroka i posljedice, navodeći osam kriterija čije je ispunjavanje potrebno da bismo govorili o uzročnoj vezi. Među tim kriterijima navodi se i sljedeći: „*Uzrok možemo definirati kao objekt, nakon kojeg slijedi drugi, gdje nakon svih objekata sličnih prvome, slijede objekti slični drugome*" [3]. Ti rani pokušaji analiziranja uzročnosti demonstriraju regulatorni pristup: ako *A* uzrokuje *B*, tada nakon *A* *uvijek* slijedi *B*, pa je uzročna veza invarijantna. S vremenom su artikulirani ozbiljni problemi regulatornog pristupa uzročnosti:

1. *Nesavršene regularnosti*. Vodeći se Humeovom definicijom, morali bismo isključiti velik broj općeprihvaćenih uzročnosti. Primjerice, morali bismo odbaciti tvrdnju da „pušenje uzrokuje rak pluća", budući da postoje pojedinci koji puše, a nisu oboljeli od raka pluća. Ili, da krenemo u još izraženiji ekstrem, možemo ustvrditi da čovjekov slobodan pad iz aviona ne uzrokuje smrt jer postoje slučajevi preživljavanja takvog pada. Regulatorna, deterministička pozicija može se braniti nepotpunim ili nedovoljnim poznavanjem faktora koji determiniraju uzročno-posljedični slijed, no postignuća na području kvantne fizike dodatno su uzdrmala povjerenje u determinizam. Stoga postoji zahtjev za formulacijom teorije uzročnosti koja neće apriori pretpostavljati determinističku prirodu procesa.

2. *Irelevantnost*. Određeni događaj može uvijek rezultirati istim efektom, iako je sam taj događaj nebitan za efekt koji slijedi. Primjerice, nakon dodavanja natrijevog heksametafosfata soli, sol se svaki put topi kada se stavi u vodu. Dakle, jedan je događaj redovito praćen drugim. Pa ipak, dodavanje natrijevog heksametafosfata nije uzrok topivosti soli u vodi jer proces dodavanja nema nikakvog utjecaja na topivost soli.

3. *Asimetrija*. Ako *A* uzrokuje *B*, uobičajeno *B* neće također uzrokovati *A*. Jedan od načina uvođenja asimetričnosti u uzročnost je propisivanje da uzroci vremenski dolaze prije posljedica. No, takvo postuliranje nailazi na svoje kontraprimjere u kvantnoj fizici (u kojoj se uzroci i posljedice događaju istodobno). Također, postuliranje asimetrije pomoću vremenskog slijeda, zbog cirkularnosti definicije onemogućava upotrebu uzročnosti kao temelj za fundiranje vremenskog slijeda. Stoga bi bilo poželjno da teorija uzročnosti dade neku argumentaciju za definiranje smjera uzročnosti, a ne da ju jednostavno postulira.

4. *Lažne regularnosti*. Pretpostavimo da nakon nekog uzroka uvijek slijede dvije posljedice. Primjerice, nakon pada pritiska u atmosferi uvijek slijede dva efekta. Prvo, pad visine stupca žive u barometru, te malo kasnije, oluja. Ukoliko pak zanemarimo zajednički uzrok tih dvaju događaja, te promatramo samo njihov međuodnos - pad visine stupca žive u barometru, te oluju koja uvijek slijedi nakon nekog vremena - prema Humeu možemo zaključiti kako pad visine stupca žive *uzrokuje* oluju. To, naravno, nije tako, stoga je važno da teorija uzročnosti u obzir uzme egzistenciju zajedničkog uzroka kao pojave koja može pružiti lažnu sliku uzročnosti.

## 1.2 Filozofske interpretacije

Danas smo u situaciji da još ne postoje stabilni i općeprihvaćeni aksiomi i teorija uzročnosti, budući da svaka od teorija ima nedostataka i kontraprimjera koji joj priječe opću prihvaćenost i primjenjivost na sve interpretacije. No, postoji nekoliko dobro utemeljenih koncepata koji, uz razumijevanje njihovih ograničenja, daju solidnu bazu za široku upotrebu. Tri su osnovna smjera razmišljanja o uzročnosti.

Prvo, postoji smjer razmišljanja koji zagovara isključivu heurističku upotrebu koncepta uzročnosti, koja kao takva ne bi trebala biti dio znanstvenog diskursa. Takvu radikalnu poziciju u početku je zauzimao Bertnard Russel koji je smatrao da se znanost bavi funkcionalnim međuodnosima, a ne zakonima uzročnosti. Drugi smjer razmišljanja je teza da je uzročnost bazična znanstvena kategorija koju ne treba posebno razmatrati. Konačno, treći smjer filozofske misli jest taj da se uzročni odnosi mogu reducirati na neke druge koncepte koji ne sadrže uzročno-posljedičnu notaciju. Taj smjer danas je ujedno i najzastupljeniji u filozofskoj literaturi [10].

Prema tom smjeru razmišljanja, postoje tri načina na koje se vrši redukcija. Prvi je *mehanistička* teorija koja uzročne odnose reducira na fizikalne procese. Drugi je *vjerojatnosni* pristup koji uzročne odnose reducira na fizikalne vjerojatnosne procese. Posljednji, *protučinjenični* (*counterfactual*) pristup reducira uzročne odnose na protučinjenične zakone. Osnovni naglasak u ovome tekstu bit će na vjerojatnosnom pristupu interpretiranja uzročnih odnosa, budući da je, unatoč svojim nedostacima, do sada pružio najšire primjenjivu strukturu. No, prije nego li uđemo u tu materiju, opišimo u par rečenica ostale modele razmišljanja.

Mehanističkim pristupom uzročnosti želi se uzročne odnose interpretirati kao način opisivanja fizikalnog procesa koji povezuje ta dva događaja (uzrok i posljedicu). Time se svi uzročni odnosi identificiraju s fizikalnim procesima. Takav je pristup objektivistički: ukoliko se dva agenta ne slažu o prirodi uzročne veze, tada je barem jedan od njih u krivu. Osnovni nedostatak ovakve interpretacije je usko područje njezine primjenjivosti. Primjerice, ako se i prihvati teza da se na taj način mogu interpretirati sve uzročne veze u fizici, pa čak i one između složenih sustava, procesi koji se na taj način opisuju nisu dovoljni za interpretiranje uzročnih veza, primjerice u ekonomiji. Pokušaji daljnje redukcije odnosa u ekonomiji na odnose u fizici nosi svoj paket filozofskih problema (redukcija sustava koji nastaje pod utjecajem čovjeka, odnosno društva, na sustav koji postoji nezavisno od njega), u koje ovom prilikom nećemo ulaziti.

Protučinjenični pristup, koji je detaljno razvio David Lewis, uzročne odnose reducira na subjunktivne kondicionale (u tim se kondicionalima iskazuje uvjet za koji se zna da je lažan ili nevjerojatan). Tada kažemo da *P uzročno ovisi o U* ako i samo ako (i) *ako bi se dogodio U, tada bi se dogodio i P* (ili se vjerojatnost njegovog događanja značajno povećala) i (ii) *ako se U ne bi dogodio, tada se ni P ne bi dogodio* (ili se vjerojatnost njegovog događanja značajno smanjila). Uzročna relacija tada je tranzitivno zatvorenje *uzročne ovisnosti*. Primjer subjunktivnog (protučinjeničnog) kondicionala je rečenica *Da si me nazvao, došao bih*.

Lewis je svoju teoriju protučinjeničnih kondicionala razvio oslanjajući se na notaciju „mogućih svjetova” za koje se podrazumijeva da postoje neovisno o našem svijetu, odnosno na notaciji „bliskosti” između različitih svjetova. No, kako ne postoji stvarna veza između ovog i ostalih mogućih svjetova, tako ne postoji ni temelj za objektivističku interpretaciju takve teorije. S druge strane, uzročne veze ne možemo lako promatrati kao subjektivne.

U novije vrijeme protučinjeničnu teoriju uzročnosti revitalizirao je Pearl, odbacujući Lewisovu notaciju mogućih svjetova, te definirajući protučinjenične kondicionale u terminima „modela strukturalnih jednadžbi” - skupa jednadžbi u kojima je svakoj varijabli dodijeljena vrijednost koja je eksplicitna funkcija drugih varijabli u sustavu. U tom modelu rečenicu *Y bi bio y da je X jednak x* definiramo s: *ako jednadžbu koja trenutno određuje X zamijenimo konstantom X = x, te riješimo sustav jednadžbi za varijablu Y, rješenje koje dobivamo je Y = y* [5]. Značaj protučinjenične teorije uzročnosti (unatoč manjem stupnju praktične primjenjivosti) je taj što daje model kojim se rješavaju problemi koje ostale interpretacije uzročnosti (u prvom redu, vjerojatnosnu) ne mogu riješiti.

## 1.3 Vjerojatnosna interpretacija uzročnosti

Vjerojatnosna interpretacija uzročnosti pokriva šire područje od mehanističke; ideja je da se uzročne veze promatraju u terminima vjerojatnosnih veza između varijabli, bile te varijable iz područja fizike, ekonomije ili neke druge znanosti. Vjerojatnosni pristup uzročnosti eliminira probleme s kojima se susreće regulatorni (deterministički) pristup teoriji (odlomak 1.1).

Neformalno, za događaj $U$ kažemo da (vjerojatnosno) uzrokuje $P$ ako ispunjenje događaja $U$ podiže vjerojatnost ispunjenja događaja $P$, tj. da je

$$p(P\,|\,U) > p(P)\,. \tag{1}$$

Lako je pokazati da je nejednadžba (1) ekvivalentna nejednadžbi

$$p(P\,|\,U) > p(P\,|\,\neg U) \tag{2}$$

koja se uobičajeno uzima za definiciju (naivne) uzročne veze između događaja $U$ i $P$.

Taj se pristup može koristiti kao objašnjenje za nesavršeno poznavanje svih faktora determinističkog sustava, ali i kao teorija koja opisuje uzročne veze kao nedeterminističke po svojoj prirodi. Time je izbjegnut problem nesavršenih regularnosti – budući da povećanje vjerojatnosti posljedice i dalje dopušta slučajeve u kojima se posljedica nije dogodila. Ujedno, riješen je i problem irelevantnosti, budući da irelevantni događaji uopće ne mijenjaju vjerojatnost ishoda posljedice, pa su time diskvalificirani kao mogući uzroci. Problem lažnih regularnosti ovakvom (neformalnom) definicijom nije riješen, kao ni problem asimetrije.

### 1.4 Problem asimetrije

Jedan od ozbiljnijih pokušaja rješavanja problema asimetrije kroz jezik teorije vjerojatnosti - pružio je Reichenbach. Neka su $A$ i $B$ u korelaciji, $p(A \wedge B) > p(A)p(B)$. Neka postoji događaj $C$ za koji vrijedi:

1. $0 < p(C) < 1$
2. $p(A \wedge B \mid C) = p(A \mid C)p(B \mid C)$
3. $p(A \wedge B \mid \neg C) = p(A \mid \neg C)p(B \mid \neg C)$
4. $p(A \mid C) > p(A \mid \neg C)$
5. $p(B \mid C) > p(B \mid \neg C)$

Trojku događaja $A$, $B$, i $C$ Reichenbach naziva *konjuktivnim račvanjem*, i oni zapravo predstavljaju strukturu događaja $A$ i $B$ čija korelacija proističe iz zajedničkog uzroka $C$. Upravo uvjeti (4) i (5) postavljaju događaj $C$ kao uzrok i događaja $A$ i događaja $B$, dok se uvjetima (2) i (3) poznavanje događaja $C$ uklanja korelacija između $A$ i $B$ (uz malo drugačiju notaciju, za te se uvjete u literaturi kaže da $C$ *zaklanja* (*screen off*) $A$ od $B$, i obrnuto). Za događaje $A$ i $B$ koji su u korelaciji, a nisu jedan drugome uzrok, zadovoljavanje ovih pet uvjeta naziva se zadovoljavanjem *principa zajedničkog uzroka* (*Common Cause Principle*).

Reichenbach sada postavlja sljedeću strukturu: ako se $C$ pojavljuje prije događaja $A$ i $B$, te ne postoji događaj koji zadovoljava iste uvjete, a pojavljuje se nakon $A$ i $B$, tada za konjuktivno račvanje kažemo da je *otvoreno prema budućnosti*. Analogno, ako se $C$ pojavljuje nakon događaja $A$ i $B$, te ne postoji događaj koji zadovoljava iste uvjete, a pojavljuje se prije $A$ i $B$, za konjuktivno račvanje kažemo da je *otvoreno prema prošlosti*. Ako pak postoje događaji $C$ i $D$, koji oba zadovoljavaju navedene uvjete, te se jedan pojavljuje prije, a jedan nakon $A$ i $B$, tada četvorku $ABCD$ nazivamo zatvorenim račvanjem. Reichenbachov prijedlog je da se tada smjer (u vremenu) djelovanja uzroka prema posljedici odredi prema predominantnom smjeru otvorenih račvanja. U našem svijetu, primjerice, postoji izrazito mnogo račvanja otvorenih prema budućnosti, a tek par ili nijedna otvorenih ka prošlosti. Reichenbach takvu asimetriju smatra analognom drugom zakonu termodinamike.

### 1.5 *do*-računica

Ozbiljnu kritiku naivne definicije uzročnosti pružio je Pearl. Njegova je tvrdnja da nejednadžba (2) nije izgrađena na intuitivnom sadržaju uzročnosti. Pearlova je teza da se princip „povećanja vjerojatnosti" ne može iskazati jezikom teorije vjerojatnosti, te da bi definiciju trebalo formulirati na sljedeći način:

$$p(P \mid do(U)) > p(P \mid do(\neg U)), \tag{2}$$

gdje *do(U)* predstavlja vanjsku intervenciju koja određuje istinitost događaja U. Nasuprot tome, vjerojatnost *p(P|U)* daje vjerojatnost kao rezultat pasivnog promatranja događaja U, što se rijetko poklapa sa *p(P|*do*(U))*. Takav pristup definiranju uzročnosti Pearl aksiomatizira kao *do*-računicu (*do-calculus*).

Promotrimo tu razliku na slučaju kojim smo opisali lažne regularnosti. U tom je primjeru zapažanje da pad visine stupca žive u barometru uistinu povećava vjerojatnost dolaska oluje – iako znamo da pad visine stupca žive nije uzrok oluje. S druge strane, ukoliko bismo *manipulacijom* nad visinom stupca žive podignuli vjerojatnost dolaska oluje, tada bi se uistinu radilo o uzroku, pa nam Pearlova notacija daje dodatan alat za eliminaciju lažnih regularnosti. Generalno gledano, formuliranje „povećanja vjerojatnosti" unutar *do*-računa eliminiralo je nekoliko problema s kojima se vjerojatnosna uzročnost susrela u posljednjih pedeset godina, uključujući i Simpsonov paradoks [7].

Opišimo na jednom primjeru pojavu Simpsonovog paradoksa u teoriji uzročnosti. Recimo, na primjer, da postoji visoka korelacija između pušenja (*A*) i života na selu (*B*) - za osobe koje žive na selu znatno je veća vjerojatnost da ujedno puše. Recimo također da je život u gradu ( $\neg B$ ), zbog kvalitete zraka i količine ispušnih plinova, također uzrok raka pluća (*C*), čak i u većem postotku od pušenja. Tada, promatrajući populaciju u globalu, podatak da neka osoba puši smanjuje vjerojatnost obolijevanja od raka pluća. Naime, zbog visoke korelacije između pušenja (*A*) i življenja na selu (*B*), visoka je vjerojatnost da je osoba izvrgnuta manjem riziku od raka pluća (kroz pušenje) od osobe koja ne puši a živi u gradu, tj. $p(C \mid A) < p(C \mid \neg A)$.

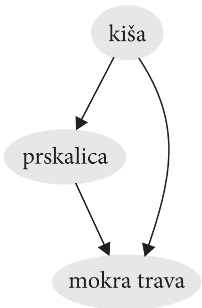
S druge strane, ukoliko promotrimo zasebno svaku populaciju – onu koja živi u gradu, kao i onu koja živi na selu, jasno je da su svi pojedinci unutar te populacije izvrgnuti istom riziku oboljenja od raka pluća koji donosi okoliš u kojemu žive, a da pušači svojim pušenjem taj rizik samo uvećavaju, pa stoga u svakoj podpopulaciji vrijedi obrnuto, $p(C \mid A) > p(C \mid \neg A)$.

Uvedemo li, pak, u definiciju uzročnosti Pearlov *do*-račun, jasno je da se podatak o povećanju (ili smanjenju) vjerojatnosti kroz manipulaciju uzrokom – dakle djelovanjem na aktivnost svakog pojedinca da počne, odnosno prestane pušiti, bez obzira na kontekst i podpopulaciju kojoj pripada – dobiva jednoznačan odgovor na vezu između pušenja i povećanja rizika obolijevanja od raka pluća. Točnije, Pearl u svom radu dokazuje [7] da (uz definiciju uzročnosti kroz *do*-račun) povećanje vjerojatnosti posljedice u svakoj podpopulaciji – rezultira povećanjem vjerojatnosti posljedice i u cijeloj populaciji, čime se rješava Simpsonov paradoks.

## 2. Modeliranje uzročnosti

Pod modeliranjem uzročnosti smatramo proces kojim se na egzaktan način, i uz poznate pretpostavke, može opisati struktura uzročnog djelovanja. U prvom redu, to se odnosi na opisivanje sustava koji imaju neposrednu primjenu, ali i bogatu statističku podlogu za izvođenje zaključaka, poput, primjerice, procesa uzročnosti u medicini, gdje je cilj (kroz niz testiranja ili promatranja) utvrditi korelacije, a na čijem se temelju (uobičajeno, uz neke dodatne pretpostavke) gradi struktura uzročne povezanosti.

Uzročni model sastoji se od skupa varijabli ***V*** i dvije matematičke strukture definirane nad njima. Prva je *aciklični usmjereni graf* (*directed acyclic graph – DAG*) ***G*** nad ***V***: graf s usmjerenim stranicama, ili „strelicama", kojemu su elementi skupa ***V*** vrhovi, pri čemu u grafu ne smiju postojati petlje, kao ni strelice koje iz vrha vode u njega samoga. Druga je *distribucija uvjetnih vjerojatnosti **P*** definirana nad tvrdnjama o vrijednostima varijabli iz skupa ***V***. Varijable u ***V*** mogu predstavljati različite stvari – od binomne varijable koja, primjerice, govori je li uključen stroj za navodnjavanje travnjaka, do realne varijable koja opisuje prihode. Ipak, u svrhu jednostavnosti opisa, za ***V*** se u osnovnoj teoriji uzimaju varijable koje poprimaju diskretan broj vrijednosti.




*Slika 1. Primjer jednostavnog direktnog acikličkog grafa*

Ilustrirajmo opisanu strukturu primjerom. Na Slici 1. prikazan je direktni aciklički graf koji modelira odnose između tri binarne varijable, koje nam pak govore pada li kiša ili ne, je li prskalica uključena, te je li trava mokra. Modelom su opisane opće razumljive uzročnosti; ako pada kiša ili ako je uključena prskalica, tada kao posljedicu očekujemo da je trava mokra, te da – ukoliko pada kiša – nećemo uključivati prskalicu. No, da bi ovaj uzročni usmjereni graf modelirao vjerojatnosne uzročne veze, moramo mu pridružiti i distribuciju vjerojatnosti danih na Slici 2.

**p (prskalica)**

| kiša | prskalica | |
|---|---|---|
| | I | L |
| I | 0,05 | 0,95 |
| L | 0,65 | 0,35 |

**p (kiša)**

| kiša | |
|---|---|
| I | 0,25 |
| L | 0,75 |

**p (mokra trava)**

| prskalica | kiša | mokra trava | |
|---|---|---|---|
| | | I | L |
| I | I | 0,98 | 0,02 |
| I | L | 0,95 | 0,05 |
| L | I | 0,81 | 0,19 |
| L | L | 0,01 | 0,99 |

*Slika 2. Distribucija uvjetnih vjerojatnosti*

U opisivanju strukture grafa, uobičajeno se koriste porodični izrazi. Tako se za vrh $V_i$ kaže da je *roditelj* vrhu $V_j$ ukoliko postoji strelica koja počinje u $V_i$ a završava u $V_j$. Analogno, za vrh $V_j$ kaže se da je *dijete* vrha $V_i$. Ukoliko pak od vrha $V_i$ do vrha $V_j$ postoji usmjereni put, tj. niz strelica od kojih svaka počinje u vrhu u kojem prethodna završava (osim prve strelice koja počinje u $V_i$), a posljednja završava u vrhu $V_j$, tada za vrh $V_i$ kažemo da je *predak* vrha $V_j$, odnosno da je vrh $V_j$ *potomak* vrha $V_i$. Jedino odstupanje od te razumljive terminologije predstavlja definicija po kojoj se svaka varijabla smatra svojim potomkom.

Namjera je grafom nad skupom varijabli opisati uzročnu strukturu među varijablama. Strelica koja vodi iz vrha $V_i$ u vrh $V_j$ predstavlja neposredan uzročni utjecaj varijable $V_i$ na varijablu $V_j$ koji ne ovisi o drugim varijablama. Tada kažemo da je $V_i$ *neposredan uzrok* od $V_j$.[2]

Distribucija vjerojatnosti $\boldsymbol{P}$ definirana je nad propozicijama oblika $X = x$, gdje je $X$ varijabla iz skupa $\boldsymbol{V}$, a $x$ vrijednost u dometu varijable $X$. $\boldsymbol{P}$ je također definirana nad konjukcijama, disjunkcijama i negacijama takvih propozicija. Odavde slijedi da će uvjetne vjerojatnosti nad takvim propozicijama biti dobro definirane kada uvjet ima strogo pozitivnu vjerojatnost. Uobičajena je upotreba pokrate u formi vjerojatnosnih tvrdnji koje sadrže samo varijable ili skupove varijabli, ali ne i njihove vrijednosti. Tako $p(X \mid Y) = p(X)$ koristimo kao pokratu za pisanje izraza

$$\forall x_1,\dots,\forall x_m,\forall y_1,\dots,\forall y_n\left[p(X_1=x_1,\dots,X_m=x_m \mid Y_1=y_1,\dots,Y_m=y_m)=p(X_1=x_1,\dots,X_m=x_m)\right]$$

## 2.1 Markovljev uvjet

Distribucija vjerojatnosti $\boldsymbol{P}$ nad $\boldsymbol{V}$ zadovoljava Markovljev uvjet ako je na sljedeći način povezana s grafom $\boldsymbol{G}$:

$$\forall X \in V\ ,\ \forall Y \subset V \setminus PO(X)\ ,\ p(X \mid RO(X) \wedge Y) = p(X \mid RO(X))\ , \qquad (4)$$

gdje je $PO(X)$ skup svih potomaka, a $RO(X)$ skup svih roditelja varijable $X$. Riječima, kažemo da zadovoljavanjem Markovljevog uvjeta, roditelji svake varijable *zaklanjaju* svaki drugi podskup varijabli, osim potomaka varijable $X$.

Smisao uvjeta je u tome da je poznavanje vrijednosti roditelja varijable $X$ dovoljno kako bi se jednoznačno odredila vjerojatnost da će varijabla $X$ poprimiti vrijednost $x$. Time zapravo tvrdimo da je varijabla $X$ nezavisna od svih ostalih varijabli, osim svojih potomaka, ukoliko joj poznajemo roditelje. Uzročni model, koji se sastoji od acikličkog usmjerenog grafa i distribucije vjerojatnosti koja zadovoljava Markovljev uvjet, nazivamo *uzročnom Bayesovom mrežom*.

[3]S napomenom da je ta tvrdnja relativna s obzirom na skup varijabli $\boldsymbol{V}$; uvođenjem neke nove varijable u model, kojom se detaljnije opisuje interakcija između varijabli $V_i$ i $V_j$, $V_i$ može prestati biti neposredan uzrok.

Dvije neposredne posljedice zadovoljavanja Markovljevog uvjeta važne su za daljnji razvoj teorije. Prva je mogućnost faktorizacije koja omogućava brži izračun vjerojatnosti, čija važnost dolazi do izražaja prilikom modeliranja sustava s više desetaka varijabli.

**Propozicija 2.1** (o faktorizaciji)

Neka distribucija vjerojatnosti ***P*** nad skupom varijabli ***V*** zadovoljava Markovljev uvjet, te neka je ***V***$=\{X_1, X_2, \dots, X_n\}$. Tada vrijedi:

$$p(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} p(X_i \mid RO(X_i)) \,. \tag{5}$$

*Dokaz*: Bez smanjenja općenitosti, varijable u skupu ***V*** možemo indeksirati prema grafu ***G*** na način da prvo indekse dobiju varijable koje nisu ničiji potomci, te da za svake dvije varijable $X_i$, $X_j$, $i < j$, varijabla $X_i$ može biti roditelj ili predak varijabli $X_j$, ali $X_j$ ne može biti ni roditelj ni predak varijabli $X_i$.[3] Krenuvši od formule za uvjetnu vjerojatnost, $p(A, B)=p(A \mid B)p(B)$ iterativno raspisujemo:

$$\begin{aligned} p(X_1, X_2, ..., X_n) &= p(X_2, ..., X_n \mid X_1) p(X_1) \\ &= p(X_3, ..., X_n \mid X_1, X_2) p(X_2 \mid X_1) p(X_1) \\ &= p(X_4, ..., X_n \mid X_1, X_2, X_3) p(X_3 \mid X_1, X_2) p(X_2 \mid X_1) p(X_1) \\ &\partial \\ &= p(X_n \mid X_1, ..., X_{n-1}) \cdot ... \cdot p(X_2 \mid X_1) p(X_1) \,. \end{aligned}$$

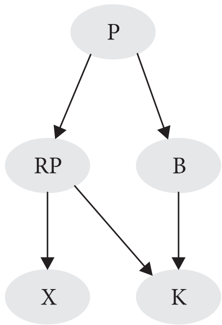
Zadovoljavanje Markovljevog uvjeta sad osigurava da skup $RO(X_i)$ zaklanja sve ostale varijable o kojima uvjetno ovisi $X_i$, iz čega slijedi tvrdnja propozicije.

Ukoliko graf ***G*** ima znatno manje strelica od punog grafa (u kojemu su svake dvije varijable povezane), faktorizacija (5) daje značajno brži izračun vjerojatnosti smanjujući broj potrebnih operacija u izračunu. Koncept se također koristi i kao način aproksimativnog izračuna vjerojatnosti u složenim sustavima sa mnogo varijabli; za zadane statističke podatke modelira se aproksimativan graf, najčešće nekom od «greedy" metoda, poput metode dodavanja strelica [10], ili se pak iz punog grafa metodom brisanja strelica dolazi do aproksimativne mreže [1] iz koje se potom brzo računaju aproksimativne vrijednosti traženih vjerojatnosti. Iako se radi o ekstenzivno upotrebljavanom svojstvu Bayesovih mreža, na ovom mjestu nećemo detaljnije ulaziti u tu temu, već preporučamo daljnje čitanje literature.

[3] Poredak, dakle, ovisi o usmjerenju strelica u grafu i nije jedinstven.

## 2.2 *d*-separacija

Druga posljedica Markovljevog uvjeta na Bayesovim mrežama je egzistencija grafičkog kriterija za određivanje uvjetnih nezavisnosti, koji se naziva *d-separacija*. Radi se o kriteriju koji omogućava uvid u nezavisnost dviju varijabli, uz poznavanje vrijednosti treće (ili skupa trećih). Primjerice, promotrimo uzročnu mrežu danu na Slici 3.




*Slika 3. Uzročna mreža procesa dijagnosticiranja*

U ovom primjeru varijable su binarne i redom označavaju: *P* pušenje, *RP* rak pluća, *B* bronhitis, *K* kašljanje, te *X* pozitivan rendgenski test. Pitanje na koje tražimo odgovor je pitanje uvjetne nezavisnosti varijabli, odnosno koje su varijable nezavisne ukoliko znamo vrijednost neke treće. Za očekivati je, primjerice, da su, uz znanje boluje li osoba od raka pluća, varijable *X* i *K* nezavisne; kako je poznat njihov jedini zajednički uzrok, tako je eliminirana i korelacija između te dvije varijable. Ta se tvrdnja dokazuje iz faktorizacije (5) koju inducira graf sa Slike 3.

Faktorizacija distribucije vjerojatnosti inducirana grafom sa Slike 3. dana je s:

$$p(P,RP,B,X,K) = p(X \mid RP) \cdot p(K \mid RP,B) \cdot p(B \mid P) \cdot p(RP \mid P) \cdot p(P) \qquad (6)$$

Uvjetnu vjerojatnost $p(X, K \mid RP = rp)$ dobivamo iz jednadžbe (6) sumacijom po svim vrijednostima za varijable *P* i *B*, pri čemu varijablu *RP* držimo fiksiranom na vrijednosti *rp*.

$$p(X,K \mid RP = rp) = \sum_{P,B} p(X \mid RP = rp) \cdot p(K \mid RP = rp, B) \cdot p(B \mid P) \cdot p(P)$$

Faktor $p(RP \mid P)$ prestali smo pisati budući da znamo vrijednost koju je poprimila varijabla *RP*, pa je pripadna vjerojatnost ili jednaka nuli (čime cijeli umnožak u traženom zbroju postaje jednak nuli) ili jednaka jedan, čime se pak gubi potreba za pisanjem tog faktora. Izlučivanjem slijedi:

$$p(X,K \mid RP = rp) = p(X \mid RP = rp) \sum_{B} p(K \mid RP = rp, B) \sum_{P} p(B \mid P) \cdot p(P)$$

Unutarnja suma po svim vrijednostima varijable $P$ daje $p(B)$, čime jednadžba prelazi u

$$p(X,K \mid RP=rp) = p(X \mid RP=rp)\sum_B p(K \mid RP=rp,B)\cdot p(B),$$

što sumiranjem po svim vrijednostima varijable $B$ daje

$$p(X,K \mid RP=rp) = p(X \mid RP=rp)\cdot p(K \mid RP=rp).$$

Dakle, varijable $X$ i $K$ uistinu su nezavisne uvjetno na poznavanje vrijednosti varijable $RP$. Takav način dokazivanja uvjetne nezavisnosti varijabli u uzročnom modelu prilično je nezgrapan za upotrebu, pa se pokazala potreba za definiranjem jasnog grafičkog kriterija za utvrđivanje uvjetnih nezavisnosti, koji je nazvan *d*-separacijom.

Definirajmo sada kriterij *d*-separacije. Neka je $\boldsymbol{G}$ usmjeren aciklički graf nad $\boldsymbol{V}$. *Put* u $\boldsymbol{G}$ je niz varijabli $(X_1,\ldots, X_k)$ za koji vrijedi da za svake dvije uzastopne varijable $X_i$ i $X_{i+1}$ postoji strelica iz $X_i$ u $X_{i+1}$, ili strelica iz $X_{i+1}$ u $X_i$. Kažemo da taj put povezuje $X_i$ i $X_k$. Za varijablu $X_i$ kažemo da je *sudar* na tom putu, ako je $i \neq 1, k$, te strelice iz prethodnog člana $X_{i-1}$ i sljedećeg člana $X_{i+1}$ vode u $X_i$. Za varijablu $X_i$ kažemo da je *izvor* na tom putu ako je $i \neq 1, k$, te strelice iz $X_i$ vode ka prethodnom članu $X_{i-1}$ i sljedećem članu $X_{i+1}$. Na Slici 3. varijabla $RP$ je *izvor* na putu od $X$ do $K$, dok je varijabla $K$ *sudar* na putu od $RP$ do $B$.

**Definicija** *(d-separacija)*

Neka su $X,Y \in V$, $Z \subset V$, te neka su $X_i$, $Y_i$ i $M$ varijable na nekom putu od $X$ do $Y$. Skup $Z$ *d*-separira varijable $X$ i $Y$ ako i samo ako za svaki put od $X$ do $Y$ vrijedi barem jedan od sljedećih uvjeta:

1. put od $X$ do $Y$ sadrži $X_i \to Z_i \to Y_i$, gdje je $Z_i \in Z$
2. put od $X$ do $Y$ sadrži $X_i \leftarrow Z_i \leftarrow Y_i$, gdje je $Z_i \in Z$
3. put od $X$ do $Y$ sadrži *izvor* $X_i \leftarrow Z_i \to Y_i$, gdje je $Z_i \in Z$
4. put od $X$ do $Y$ sadrži *sudar* $X_i \to M \leftarrow Y_i$, gdje varijabla $M$ nije element skupa $Z$, niti je ijedan potomak varijable $M$ element skupa $Z$.

Pokazuje se da je kriterijem *d*-separacije dan potpuni grafički kriterij za određivanje nezavisnosti dviju varijabli, uz poznavanje vrijednosti nekog skupa trećih varijabli [5], čime je dokazano da je proučavanje grafičke strukture Bayesove mreže dovoljno za utvrđivanje svih uvjetnih vjerojatnosti u danom modelu.

Vratimo se još jednom primjeru sa Slike 3. Istaknimo uvjetnu nezavisnost između varijabli $RP$ i $B$: između te dvije varijable postoje dva puta; jedan koji prolazi varijablom $P$ i sadrži *izvor*, te drugi koji prolazi varijablom $K$ i sadrži *sudar*. Prema definiciji, tada varijabla *P*, koja je zajednički uzrok varijabli *RP* i *B* *d*-separira te varijable, no skup {*P*, *K*} ih ne *d*-separira. Zaključak je očekivan, budući da poznavanjem vrijednosti zajedničkog uzroka uklanjamo korelaciju između varijabli *RP* i *B*, no ukoliko pored poznavanja vrijednosti *P* poznajemo i vrijednost varijable *K*, koja je zajednička posljedica spomenutih varijabli, tada korelacija nije uklonjena, pa ni varijable nisu uvjetno nezavisne. Zašto je to tako? Uz pretpostavku da i rak pluća i bronhitis pozitivno utječu na kašljanje, poznavanje informacije o kašljanju stavlja vjerojatnosti raka pluća i bronhitisa u (negativnu) korelaciju; ako znamo da osoba ne boluje od raka pluća (a kašlje), tada se povećava vjerojatnost bronhitisa, i obrnuto.

### 2.3 Uzročna mreža

Izgradnjom strukture Bayesove mreže stvoren je temelj za opisivanje uzročnih odnosa, no Bayesova mreža sama po sebi nije dovoljna za jednoznačno definiranje uzročnih odnosa između promatranih varijabli; usmjerena strelica iz varijable *X* u varijablu *Y* u Bayesovoj se mreži ne mora nužno interpretirati kao uzročnu ovisnost varijable *Y* o varijabli *X*. To nam demonstrira jednostavan primjer dviju mreža, mreže $X \to Y \to Z$ i mreže $X \leftarrow Y \leftarrow Z$. Obje te Bayesove mreže, naime, generiraju iste uvjetne nezavisnosti.

Da bi se Bayesova mreža mogla interpretirati kao uzročna mreža, mora postojati eksplicitan semantički zahtjev da se strelice u danoj Bayesovoj mreži interpretiraju kao uzročna ovisnost. Taj je zahtjev semantičke prirode, i kao takav predstavlja intervenciju „izvana". Primjerice, već spomenuti algoritam dodavanja strelica kao rezultat daje više Bayesovih mreža koje ekvivalentno opisuju uvjetne vjerojatnosti među zadanim varijablama. No, da bi se dobiveni rezultat (jednoznačno) interpretirao kao uzročna mreža, potrebna je vanjska intervencija kojom se determiniraju uzročni odnosi među varijablama. U pravilu, vanjska se intervencija temelji na temporalnom slijedu - uzrokom se proglašava varijabla koja prethodi u vremenu (što, naravno, sa sobom nosi filozofske posljedice opisane u poglavlju 1.3), no ponekad ni to nije moguće pa se (semantička) utemeljenost definiranja uzroka mora dodatno argumentirati.

## 3. Zaključak

U ovom tekstu dali smo kratak pregled filozofskih i matematičkih temelja na kojima se gradi teorija uzročnosti. Opisani vjerojatnosni pristup ima svojih ograničenja – primjerice, pomoću njega ne možemo kvalitetno opisati uzročno-posljedične veze između izoliranih, pojedinačnih i neponovljivih slučajeva, kao ni ponuditi odgovore na protučinjenična pitanja, poput *Bi li Titnic potonuo da kapetan kormilo nije prepustio prvom časniku?* U traženju odgovora na to pitanje nije moguće, primjerice, primijeniti *do*-računicu jer ne postoji način da se promijeni istinitost pretpostavljene varijable – danas ni na koji način ne možemo postulirati kapetanov ostanak za kormilom *Titanica* kako bismo saznali posljedice istog.

No, unatoč ograničenjima, vjerojatnosni pristup teoriji uzročnosti dao je snažan matematički aparat za analiziranje uzročnih odnosa u složenim sustavima poput dijagnosticiranja bolesti, ili opisivanja sustava za donošenje odluka. Primjerice, bogata struktura teorije omogućava traženje odgovora na pitanje *identifikabilnosti,* tj. na ocjenjivanje uzročnog efekta u sustavima u kojima nisu poznate vrijednosti svih varijabli. Pored toga, koncept uzročnosti, svojim bogatim sadržajem, otvara prostor za daljnje interpretacije i razvijanje teorije koja nam u konačnici može puno reći o svakodnevnim odnosima u svijetu oko nas.

## Literatura


[1] A. Choi, A. Darwiche *On Bayesian Network Approximation by Edge Deletion*, Proceeding AAAI'06 proceedings of the 21st national conference on Artificial intelligence - Volume 2, stranice 1107-1114, 2006.

[2] C. Hitchcock *Probabilistic Causation*, The Stanford Encyclopedia of Philosophy (Winter 2011. Edition)

[3] D. Hume *An Enquiry Concerning Human Understanding*, Harvard Classics Volume 37, 1910. P. F. Collier & Son., 1748.
\texttt{http://ebooks.adelaide.edu.au/h/hume/david/h92e/}

[4] W. E. May *Knowledge of Causality in Hume and Aquinas*, The Thomist 34.2, 1970.

[5] J. Pearl *Causality: models, reasoning and inference*, Cambridge University Press, 2000.

[6] J. Pearl, J. Tian *On the Identifcation of Causal Effects*, UCLA Cognitive Systems Laboratory, Technical Report (R-290-L), 2003.

[7] J. Pearl *Simpson's paradox: An anatomy*, UCLA Cognitive Systems Laboratory, Technical Report (R-264), November 1999.
\texttt{http://ftp.cs.ucla.edu/pub/stat_ser/R264.pdf}

[8] J. F. Sowa *Processes and Causality*, 2000.
\texttt{http://www.jfsowa.com/ontology/causal.htm}

[9] M. Valtorta *Identifiability in Causal Bayesian Networks: A Gentle Introduction*, Cybernetics and Systems: An International Journal, Volume 39, Issue 4, 2008.

[10] J. Williamson *Bayesian nets and Causality: Philosophical and Computational Fundations*, Oxford University Press, Oxford, 2005.