

Kandidaatintutkielma
Kausaalipäättely valikoitumisharhan korjaamisessa

Riku Laine
Valtiotieteellinen tiedekunta
Helsingin yliopisto

14. toukokuuta 2019

Sisältö

1	Kiitokset – Acknowledgements	2
2	Johdanto	3
2.1	”Kausaalipäättely uutena paradigmana”	3
2.2	Valikoitumisharha – seulotun aineiston ongelma	4
3	Aineiston generointi	6
4	Menetelmät	8
4.1	Verkkoteoria	8
4.2	Kausaalipäättely	10
4.2.1	Merkinnät ja keskeiset lauseet	11
4.2.2	Malli	12
4.2.3	algo	14
5	Tulokset	15
6	Johtopäätökset	16

Luku 1

Kiitokset – Acknowledgements

Tämän tutkielman aikana on tullut esiin takuujärjestelmään liittyvät ongelmat ja sovel-
lusalueen yhteiskunnallinen merkitys. Tutkielman teko on ollut minulle erityisen miele-
kässtä antoisan aiheen ja mieleisten yhteistyökumppanien vuoksi. Olen kirjoittanut tämän
kandidaatintutkielman yhteistyössä Helsingin yliopiston tietojenkäsittelytieteen osaston
apulaisprofessorin Michael Mathioudakis ja tohtoritutkijan Antti Hyttisen kanssa. He
tarjosivat minulle aiheen ja merkittävää tukea sekä tärkeitä kommentteja tämän tutkiel-
man kirjoittamisen aikana.

Tämän tutkielman on tarkastanut XYZ.

Helsingissä 14. toukokuuta 2019,
Riku Laine

I would like to wholeheartedly thank assistant professor Michael Mathioudakis from Uni-
versity of Helsinki's Department of Computer Science for numerous things. He provided me
this extremely interesting thesis topic and provided insightful and encouraging comments
throughout the process. Antti Hyttinen from the same department also gave important
insight in the causal modelling and commented on the content.

Luku 2

Johdanto

Tämän tutkielman tavoitteena on luoda kausaalipäätelyn avulla algoritmi, jolla voimme arvioida ennustavien mallien tarkkuutta, kun käytettävissä on ainoastaan valikoitumisharhasta kärsivää aineistoa. Samankaltaista asetelmaa ovat julkaisuissaan käsitelleet muun muassa Lakkaraju ja Madras [?, ?]. Pyrin tutkielmassani luomaan joustavamman ja tarkemman vaihtoehdon Lakkarajun luomalle supistusalgoritmille, mutta esitän ensin yleistä taustaa kausaalipäätelystä ja valikoitumisharhasta.

2.1 ”Kausaalipäätely uutena paradigmana”

Kuten Pearl ja Mackenzie esittävät kirjassaan Miksi, ihmisillä on luontainen kausaalisen päätelyn taito [?]. Tavalliset tilastollisen päätelyn menetelmät eivät tarjoa tapaa määritellä kausaalista yhteyttä: aineistosta voidaan päätellä erilaisia *korrelaatioita*, mutta kausaalinen päätely *A johtuu B:stä* vaatii uudenlaista lähestymistapaa. Käytännön tutkimuksessa kausaaliset yhteydet kiinnostavat erityisesti lääketieteen alalla [?]. Kuten Kalisch toteaa, aiemmin kausaalisuuden päätely on perustunut korrelaatioiden havaitsemiseen. On hypotetisoitu, että biomarkkerin ja taudin samanaikainen ilmaantuminen viittaisi siihen, että markkeri aiheuttaa taudin. Voimmeko siis markkeria käsittelemällä vaikuttaa tautiin tai jopa parantaa se? [?]

Syy-seuraussuhteen matemaattinen määrittely vaatii uutta lähestymistä myös todennäköisyyslaskennan merkintöihin. Pearl käyttää alkuperäisessä, englanninkielisessä kirjallisuudessa merkintää 'do' ilmaisemaan interventiota. Merkinnällä halutaan erottaa tavanomainen ehdollinen todennäköisyys $\mathbb{P}(Y|X = x)$ interventiosta, jossa asetamme muutujan X arvoon x : $\mathbb{P}(Y|\text{do}(X = x))$. Kimmo Pietiläinen käyttää kirjan suomennoksessa do-operaattorista käännöstä *tee*, mutta seuraan tässä tutkielmassa Pearl'n merkintöjä, ellen erikseen muuta mainitse [?]. Alalla käytetään myös muita, alaindekseillä rikastettuja

merkintätapoja [?]. Esittelen käyttämäni merkinnät tarkemmin kappaleessa 4.2.1.

Kausaalipäätelyssä mallit voidaan esittää graafeina, eli verkkoina. Verkoista voidaan suoraan lukea eri muuttujien relaatiot kausaalisuuden suuntien ja riippuvuuksien suhteen.

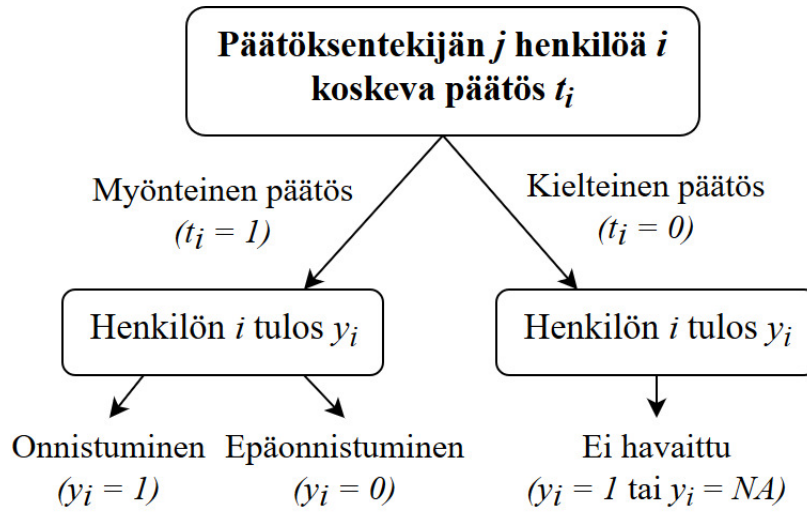
2.2 Valikoitumisharha – seulotun aineiston ongelma

Havaintoja voi puuttua erilaisissa tutkimuksissa useista eri syistä. Kyselytutkimuksissa vastauskatoa voi syntyä esimerkiksi vastaajan haluttomuudesta vastata kysymykseen tai yksinkertaisesti siitä syystä, että vastaajaa ei tavoiteta. Jos aineiston puuttuneisuusmekanismi on luonteeltaan täysin satunnainen, eli vastauksen puuttuneisuus ei liity millään tavalla mitattuihin muuttujiin, voidaan sanoa aineistoa puuttuvan *täysin satunnaisesti*. Käänteisessä tapauksessa voidaan puhua *ei-satunnaisesta puuttuvuudesta*. [?]

Tässä tutkielmassa tarkasteltavassa asetelmassa havaintojen puuttuminen liittyy sekä havaittuihin että havaitsemattomiin muuttujiin. Puuttuneisuuden voidaan sanoa olevan *satunnaista ehdollisesti*, koska aineistoa puuttuu vain yksilöiltä, joilla on korkea todennäköisyys haitalliseen tulokseen. (Erilaisia aineiston puuttuneisuusmekanismeja esitellään laajemmin esimerkiksi Laaksosen kirjassa *Surveyymetodiikka*.) Puuttuneisuutta voidaan korjata imputoinnilla, jolla yritetään tehdä mahdollisimman hyvä arvaus puuttuvasta arvosta. Todistan tutkielmassani myöhemmin, että kausaalipäätelyä hyödyntämällä voimme estimoida havaitusta, valikoitumisharhaisesta aineistosta haluttuja tunnuslukuja ilman imputointia harhattomasti. [?] Englanninkielisessä kirjallisuudessa seulotun aineiston ongelmasta on alettu käyttää Lakkarajun esittämää termiä *selective labels* [?].

Aineiston luova mekanismi on esitetty kuvassa 2.1 ja toimii siten, että aluksi jokin henkilö tai muu entiteetti saapuu päätöksentekijän eteen seulottavaksi. Päätöksentekijän tavoitteena on estää haitallinen tulos ($y = 0$) pitäen samalla myönteisten päätösten ($t = 1$) määrä mahdollisimman pienenä. Seuloja pyrkii siis antamaan kielteisen päätöksen kaikille niille, joilla epäonnistuminen on todennäköisin. Päätöksen jälkeen Kohtalo määrittää havainnolle tuloksen $y \in \{0, 1\}$. Kielteisen päätöksen saaneille tulos voidaan merkitä puuttuvaksi tai onnistuneeksi, koska haitallista tapahtumaa ei havaita.

Aineiston generoivaa mekanismia voidaan havainnollistaa lääke- ja oikeustieteen alan esimerkeillä. Henkilö on ensin mainitussa potilas ja jälkimmäisessä epäilty. Seuloja voi olla esimerkiksi lääkäri, joka päättää annetaanko potilaalle vahvempaa ja samalla kalliimpaa lääkettä, jolloin relapsia ei havaita. Oikeudellisessa asetelmassa seulojalla voidaan tarkoittaa tuomaria, joka päättää epäillyn vapauttamisesta takuita vastaan ilman pelkoa rikoksen uusimisesta. Molemmilla päättäjillä on selkeä kannustin estää haitalliset tulokset – sairauskohtaukset tai rikokset – pitäen samalla päätöksistä aiheutuvat rasitteet yhteiskunnalle ja yksilöiden elämille mahdollisimman pienenä. Lisäksi erityisesti oikeudellisessa asetelmassa on selvää, kuinka takuukäsittelystä kielteisen tuloksen saaneet eivät



Kuva 2.1: Valikoitumisharha aineiston generoivana mekanismina [?]

voi syyllistyä uuteen rikokseen, joten heidän tulosmuuttujan arvo voidaan koodata joko onnistumiseksi tai havaitsemattomaksi.

Luku 3

Aineiston generointi

Synteettinen aineisto luotiin Lakkarajun selostamalla tavalla. Aineistoon simuloitiin kolme muuttujaa X , Z , ja W . Näistä muuttujista X vastaa informaatiota, joka on sekä mallin että päätöksentekijän havaittavissa. Käytännössä muuttuja X vastaa kirjallista informaatiota, joka on kirjattu erilaisiin pöytäkirjoihin tai rekistereihin. Muuttujalla Z kuvataan tietoa, jonka vain päätöksentekijä voi havaita: kuten Lakkaraju havainnollistaa, tällaista voi olla oikeudessa esimerkiksi tieto siitä, onko vastaajalla perhettä mukana oikeussalissa. W tuo malliin kohinaa. Muuttujalla esitämme aineistossa informaatiota, joka ei ole sacaatavilla päätöksentekijöille eikä mallille, mutta vaikuttaa silti epätoivottavan tuloksen riskiin. Aineistossa nämä ovat kaikki riippumattomia standardinormaalijakautuneita satunnaismuuttujia, eli $X, W, Z \sim N(0, 1) \perp$. [?]

Aineistossa jyvitämme jokaiselle $M = 100$ päätöksentekijälle 500 arvioitavaa. Kaikille päättäjille arvotaan hyväksymisprosentti ottamalla arvoja tasajakaumasta suljetulta väliltä $[0,1; 0,9]$ ja sitten pyöristämällä saadut arvot 10 desimaalin tarkkuuteen. Tulosuuttuja Y määritetään ehdollisen todennäköisyyden

$$(3.1) \quad \mathbb{P}(Y = 0|X, Z, W) = \frac{1}{1 + \exp\{-(\beta_X X + \beta_Z Z + \beta_W W)\}}$$

mukaisesti. Jos $\mathbb{P}(Y = 0|X, Z, W) \geq 0,5$, tulosuuttujan arvoksi asetetaan 0 ja vastaavasti jos $\mathbb{P}(Y = 0|X, Z, W) < 0,5$ muuttujan arvoksi asetetaan 1. Lausekkeissa 3.1 ja 3.2 olevat kertoimet β_X , β_Z ja β_W ovat 1, 1 ja 0,2 vastaavassa järjestyksessä. [?]

Päätösmuuttuja T määritetään kaksivaiheisesti: ensin määritetään todennäköisyys kielteiselle päätökselle ja sitten muuttujan arvo asetetaan näiden todennäköisyyksien keskinäisen suuruuden mukaisesti. Muuttujan T ehdollinen todennäköisyys

$$(3.2) \quad \mathbb{P}(T = 0|X, Z) = \frac{1}{1 + \exp\{-(\beta_X X + \beta_Z Z)\}} + \epsilon,$$

missä $\epsilon \sim N(0, 0, 1)$ vastaa pientä määrää kohinaa. Henkilölle i annetaan kielteinen päätös, eli $T_i = 0$, jos ehdollinen todennäköisyys $\mathbb{P}(T = 0)$ on seulojan j suurimman $(1 - r) \cdot 100\%$ joukossa. Toisin sanoen seuloja j antaa myönteisen päätöksen r prosentille hänen arvioitavakseen annetuista henkilöistä, joilla on alin todennäköisyys kielteiseen päätökseen, oli se sitten rikoksen uusinta tai relapsi. [?]

Kun aineisto oli simuloitu, se jaettiin niin sanottuihin koulutus- ja testiaineistoihin. Lopuksi molempia aineistoja muokattiin siten, että tulostuuttujan arvo oli saatavissa vain yksilöille, joille oli annettu positiivinen päätös ($T = 1$). Kielteisen päätöksen saaneille tulostuuttujan arvo asetettiin arvoon NA, kuten kuvassa 2.1. Syntetisoidun aineiston keskeisimmät hajontaluvut on esitetty taulukossa 3.1. [?]

Muuttuja	Keskiarvo	Keskihajonta	Minimi	25%	50%	75%	Maksimi
acceptanceRate_R	0.48	0.23	0.10	0.26	0.47	0.65	0.89
X	0.00	1.00	-4.66	-0.67	0.00	0.67	3.83
Z	0.01	1.00	-4.85	-0.67	0.00	0.68	4.24
W	0.01	1.00	-4.03	-0.67	0.01	0.68	4.29
result_Y	0.50	0.50	0.00	0.00	0.00	1.00	1.00
decision_T	0.48	0.50	0.00	0.00	0.00	1.00	1.00

Taulukko 3.1: Synteettisen aineiston muuttujien hajontalukuja

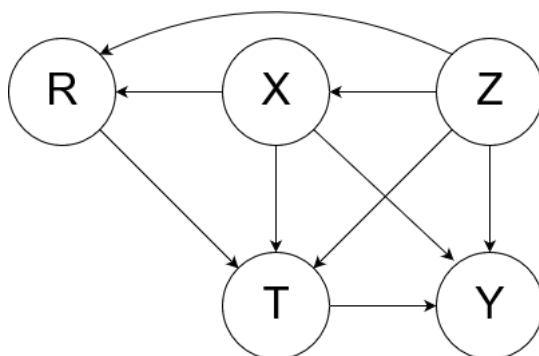
Luku 4

Menetelmät

Tässä kappaleessa selostan mallin laatimisessa ja arvioinnissa käyttämäni teoreettisen taustan. Koska kausaalinen malli esitetään verkkona, käyn aluksi läpi vaadittavat verkko-teoreettiset määritelmät. Esitän sen jälkeen mallini graafina ja osoitan kausaalisen vaikutuksen olevan identifioituva.

4.1 Verkkoteoria

Verkot koostuvat *solmuista* ja *kaarista*, joita voidaan havainnollistaa pisteinä ja viivoina tai nuolina näiden pisteiden väliillä. Kaaret ovat järjestettyjä pareja, kuten verkot itsekin, mutta oletan tavallisimmat joukko-opin merkinnät ja käsitteet tunnetuiksi. Noudatan määritelmissä Oinosta [?] ja erikseen merkityissä kohdissa Kivistä [?].



Kuva 4.1: Eräs verkko $H = (V, E)$, missä $V = \{R, X, Z, T, Y\}$.

Määritelmä 4.1 (Suunnattu verkko). *Suunnattu verkko* G on pari (V, E) , missä $V \neq \emptyset$ on *solmujen* joukko ja

$$E = \{(a, b) \in V \times V \mid \text{solmusta } a \text{ on nuoli solmuun } b\}$$

on *kaarien* joukko.

Kuvassa 4.1 näkyvässä verkossa esimerkiksi $(X, R) \in E$, mutta $(T, Z) \notin E$, koska solmusta T ei ole nuolta solmuun Z . Lisäksi voidaan todeta, että kaarien joukkoon kuuluu yhdeksän järjestettyä paria ja solmujen joukko V käsittää viisi alkioita.

Määritelmä 4.2. Oletetaan, että $G = (V, E)$ on suunnattu verkko ja $a, b \in V$.

Merkintä $a \rightarrow b$ tarkoittaa, että $(a, b) \in E$. Tällöin sanotaan, että a on kaaren (a, b) *lähtösolmu* ja b on kaaren (a, b) *maalisolmu*. Sanotaan myös, että solmu b on solmun a *vierussolmu* tai että solmut a ja b ovat *vierekkäisiä*.

Jos $(a, a) \in E$, sanotaan suunnatussa verkossa olevan *silmukka* solmussa a .

Esimerkkiverkossa H kaaren (Z, T) lähtösolmu on solmu Z ja maalisolmu solmu T . Lisäksi huomataan, että verkossa H ei ole yhtään silmukkaa. Kuvan 4.1 verkosta havaitaan, että melkein kaikki solmut ovat toistensa vierussolmuja. Ainoa poikkeus on solmut R ja Y , joiden välillä ei ole nuolta ja jotka eivät siten ole vierekkäisiä.

Määritelmä 4.3 (Yksinkertainen suunnattu verkko). Oletetaan, että $G = (V, E)$ on suunnattu verkko, jossa ei ole yhtään silmukkaa eli $(v, v) \notin E$ kaikilla $v \in V$.

Tällöin sanotaan, että G on yksinkertainen suunnattu verkko.

Esimerkkinä käytetystä verkosta H nähdään heti, että se on yksinkertainen suunnattu verkko, koska siinä ei ole yhtään silmukkaa. Yksinkertaisesta suunnatusta verkosta käytetään englanniksi nimitystä *directed acyclic graph* ja se saatetaan lyhentää DAG.

Määritelmä 4.4 (Polku ja suunnattu polku). Oletetaan, että G on yksinkertainen verkko ja $n \in \mathbb{N}, n \geq 1$.

- (a) Verkon G solmujen jono v_1, \dots, v_n on *polku* solmusta v_1 solmuun v_n , jos jonon jokaisesta solmusta on kaari jonon seuraavaan solmuun. Polkua voidaan merkitä

$$v_1 \rightsquigarrow v_n.$$

- (b) Jos verkko G on suunnattu verkko, $a, b \in V$ ja kaikki polun $a \rightsquigarrow b$ kaaret kulkevat kaarien suuntien mukaisesti, voidaan täsmentää, että polku $a \rightsquigarrow b$ on *suunnattu polku*.
- (c) Polku on *yksinkertainen*, jos kukin solmu esiintyy polussa vain kerran, paitsi että viimeinen ja ensimmäinen saavat olla sama solmu. [?]
- (d) Yksinkertainen polku on *sykli* eli *kehä*, jos viimeinen ja ensimmäinen solmu ovat samat. [?]

Huomataan, että verkossa H on useita polkuja solmusta R solmuun Y . Polku $R \rightarrow T \rightarrow Y$ on ainut suunnattu polku ja $R \leftarrow X \rightarrow Y$ on tavallinen polku, sillä solmujen R ja X välillä kuljetaan nuolen suunnan vastaisesti. Verkossa ei ole yhtään sykliä eli se on *sykliton*. Suunnatuista ja syklittömistä verkoista voidaan käyttää englannin kielestä juontuvaa lyhennettä DAG (*directed acyclic graph*) [?].

Määritelmä 4.5 (Jälkeläisyys). Oletetaan, että $G = (V, E)$ on suunnattu verkko ja $a, b \in V$.

Jos on olemassa suunnattu polku $a \rightsquigarrow b$, niin solmun b sanotaan olevan solmun a *jälkeläinen*. Vastaavasti tällöin sanotaan solmun a olevan solmun b *vanhempi*.

Esimerkiksi kuvan 4.1 verkossa solmulla Y ei ole jälkeläisiä ja solmun Z jälkeläiset ovat kaikki muut verkon solmut poislukien se itse, eli solmun Z jälkeläiset on joukko $V \setminus \{Z\}$.

Kausaalipäätelyssä kausaalisten vaikutusten identifioimiseksi tarvitaan usein selvittää niin sanotut *haarukka-* ja *käänteiset haarukkasolmut*. Määritellään ne seuraavaksi.

Määritelmä 4.6 (Haarukkasolmu). Oletetaan, että suunnatussa verkossa on polku $A \leftarrow B \rightarrow C \leftarrow D$. Tällöin solmua B sanotaan *haarukkasolmuksi* ja solmua C *käänteiseksi haarukkasolmuksi*.

4.2 Kausaalipäätely

Judea Pearl esittää artikkelissaan [?], että kaikessa tutkimuksessa, joka hyödyntää kausaalipäätelyä, tulisi edetä järjestelmällisesti neljässä vaiheessa:

1. Määrittely: Määritetään tavoitesuuruus Q funktiona $Q(\mathcal{M})$, joka voidaan laskea kaikille malleille \mathcal{M} .

2. Oletuksien esitys: Esitä kausaaliset oletukset luonnollisella kielellä ja ilmaise niiden rakenteellinen osa verkkona.
3. Identifioituvuus: Osoita, onko tavoitesuuruus määritettävissä (ilmaistavissa estimoitavina parametreina).
4. Estimointi: Estimoi tavoitesuuruutta, jos se on identifioituva tai approksimoi sitä jos se ei ole. Tarkista mallin mahdolliset (tilastolliset) oletukset ja implikaatiot ja muuta mallia, jos oletukset osoittautuvat paikkaansa pitämättömiksi.

Tutkielmani tavoitteena on esittää algoritmi, jolla voimme paremmin ennustaa riskiä populaatiotasolla, kun muutamme myönteisten päätösten osuutta jakun käytössä on valintaharhasta kärsivää aineistoa. Todennäköisyyslausekkein ilmaistuna haluamme siis selvittää vapautusprosentin muutoksen vaikutusta epätoivottavan tapahtuman $Y = 0$ todennäköisyyteen, mikä voidaan kirjoittaa muotoon

$$(4.7) \quad \mathbb{P}(Y = 0 | \text{do}(R = r)).$$

Huomataan, että lauseke 4.7 ei riipu mistään mallista \mathcal{M} , joten se täyttää Pearlin tavoitesuuruuden Q määritelmän mukaiset ehdot.

Kausaalipäättelyssä mallit määritellään usein yksinkertaisina suunnattuina verkkoina. Mallin määrittämistä verkosta voidaan suoraan lukea kausaaliset riippuvuussuhteet ja malliin kuuluvat muuttujat. Jos mallissa on solmut A ja B ja jos solmu B on solmun A jälkeläinen, niin muuttujalla A on mallin mukaan jonkinlainen kausaalinen vaikutus muuttujaan B . Jos verkossa muuttujien välillä ei ole jälkeläisyssuhdetta, niin ne ovat toisistaan riipumattomat. Kausaalisen vaikutuksen funktionaalista muotoa ei usein määritellä.

4.2.1 Merkinnät ja keskeiset lauseet

Kausaalipäättelyssä käytettävät merkinnät noudattelevat pitkälle tavallisia todennäköisyyslaskennan merkintöjä. Kun selvitetään muuttujan X vaikutusta muuttujaan Y ja tehdään interventio asettamalla muuttuja X arvoon x_0 , sitä merkitään $\mathbb{P}(Y | \text{do}(X = x_0))$.

Käydään seuraavaksi läpi kausaalilaskennan kannalta keskeisimmät lauseet. Lauseiden todistukset sivuutetaan, mutta ne on löydettävissä Pearlin artikkelin lähteistä [?]. Määritelmät 4.8 ja 4.9 **JNE**.

Määritelmä 4.8 (d-separoituvuus [?]). Joukko \mathcal{S} katkaisee (blocks) polun p , jos vähintään toinen seuraavista ehdoista on voimassa:

- (a) Polku p sisältää vähintään yhden solmun, joka on jonkin polun kulkusuuntaisen kaaren lähtösolmu ja kuuluu joukkoon \mathcal{S} . (arrow-emitting)
- (b) Polku p sisältää vähintään yhden käänteisen haarukkasolmun (collision node), joka ei kuulu joukkoon \mathcal{S} ja jolla ei ole jälkeläisiä joukossa \mathcal{S} .

Jos joukko \mathcal{S} katkaisee kaikki polut muuttujasta X muuttujaan Y , sanotaan joukon \mathcal{S} d-separoivan muuttujat X ja Y . Tällöin X ja Y ovat riippumattomia ehdolla \mathcal{S} , eli $X \perp\!\!\!\perp Y | \mathcal{S}$.

Määritelmä 4.9 (Takaovikriteeri (*back-door criterion*) [?]). Oletetaan, että halutaan selvittää muuttujan X kausaalista vaikutusta muuttujaan Y . Joukko \mathcal{S} on *riittävä* vaikutuksen selvittämiseen (sufficient for adjustment), kun seuraavat ehdot ovat voimassa:

- (1) Yksikään joukon \mathcal{S} alkioista ei ole solmun X jälkeläinen.
- (2) Joukon \mathcal{S} alkiot katkaisevat kaikki määritelmän 4.8 mukaiset kiertoreitit solmusta X solmuun Y . Kiertoreittejä ovat polut, jotka päättyvät muuttujaan X osoittavaan nuoleen.

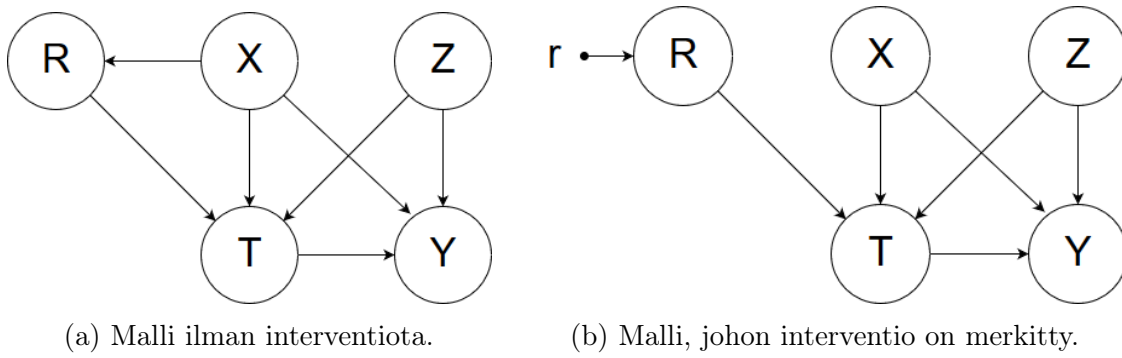
4.2.2 Malli

Malli sisältää viisi muuttujaa, jotka on esitelty lyhyesti taulukossa 4.1. Muuttujalla R kuvataan päätöksentekijän hyväksymisprosenttia, eli sitä prosentuaalista osuutta henkilöistä, joilla on pienin vaara epätoivottavaan tulokseen ja joille siten voidaan antaa myönteinen päätös. X ilmentää henkilön henkilökohtaisia ominaisuuksia, jotka ovat sekä päätöksentekijän että mallin havaittavissa. Muuttuja X voi olla esimerkiksi jonkinlainen rekisteritieto, kuten ikä tai sukupuoli. Muuttuja Z on muuttuja, jonka tuomari tai muu asiantuntija voi havaita, mutta joka on mallilta piilotettu. Muuttujan Z voidaan ajatella esimerkiksi oikeuskäsittelyjen tapauksessa kuvaavan epäillyn kääntöä oikeussalissa. Tulosuuttuja Y ja päätösmuuttuja T ovat kaksiarvoisia ja niiden määrittelyt on esitelty kuvassa 2.1: myönteistä päätöstä merkitään $t = 1$, kielteistä $t = 0$. Vastaavasti myönteinen tulos määritellään muuttujan y arvoksi 1, kielteinen arvoksi 0.

Mallin määrittelevä graafi on esitetty kuviossa 4.2a ilman virhemuuttujia. Graafista voidaan suoraan lukea oletukset: oletetaan, että $Z \perp\!\!\!\perp X, R$ mutta laajennetaan Lakkarajun oletuksia sallimalla muuttujan X vaikutus muuttujaan R [?]. Mallin oletetuilla kausaalisilla vaikutuksilla on lisäksi selkeästi ilmaistavat realisaatiot: kuinka osuuden R muuttaminen vaikuttaa päätökseen ja edelleen päätös tulokseen ja niin edelleen.

Muuttuja	Kuvaus
R	Myönteisten päätösten osuus prosentteina $r \in [0, 1]$
X	Kirjatut muuttujat, havaittavissa kaikille
Z	Kirjaamattomat muuttujat, vain päättäjän havaitsemat
Y	Tulosmuuttuja, $y \in \{0, 1\}$
T	Päätösmuuttuja, $t \in \{0, 1\}$

Taulukko 4.1: Mallin muuttujien selitteet



Kuva 4.2: Kausaalimallit graafeina.

Johdetaan muuttujan R kausaalivaikutus muuttujaan Y yli kaikkien ositteiden X . Huomataan, että osuuden R kausaalinen vaikutus voidaan ilmaista suoraan lausekkeella 4.7, sillä $\mathbb{P}(Y = 0 | \text{do}(R = 0)) = 0$ ja siten edelleen

$$\mathbb{P}(Y = 0 | \text{do}(R = r)) - \mathbb{P}(Y = 0 | \text{do}(R = 0)) = \mathbb{P}(Y = 0 | \text{do}(R = r)).$$

Osoitetaan seuraavaksi, että X on riittävä vaikutusten korjaamiseen määritelmän 4.9 mukaisesti, kun selvitetään muuttujan R kausaalista vaikutusta muuttujaan Y . Mallista voidaan suoraan lukea, että takaovikriteerin ensimmäinen ehto on voimassa: X ei ole muuttujan R jälkeläinen. Polut, jotka muuttujan X pitää katkaista ollakseen riittävä vaikutusten korjaamiseen ovat $R \leftarrow X \rightarrow Y$, $R \leftarrow X \rightarrow T \rightarrow Y$ ja $R \leftarrow X \rightarrow T \leftarrow Z \rightarrow Y$. Muuttuja X täyttää kuitenkin määritelmän 4.8 (a)-kohdan ehdon ja siten d-separoi muuttujat R ja Y . Tällöin X on riittävä vaikutusten korjaamiseen ja voidaan hyödyntää Pearlin kaavaa 25 [?]:

$$(4.10a) \quad \mathbb{P}(Y = 0 | \text{do}(R = r)) = \sum_x \mathbb{P}(Y = 0 | R = r, X = x) \mathbb{P}(X = x)$$

$$(4.10b) \quad = \sum_x \left(\sum_t \mathbb{P}(Y = 0, T = t | R = r, X = x) \right) \mathbb{P}(X = x)$$

$$(4.10c) \quad = \sum_x \left(\sum_t \mathbb{P}(Y = 0 | T = t, R = r, X = x) \mathbb{P}(T = t | R = r, X = x) \right) \mathbb{P}(X = x)$$

$$(4.10d) \quad = \sum_x \mathbb{P}(Y = 0 | T = 1, R = r, X = x) \mathbb{P}(T = 1 | R = r, X = x) \mathbb{P}(X = x)$$

$$(4.10e) \quad = \sum_x \mathbb{P}(Y = 0 | T = 1, X = x) \mathbb{P}(T = 1 | R = r, X = x) \mathbb{P}(X = x)$$

Yllä oleva lauseke on yhtäpitävä myös jatkuville muuttujan x arvoille, kun korvaamme summaukset integraalilla parametriavaruuden yli:

$$\mathbb{P}(Y = 0 | \text{do}(R = r)) = \int_x \mathbb{P}(Y = 0 | T = 1, X = x) \mathbb{P}(T = 1 | R = r, X = x) \mathbb{P}(X = x).$$

4.2.3 algo

Pearlin mukaan:

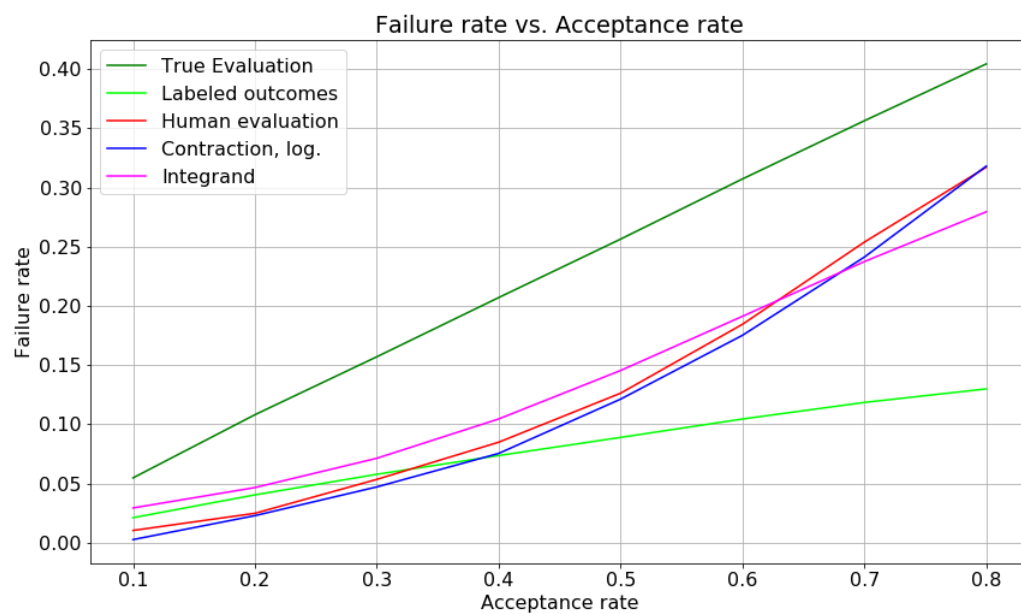
$$P(Y = 0 | \text{do}(R = r), X = x) = P(Y = 0 | R = r, X = x) = P(Y = 0 | R = r, X = x, T = 1) P(T = 1 | R = r, X = x)$$

Mallit vaikutukset laskettiin Pythonilla versio 3.6. Syötteett sklinear malliin , joka fitattiin testi dataan ja sitten integroitiin eri leniencyn tasoilla muuttujan X parametria-
varuuden eli reaaliakselin ylitse.

Luku 5

Tulokset

- se pääkuvaaja vertailuineen
 - beta ztan vaikutus?
 - erilaiset mallit ja koko käyrä aina 1 asti -> kuinka meillä parempi
 - voidaanko antaa estimaateille mitään luottusvälejä tjsp?



Kuva 5.1: Tulokset kuvana

Luku 6

Johtopäätökset

- Jatkosuunnitelmat: tutkitaan beta zetan vaikutusta tuloksiin, kuinka hyvin estimoituu. Sovelletaan oikeaan data settiin. Mielenkiintoiseksi on osoittautunut propublica julkaisun artikkelissa machine bias käyttämä COMPAS-aineisto.

- Ongelmat / muut huomiot: Tällä aikataululla en ole tehnyt mallin validointeja: onko kausaaliset pathwayt reasonable. Malli itsessään on suhteellisen yksinkertainen joten (KÄSIENHEILUTTELU) on jokseenkin luultavaa, että sinällään mallin spesifionnissa tuskin on mitään virheitä. Voitaisiin ehkä tietenkin koostaa jokseenkin hienosyisempi malli (erilaiset rikoshistoria yms erikseen) ja jotain. Jvat muuttujat? P-uloitteinen parametriavaruus???

- Mallin validointi epäeettistä, koska vaatisi huonoja päätöksiä > meillä kyllä synteettinen?

- Implikaatiot: parempia malleja???