УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Катарина Јелесијевић

БАЈЕСОВЕ МРЕЖЕ У НЕУРОНАУКАМА

мастер рад

Београд, 2024.

Ментор:

др Марија Цупарић, доцент
Универзитет у Београду, Математички факултет

Чланови комисије:

проф. др Бојана Милошевић, ванредни професор
Универзитет у Београду, Математички факултет

др Марко Обрадовић, доцент
Универзитет у Београду, Математички факултет

Датум одбране:

Захваљујем се професорки и својој менторки Марији Цупарић на помоћи око рада, као и осталим члановима комисије, професору Марку Обрадовићу и професорки Бојани Милошевић на уложеном времену приликом читања рада.

Посвећено породици, а посебно родитељима.

Садржај

1 Увод **1**

1.1 Мотивациони пример 3

2 Основни појмови **4**

2.1 Теорија вероватноће 4

2.2 Теорија графова 6

2.3 Марковљев модел 8

3 Бајесове мреже **10**

3.1 d-сепарација 10

3.2 Дефиниција Бајесових мрежа 13

3.2.1 Марковљево својство 17

3.3 Гаусове Бајесове мреже 28

3.3.1 Основна параметризација 29

3.3.2 Трансформација Гаусове Бајесове мреже у вишедимензиону нормалну расподелу 29

3.4 Динамичке Бајесове мреже 31

3.4.1 Скривени Марковљеви модели 34

4 Учење Бајесових мрежа из података **36**

4.1 Параметарско учење 36

4.1.1 Метод максималне веродостојности 37

4.1.2 Бајесова оцена 39

4.1.3 ЕМ алгоритам 42

4.2 Структурно учење 43

4.2.1 Методе учења засноване на ограничењима 44

4.2.2 Методе учења засноване на оцењивању 47

5 Примена Бајесових мрежа користећи *R* **55**

5.1 Графички приказ 56

5.2 Вероватносни приказ 58

5.3 Оцене мреже 61

6 Закључак **70**

Библиографија **71**

Глава 1

Увод

Бајесове мреже су компактан приказ расподеле вероватноће из скупа, углавном, дискретних промењљивих. Свака од промењљивих утиче на неку другу промењљиву или зависи од неке промењљиве, тако да су оне приказане као чворови мрежа. Структура Бајесових мрежа је усмерени, ациклични граф где чворови имају улогу промењљивих, а гране представљају њихову међуповезаност. Квантитативни део Бајесових мрежа је колекција табела условне вероватноће, од којих је свака придружена чвору, изражавајући вероватноћу промењљиве у чвору која је условљена својим родитељима у мрежи.

Нека су $(X_1, X_2, ..., X_n)$ случајне величине. Функција $P : B^n \rightarrow R$ назива се заједничка расподела вероватноће случајног вектора и може се одредити уз помоћ формуле множења условних вероватноћа. Ова расподела укључује довољно информација да се формира вероватноћа било ког догађаја израженог промењљивама мреже. Штавише, постоје ефикасни алгоритми за израчунавање такве вероватноће без потребе да се генерише заједничка расподела вероватноће (у многим случајевима би било неизводљиво израчунати n-димензиону расподелу). Бајесове мреже су енормно напредовале током последњих неколико деценија што је довело до примена које обухватају сва поља.

Неуронауке су широка биолошка област која се бави научним, систематским, експерименталним и теоријским истраживањима нервног система. Предмет истраживања неуронаука су структура, функције, генетика и еволуција, развиће, биохемија, физиологија, фармакологија и патологија нервног система. Употреба математике у проучавању мозга имала је велики утицај на области неуронауке и истовремено, мотивисала важна истраживања у математици.

Основни циљ неуронауке је да разуме како нервни систем комуницира и обрађује информације. Основна структурна јединица нервног система је појединачни неурон који преноси неуронске информације преко електричних и хемијских сигнала. Све активности мозга се заснивају на неуронским сигналима. Ове активности обухватају једноставне моторичке радње као нпр. ходање и дисање, као и виша когнитивна понашања као што су размишљање, осећање и учење.

Важан циљ математичке неуронауке је развој и анализа математичких модела за неуронске активности. Модели се користе да би се разумело како се активности генеришу и како се активности мењају када се параметри у систему мењају. Такође, модели служе за тумачење података, тестирање хипотеза и предлагање нових експеримената. Пошто су неуронски системи обично компликовани, морамо бити пажљиви да моделујемо систем на одговарајућем нивоу. Модел мора бити довољно компликован да укључује процесе који играју важну улогу у стварању одређене активности, међутим, модел не сме бити превише компликован тако да га је немогуће анализирати, било аналитички или рачунски.

Предности моделовања мозга су различите. (*i*) Захваљујући моделу, сазнања о последицама сложеног, нелинеарног можданог система са многим интерактивним компонентама су доступнија. (*ii*) Нове појаве се могу открити упоређивањем симулација предвиђања са експерименталним резултатима. Такође, нови експерименти могу бити конструисани на основу ових предвиђања. (*iii*) Експерименти који су компликовани или које је готово немогуће извести у живом ткиву, као што је нпр. селективна лезија одређених канала, синапси, неурона или путева, могу се симулирати коришћењем модела.[15]

У овом раду биће указано на проблеме неуронаука који су решени помоћу Бајесових мрежа. Такође, биће приказ опште теорије у вези са Бајесовим мрежама. Додатно, биће приказани алгоритми учења који се користе за конструисање структуре и процену вероватноћа који дефинишу Бајесове мреже и на крају, биће приказано како се Бајесове мреже користе у пракси кроз рад у R-у на конкретној бази података.

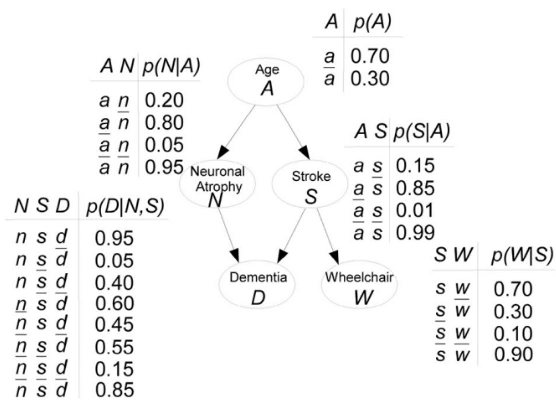
1.1 Мотивациони пример

Слика 1.1 приказује хипотетички пример Бајесових мрежа, моделовање ризика од деменције. Све променљиве су бинарне, x означава „присуство”, y „одсуство”, D деменцију, N неуронску атрофију, S мождани удар и W везан за инвалидска колица. За чвор *Age* A, a означава „старији од 65 година”, у супротном, $\overline{a}$. Променљива *Age* A утиче на променљиве *Neuronal Atrophy* N и *Stroke* S, и њихов је родитељ. Ова два стања утичу на променљиву *Dementia* D, њихово дете. Променљива *Wheelchair* W је директно повезана са можданим ударом. За сваки чвор, табела условне вероватноће указује на специфичне условне вероватноће, на пример, ако неко има неуронску атрофију и имао је мождани удар, вероватноћа да ће бити дементан је 0.95: $P(d|n,s) = 0,95$.

Заједничка расподела вероватноће:

$$P(A,N,S,D,W) = P(A)P(N|A)P(S|A)P(D|N,S)P(W|S).$$

Да би била у потпуности израчуната, заједничка расподела вероватноће захтева $2^5 - 1 = 31$ параметар, док је Бајесовим мрежама потребно само 11 иницијалних вероватноћа. Приметимо да у овако постављеној мрежи нема циклуса, тј. не постоји низ чворова тако да мрежа почиње и завршава се у истом чвору пратећи правац грана мреже, што је и један од предуслова за коришћење Бајесових мрежа. Потомци чвора S су D и W, док су сви чворови потомци од A. Редослед чворова може бити нпр. $A-N-S-D-W$ или $A-S-W-N-D$.



Слика 1.1: Хипотетички пример Бајесове мреже који моделира ризик од деменције.[2]

Глава 2

Основни појмови

У овом делу рада дефинисани су основни појмови који су потребни како би боље приближили појам *„Бајесове мреже"*.

2.1 Теорија вероватноће

Нека је Ω скуп елементарних догађаја и нека је на њему дата σ-алгебра $\mathcal{F}$. Скупове из σ-алгебре зовемо случајним догађајима, скуп Ω сигуран догађај, а $\emptyset$ немогућ догађај. Уређен пар $(\Omega, \mathcal{F})$ зовемо мерљив простор догађаја.

Дефиниција 1 *Нека је $(\Omega, \mathcal{F})$ мерљив простор догађаја. Скуповна функција $P : \mathcal{F} \to [0,1]$, где је $P \geq 0$, зове се* ***вероватноћа*** *ако има особине:*

1. $P(\Omega) = 1$,
2. $P(A) \geq 0$ *за сваки догађај* $A \in \mathcal{F}$,
3. *Ако су $A_1, A_2, ...$ дисјунктни догађаји из $\mathcal{F}$ тада важи* $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

Уређена тројка $(\Omega, \mathcal{F}, P)$ назива се ***простор вероватноће****.*

Понекад информације о догађају B не мењају наша уверења везана за догађај A, тј. догађај B не утиче на догађај A. У том случају кажемо да су A и B *независни* ($P(A|B) = P(A)$). Концепт независности се такође појављује када се условља више догађаја. Конкретно, ако информације о догађају B не промене наша уверења о догађају A, а да је при том, познат догађај C, онда кажемо да су A и B *условно независни* од C.

Дефиниција 2 (Условна независност) *Догађаји A и B су условно независни од догађаја C ако*

$$P(A|B,C) = P(A|C).$$

Условну независност такође можемо записати у облику

$$I_P(A,B|C)$$

где I_P означава независност у P.

Напомена: Када су два догађаја условно независна, онда можемо користити правило множења када хоћемо да израчунамо вероватноћу да ће се десити оба догађаја:

$$P(AB|C) = P(A|C) \cdot P(B|C).$$

Такође, битно је напоменути да условна независност игра јако битну улогу у Бајесовим мрежама.

Теорема 2.1.1 (Формула потпуне вероватноће) *Ако догађаји $E_1, E_2, ..., E_n$ чине потпун систем догађаја у односу на догађај F, тада је*

$$P(F) = \sum_{k=1}^{n} P(E_k)P(F|E_k). \tag{2.1}$$

Теорема 2.1.2 (Бајесова формула) *Дата су два догађаја E и F тако да је $P(E) \neq 0$ и $P(F) \neq 0$. Тада важи*

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}. \tag{2.2}$$

Ако је $\{E_1, E_2, ..., E_n\}$ потпун систем догађаја такав да $P(E_i) \neq 0$ за свако i, имамо да за $1 \leq i \leq n$ важи

$$\begin{aligned} P(E_i|F) &= \frac{P(F|E_i)P(E_i)}{P(F|E_1)P(E_1)+P(F|E_2)P(E_2)+\cdots+P(F|E_n)P(E_n)} \\ &= \frac{P(F|E_i)P(E_i)}{\sum\limits_{k=1}^{n} P(E_k)P(F|E_k)}. \end{aligned} \tag{2.3}$$

Доказ. Да бисмо доказали једнакост (2.2), прво користимо дефиницију условне вероватноће на следећи начин:

$$P(E|F) = \frac{P(EF)}{P(F)}, \qquad P(F|E) = \frac{P(FE)}{P(E)}.$$

Затим ћемо помножити ове једнакости са својим имениоцима, тако да добијемо:

$$P(E|F)P(F) = P(F|E)P(E)$$

Када поделимо ову последњу једначину са $P(F)$ добијамо управо једнакост (2.2). Уколико искористимо Формулу потпуне вероватноће 2.1.1 и израз $P(F)$ заменимо у имениоцу једнакости (2.2), добијамо тражену једнакост (2.3).

□

Дефиниција 3 (Нормална (Гаусова) расподела) *Случајна променљива* X *има нормалну (Гаусову) расподелу са математичким очекивањем* $\mu \in \mathbb{R}$ *и дисперзијом* σ^2 $(\sigma > 0)$*, у ознаци* $X \sim \mathcal{N}(\mu; \sigma^2)$ *ако је њена густина расподеле*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

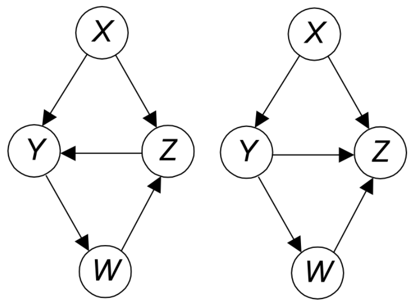
Стандарнда нормална (Гаусова) расподела је она код које је $\mu = 0$ *и* $\sigma = 1$*.*

Нормална (Гаусова) расподела има криву у облику звона, где μ представља очекивану вредност посматране случајне променљиве, на графику ће то бити центар. σ је стандарно одступање и оно одређује облик графика. Што је стандардно одступање мање, то је функција више испупчена, те су тада све тачке ближе очекиваној вредности и подаци су прецизни.

2.2 Теорија графова

У овом делу рада разматрамо предмет графичких модела који представља интеракцију између теорије вероватноће и теорије графова. Теорија графова помаже да се илуструју и користе независне структуре унутар интерактивних скупова променљивих, чиме се олакшава дизајн ефикасних алгоритама.

Дефиниција 4 ***Усмерени граф*** *је пар* (V, E)*, где је* V *коначан, непразан скуп чији се елементи називају* ***чворови****, а* E *је скуп уређених парова различитих елемената из* V*. Елементи скупа* E *се називају* ***усмерене гране****, и ако* $(X, Y) \in E$ *кажемо да постоји грана од* X *до* Y*.*



Слика 2.1: Оба графа су усмерени графови, али је само граф (b) усмерен ациклични граф.[9]

Скуп чворова за граф (a) на слици 2.1 је

$$V = \{X, Y, Z, W\},$$

а скуп грана је

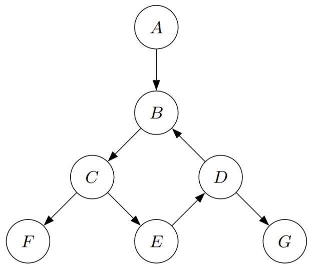
$$E = \{(X, Y), (X, Z), (Y, W), (Z, Y), (W, Z)\}.$$

Путања у усмереном графу је низ чворова $[X_1, X_2, ..., X_k]$ тако да $(X_{i-1}, X_i) \in E, 2 \leq i \leq k$. На пример, $[X, Y, W, Z]$ је путања у усмереном графу на слици 2.1 (a). **Ланац** у усмереном графу је низ чворова $[X_1, X_2, ..., X_k]$ тако да $(X_{i-1}, X_i) \in E$ или $(X_i, X_{i-1}) \in E, 2 \leq i \leq k$. На пример, $[Y, W, Z, X]$ је ланац у усмереном графу на слици 2.1 (b), али није пут. **Циклус** у усмереном графу је пут који почиње и завршава се истим чвором. На слици 2.1 (a), $[Y, W, Z, Y]$ је циклус од Y до Y, док у делу (b) $[Y, W, Z, Y]$ није циклус јер није путања.

Дефиниција 5 *Усмерени граф $\mathbb{G}$ назива се* ***директан (усмерен) ациклични граф*** *(ДАГ) ако не садржи усмерене циклусе.*

За дати ДАГ $\mathbb{G} = (V, E)$ и чворове $X, Y \in V$, Y је **родитељ** од X ако постоји грана од Y до X, односно Y је **потомак (дете)** од X и X је **предак** од Y ако постоји путања од X до Y, и за Y кажемо да је **непотомак** од X ако Y није потомак од X и Y није родитељ од X.

Бајесове мреже морају бити ацикличне. Структура ацикличног графа гарантује да не постоји чвор који може бити сопствени предак или сопствени потомак, тј. да не постоји циклус повратних информација с обзиром да их је тешко моделирати. Овај услов је значајан и за факторизацију заједничке расподеле вероватноће. Ради лакшег разумевања Бајесових структура, обично су мреже представљене у виду (наопаког) стабла, где гране иду од врха ка доле. Тада су корени на врху, а листови на дну тог стабла.



Слика 2.2: Усмерени граф са циклусом повратних информација. Ово није дозвољено у Бајесовим мрежама.[4]

2.3 Марковљев модел

Код Марковљевих модела важи да вероватноћа неког догађаја који ће се реализовати у будућности (у неком тренутку t) не зависи од прошлости (од тренутака $t_1, t_2, ..., t_{n-1}$), већ само од садашњости (од тренутка t_n). Другим речима, расподела случајне променљиве X_t у моменту $t = t_n$ зависи само од вредности x_{n-1} процеса у моменту t_{n-1}, а не зависи од осталих вредности $x_{n-2}, ..., x_1$ процеса у моментима $t_{n-2} > t_{n-3} > ... > t_1$. Ову особину Марковљевих процеса називамо *Марковљево својство.*

Дефиниција 6 *Нека је S пребројив скуп. Случајни процес $X = (X_n : n \geq 0)$ дефинисан на простору вероватноћа (Ω, P) са вредностима у скупу S је* ***Марковљев ланац првог реда*** *ако важи*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i) \qquad (2.4)$$

за свако $n \geq 0$ и за све $i_0, ..., i_{n-1}, i, j \in S$ за које су обе условне веробатноће добро дефинисане.

Једнакост (2.4) назива се **Марковљево својство**.[12]

Дефиниција 7 *Нека је $p_{ij} = P(X_{t+1} = j | X_t = i)$ вероватноћа да случајна променљива X пређе у стање j у тренутку $t+1$, ако је у тренутку t била у стању i. Вредности p_{ij} се називају* ***вероватноће прелаза*** *и морају да* *задовољавају следеће услове:*

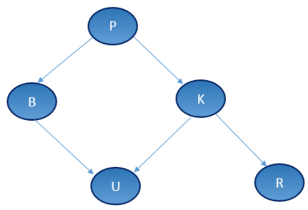
$$p_{ij} \geq 0, \qquad \sum_{i,j=1}^{n} p_{ij} = 1.$$

Марковљев ланац заједно са задатим вероватноћама прелаза називамо **Марковљевим моделом**.

Конкретно, у Бајесовим мрежама, Марковљево својство служи да поједностави вероватноћу пресека коначно много догађаја $A_1, A_2, ..., A_n$,

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1A_2...A_{n-1}),$$

која ће касније бити и теоријски образложена (Исказ 1 и Теорема 3.2.1). Марковљево својство представља једну од особина Бејесових мрежа која подразумева да сваки од чворова у мрежи зависи само од својих родитеља. То би значило да се приликом рачунања условне вероватноће за сваку променљиву у мрежи узима у обзир само вероватноћа њених родитеља. На слици испод дат је пример Бејесове мреже.



Слика 2.3: Бајесова мрежа [8]

Уколико се примени правило множења, вероватноћа пресека догађаја који чине мрежу била би следећа:

$$P(RUKBP) = P(R|UKBP)P(U|KBP)P(K|BP)P(B|P)P(P).$$

Узимајући у обзир да важи Марковљево својство, јасно је да ће, на пример, догађај K зависити само од P, јер је он његов једини родитељ. Понављајући исти поступак за сваки чвор у мрежи, вероватноћа пресека ових догађаја је следећа

$$P(RUKBP) = P(R|K)P(U|BK)P(K|P)P(B|P)P(P).$$

Глава 3

Бајесове мреже

Бајесове мреже се дефинишу као узрочно-последичне (каузалне) мреже које се састоје од узрочних веза представљених као условне вероватноће. У наставку биће представљена формула множења вероватноћа (енг. *chain rule*) за Бајесове мреже. Формула множења вероватноћа је својство које чини Бајесове мреже веома моћним алатом за представљање области са постојећим неизвесностима.

3.1 *d*-сепарација

Битно је напоменути да променљиве могу имати пребројив или непрекидан скуп стања, али у овом раду биће разматране само променљиве са коначаним бројем стања. Дакле, ако променљива има неколико стања (или исхода), она може, на пример, бити *болест* где су стања *бронхитис, туберкулоза, рак плућа*.

У каузалној мрежи, променљива представља скуп могућих стања, међутим она може бити искључиво у једном од својих стања (које некада може бити непознато).

Узрочне-последичне мреже се могу користити за праћење како промена у једној променљивој може утицати на друге променљиве, као и сам ток мреже. У наставку биће представљен скуп правила за такав начин резоновања.

Дефиниција 8 (*d*-сепарација) *Нека је дат граф* $(V, E), A \subseteq V$. *За два различита чвора* X *и* Y *важи d-сепарација (енг. d-separation или directed separation) ако важи бар једна од следећих ставки:*

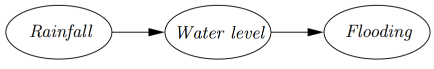
1. *На путањи између X и Y постоји чвор $Z \in A$ тако да је ланац $X \to Z \to Y$ усмерен по принципу глава - реп. Тада се за везу каже да је редна (серијска).*
   Напомена: Структура $Y \to Z \to X$ није редна веза. Уместо тога, представља обрнуту редну везу (обрнуту ланчану везу). У овом случају, чворови су повезани у линеарној секвенци, али у супротном правцу у односу на редну везу. Конкретно, ако је $Y \to Z \to X$, онда Y утиче на Z, а Z утиче на X.

2. *На путањи између X и Y постоји чвор $Z \in A$ тако да је ланац $X \leftarrow Z \to Y$ усмерен по принципу реп - реп. Тада се за везу каже да је дивергентна.*

3. *На путањи између X и Y постоји чвор Z, такав да ни Z нити било који од његових потомака није у A. Ланац $X \to Z \leftarrow Y$ је усмерен по принципу глава - глава и за овакву везу се каже да је конвергентна.*

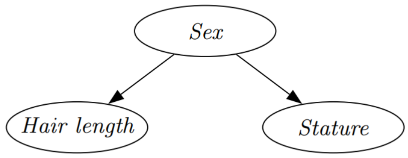
За X и Y се каже да су д-повезани уколико нису д-сепарабилни.

Пример 1 Слика 3.1 приказује модел који има серијску везу између чворова *Rainfall* (без кише, слаба, средња или јака киша), *Water level* (низак, средњи или висок ниво воде), *Flooding* (поплава - да или не). Уколико се не узме у обзир ниво воде, онда ће се, знајући да је било поплаве, повећати уверење да је ниво воде висок, при чему ће заузврат бити више информације о падавинама. Иста логика се примењује и у супротном смеру. Са друге стране, уколико је познат ниво воде, онда чињеница да је било поплаве не говори ништа о интезитету кише.



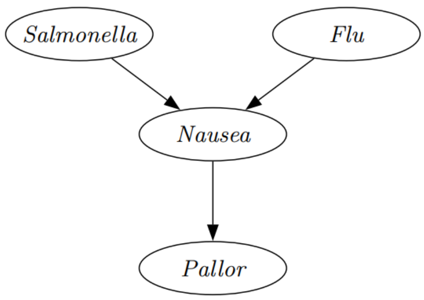
Слика 3.1: Модел са чворовима *Rainfall, Water level, Flooding* који су повезани серијском везом.[4]

Пример 2 Слика 3.2 приказује дивергентну везу између чворова *Sex* (мушки или женски пол), *Hair length* (дуга или кратка коса) и *Stature* (висина <168 или >=168 цм). Уколико се не зна пол особе, на основу дужине косе се може добити више информација о полу што ће заузврат оправдати уверења о висини особе. Са друге стране, уколико знамо да је у питању мушки пол, онда дужина његове косе не даје информације о његовој висини.



Слика 3.2: Пол утиче на дужину косе, као и на висину особе.[4]

Пример 3 Слика 3.3 приказује конвергентну везу између *Salmonella* инфекције, *Flu*, *Nausea* и *Pallor*. Ако се не зна ништа о мучнини или бледилу особе, онда информација о томе да ли особа има инфекцију салмонелом не говори ништа о томе да ли особа има грип. Међутим, ако је особа бледа, онда уз информацију да особа нема инфекцију салмонелом сматраће се да особа има грип.



Слика 3.3: Салмонела и грип могу изазвати мучнину, што заузврат изазива бледило.[4]

Три претходна случаја покривају све начине на које се евиденција о нечему може пренети кроз променљиву, и пратећи правила могуће је одлучити за било који пар променљивих у узрочној мрежи да ли су независне с обзиром на одговарајуће услове у мрежи.

3.2 Дефиниција Бајесових мрежа

За променљиву A са стањима $a_1, ..., a_n$, изражавамо нашу неизвесност о њеном стању кроз расподелу вероватноће $P(A)$ над тим стањима:

$$P(A) = P(x_1, ..., x_n); \qquad x_i \geq 0; \qquad \sum_{i=1}^{n} x_i = x_1 + \cdots x_n = 1,$$

где је x_i вероватноћа да A буде у стању a_i. Расподела се зове *униформна* ако су све вероватноће једнаке.

Напомена: Уопштено, вероватноћа да је A у стању a_i означава се са $P(A = a_i)$, а са $P(a_i)$ ако је променљива очигледна из контекста.

Нека је $U = \{A_1, ..., A_n\}$ скуп променљивих. Ако је могуће израчунати заједничку расподелу вероватноће $P(U) = P(A_1, ..., A_n)$, онда је могуће израчунати и $P(A_i)$, као и $P(A_i|e)$ где је e евиденција о некој од променљивих у Бајесовој мрежи.

Дефиниција 9 *Нека је $P(U)$ расподела вероватноће која задовољава својства Бајесових мрежа: (i) условне вероватноће за променљиву која је одређена променљивама које представљају њене родитеље морају бити унапред дефинисане у Бајесовим мрежама, и (ii) ако су променљиве A и B d-сепарабилни скупом променљивих C у Бајесовим мрежама, онда су A* ***и*** *B* ***независни од*** *C* ***у*** *$P(U)$.*

На основу ова два својства може се закључити да ако се скуп променљивих U састоји од само једне променљиве A_1, онда се Бајесовим мрежама одређује $P(A_1)$, а $P(U)$ је једнозначно одређен. Касније ћемо показати да то уопштено важи. За расподелу вероватноће $P(U) = P(A_1, ..., A_n)$ постоји једначина позната као *формула множења вероватноћа.* За Бајесове мреже ова једначина има посебан облик.

Исказ 1 (Формула множења вероватноћа) *Нека је $U = \{A_1, ..., A_n\}$ скуп догађаја. Тада за сваку расподелу вероватноће $P(U)$ важи*

$$P(U) = P(A_n|A_1, ..., A_{n-1})P(A_{n-1}|A_1, ..., A_{n-2}) \cdots P(A_2|A_1)P(A_1).$$

Доказ. Итеративно коришћење основног правила условне вероватноће:

$$\begin{aligned} P(U) &= P(A_n|A_1,...,A_{n-1})P(A_1,...,A_{n-1}), \\ P(A_1,...,A_{n-1}) &= P(A_{n-1}|A_1,...,A_{n-2})P(A_1,...,A_{n-2}), \\ &\vdots \\ P(A_1,A_2) &= P(A_2|A_1)P(A_1). \end{aligned}$$

□

Претпоставимо да за свако A_i постоји подскуп $rod(A_i) \subseteq \{A_1,...,A_{i-1}\}$ тако да је A_i условно независна од свих променљивих из скупа $\{A_1,...,A_{i-1}\} \setminus rod(A_i)$, тј.

$$P(A_i|A_1,...,A_{i-1}) = P(A_i|rod(A_i)).$$

Када ову једначину убацимо у формулу множења вероватноћа (Исказ 1), добијамо теорему која управо следи.

Теорема 3.2.1 (Правило ланца за Бајесове мреже) *Нека су Бајесове мреже дефинисане над скупом података $U = \{A_1,...,A_n\}$. Тада је јединствена заједничка расподела вероватноће $P(U)$ задата као производ свих условних вероватноћа одређених у Бајесовим мрежама дата формулом:*

$$P(U) = \prod_{i=1}^{n} P(A_i|rod(A_i)), \tag{3.1}$$

где су $rod(A_i)$ родитељи од A_i у Бајесовим мрежама.

Доказ. Претпоставимо да је $P(U)$ расподела вероватноће и да $P(U)$ задовољава својства Бајесових мрежа. Треба доказати да расподела вероватноће $P(U)$, дефинисана као у теореми 3.2.1, задовољава условне вероватноће из Бајесових мрежа, тј. да производ условних вероватноћа заиста правилно представља целокупну расподелу вероватноће унутар мреже. Такође, путем индукције, треба доказати да производ задовољава својства d-сепарације.

За базични случај, када је $n = 1$, имамо само једну променљиву A_1. Тада је

$$P(A_1) = P(A_1),$$

што очигледно важи јер је расподела вероватноће за једну променљиву једнака сама себи. У следећем, индукционом кораку, претпоставимо да теорема важи за $n-1$ променљивих. То значи да важи

$$P(U \setminus \{A_n\}) = \prod_{i=1}^{n-1} P(A_i|rod(A_i)).$$

Сада треба да покажемо да теорема важи и за n променљивих, тј. да важи

$$P(U) = \prod_{i=1}^{n} P(A_i|rod(A_i)).$$

Дакле, користећи формулу множења вероватноћа, заједничку вероватноћу за n променљивих можемо записати као

$$P(U) = P(A_1, A_2, ..., A_n) = P(A_n|A_1, A_2, ..., A_{n-1})P(A_1, A_2, ..., A_{n-1}).$$

Према претпоставци индукције, имамо да је

$$P(U) = P(A_1, A_2, ..., A_n) = P(A_n|A_1, A_2, ..., A_{n-1}) \prod_{i=1}^{n-1} P(A_i|rod(A_i)).$$

У Бајесовој мрежи, условна независност значи да, када знамо вредности родитеља променљиве A_i (тј. вредности променљивих од којих A_i директно зависи), променљива A_i је независна од свих других променљивих у мрежи које нису њени потомци. Другим речима, знајући вредности родитеља A_i, додатне информације о било којој променљивој која није потомак од A_i не пружају никакве додатне информације о A_i.Посебно, за променљиву A_n важи:

$$P(A_n|A_1, A_2, ..., A_{n-1}) = P(A_n|rod(A_n)).$$

Користећи ову условну вероватноћу, коначно добијамо

$$P(U) = P(A_n|rod(A_n)) \prod_{i=1}^{n-1} P(A_i|rod(A_i)) = \prod_{i=1}^{n} P(A_i|rod(A_i)).$$

Даље, нека је B_M Бајесова мрежа са n променљивих $\{A_1, ..., A_n\}$ и претпоставимо да A_n нема потомке и нека B'_M представља B_M без елемента A_n. Јасно је да је B'_M Бајесова мрежа са истим условним расподелама вероватноће као B_M (осим за A_n) и са истим својствима d-сепарације над $\{A_1, ..., A_{n-1}\}$ као B_M.

Дакле, ако су A и B d-сепарабилни датим C у B_M, онда су A и B d-сепарабилни и у B'_M, стога важи $P(A|B, C) = P(A|C)$. Да би доказали да ово важи и за својства d-сепарације када је укључен елемент A_n, размотрићемо случај када $A_n \in C$ и случај када је $A = A_n$. У првом случају, узимајући у обзир да A_n припада конвергентној вези, важи да ако су A и B d-сепарабилни датим C, онда су они d-сепарабилни и датим $C \setminus \{A_n\}$ и добијамо тврдњу изнад. У другом сличају, приметимо да важи

$$P(A_n|B,C) = \sum_{p(A_n)} P(A_n|B,C,rod(A_n))P(rod(A_n)|B,C).$$

Даље, ако су A_n и B d-сепарабилни датим C, онда су $rod(A_n)$ и B такође d-сепарабилни датим C. С обзиром да у Бајесовој мрежи B'_M није укључен елемент A_n, важи $P(rod(A_n)|B,C) = P(rod(A_n)|C)$, тако да је само остало да се покаже да важи $P(A_n|B,C,rod(A_n)) = P(A_n|rod(A_n))$. Користећи формулу множења вероватноће као и правило ланца, важи следеће

$$\begin{aligned}
P(A_n|B,C,rod(A_n)) &= \frac{P(A_n,B,C,rod(A_n))}{P(B,C,rod(A_n))} = \frac{\sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} P(U)}{\sum\limits_{U\setminus\{B,C,rod(A_n)\}} P(U)} \\
&= \frac{\sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} \prod\limits_{i=1}^{n} P(A_i|rod(A_i))}{\sum\limits_{U\setminus\{B,C,rod(A_n)\}} \prod\limits_{i=1}^{n} P(A_i|rod(A_i))} \\
&= \frac{P(A_n|rod(A_n)) \sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} \prod\limits_{i=1}^{n-1} P(A_i|rod(A_i))}{\sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} \prod\limits_{i=1}^{n-1} P(A_i|rod(A_i)) \sum\limits_{A_n} P(A_n|rod(A_n))} \\
&= \frac{P(A_n|rod(A_n)) \sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} \prod\limits_{i=1}^{n-1} P(A_i|rod(A_i))}{\sum\limits_{U\setminus\{A_n,B,C,rod(A_n)\}} \prod\limits_{i=1}^{n-1} P(A_i|rod(A_i)) \cdot \mathbf{1}} \\
&= P(A_n|rod(A_n)).
\end{aligned}$$

Да би доказали јединственост, нека је $\{A_1, ..., A_n\}$ тополошко уређење[1] промењљивих. Тада за сваку промењљиву A_i са родитељима $rod(A_i)$ имамо да је A_i d-сепарабилно од $\{A_1, ..., A_{i-1}\} \setminus rod(A_i)$ датим $rod(A_i)$. Ово значи да за било коју расподелу P која задовољава својства Бајесове мреже B_M мора важити $P(A_i|A_1, ..., A_{i-1}) = P(A_i|rod(A_i))$. Заменом ове једначине у опште правило ланца,

$$\begin{aligned}
P(U) &= P(A_n|A_1, ..., A_{n-1})P(A_{n-1}|A_1, ..., A_{n-2}) \cdots P(A_2|A_1)P(A_1) \\
&= P(A_n|rod(A_n)) \cdots P(A_1|rod(A_1)),
\end{aligned}$$

[1]Тополошко уређење (сортитање) усмереног графа је линеарно уређење његових чворова тако да свака усмерена грана од чвора U ка чвору V, U долази пре V у уређењу. На пример, родитељи долазе испред своје деце у низу.

доводи до тога да свака расподела која задовољава својства Бајесове мреже B_M мора бити производ условних вероватноћа.

□

Дакле, Бајесове мреже имају две компоненте: директан ациклични граф, где је сваки чвор случајна променљива A_i и скуп условних расподела вероватноћа датих са $P(A_i|rod(A_i))$ за сваку променљиву A_i, и оне одређују заједничку расподелу вероватноће дате једнакошћу (3.1) из Теореме 3.2.1. Прва квалитативна компонента се назива *структура Бајесове мреже*, а друга *параметри Бајесове мреже*.

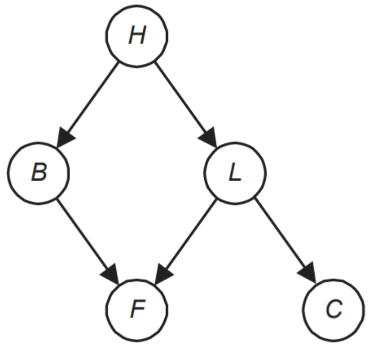
Када су сви чворови дискретне променљиве, условне вероватноће су табеларно приказане у ономе што се обично назива *табела условне вероватноће*.

3.2.1 Марковљево својство

Условне вероватноће у Бајесовим мрежама које су представљене као производ заједничке расподеле вероватноћа, као из теореме 3.2.1, су такве да сваки чвор је условно независан од својих непотомака у односу на своје родитеље. Тада се каже да $\mathbb{G}$ задовољава Марковљево својство са расподелом вероватноће P и да је $(\mathbb{G}, P)$ Бајесова мрежа. На слици 3.4, ако применимо Марковљево својство на чвор L, можемо приметити да су L и B условно независни у односу на чвор H.

Дефиниција 10 *Нека је P заједничка расподела вероватноћа случајних променљивих у коначном, непразном скупу V чији се елементи називају чворови и нека је $\mathbb{G} = (V, E)$ директан ациклични граф, где је E скуп уређених парова различитих елемената V. Кажемо да $(\mathbb{G}, P)$ задовољава **Марковљево својство** ако за сваку променљиву $X \in V$, $\{X\}$ је условно независан од скупа свих његових непотомака у односу на скуп свих његових родитеља. Када $(\mathbb{G}, P)$ задовољава Марковљево својство, онда $\mathbb{G}$ и P међусобно задовољавају Марковљево својство.*

У случају великог броја инстанци, компликовано је у потпуности израчунати заједничку расподелу вероватноће. Тада се користе погодности које пружа Марковљево својство. Теорема 3.2.2, која ће бити дефинисана након кратког увода, показује да ако $(\mathbb{G}, P)$ задовољава Марковљево својство, онда



Слика 3.4: Илустрација Марковљевог својства путем директног ацикличног графа.[10]

је P једнака производу условних расподела вероватноћа свих чворова у односу на њихове родитеље у $\mathbb{G}$, кад год ове условне расподеле постоје. Ово значи да је потребно одредити много мање вредности (параметара) него када би на стандардан начин одређивали све вредности у заједничкој расподели.

Дефиниција 11 *За сваки чвор X са родитељима $Y_1, Y_2, ..., Y_k$* ***број параметара*** *потребан за чвор X, тј. за описивање условне вероватноће $P(X|Y_1, Y_2, ..., Y_k)$ је $(|X| - 1) \cdot (|Y_1| \cdot |Y_2| \cdots |Y_k|)$, где је $|X| - 1$ због нормализације (јер збир вероватноћа за све могуће вредности мора бити 1), а $|X|$ представља број могућих вредности за X. Укупан број параметара у мрежи је збир броја параметара за све чворове.*

У наставку следи илустрација горе дефинисане заједничке расподеле вероватноће P у директном ацикличном графу $\mathbb{G}$. На слици 3.4, за све реализоване вредности f, c, b, l и h важи

$$P(f, c, b, l, h) = P(f|b, l)P(c|l)P(b|h)P(l|h)P(h), \tag{3.2}$$

кад год условне вероватноће са десне стране постоје. Ако се примети да једна од њих не постоји за одређену комбинацију променљивих, онда је $P(b, l) = 0$ или $P(l) = 0$ или $P(h) = 0$, што имплицира $P(f, c, b, l, h) = 0$ за ту комбинацију вредности. Међутим, постоје случајеви у којима је $P(f, c, b, l, h) = 0$ и условне вероватноће постоје. На пример, ово би био случај ако све условне вероватноће са десне стране постоје и $P(f|b, l) = 0$ за неке комбинације вредности f, b и l. Дакле, једнакост 3.2 важи за све вредности различите од нуле заједничке расподеле вероватноће.

Теорема 3.2.2 *Ако* $(\mathbb{G}, P)$ *задовољава Марковљево својство, онда је* P *једнака производу условних расподела вероватноће свих чворова у односу на вредности њихових родитеља, кад год постоје ове условне расподеле.*

Доказ. Нека је P дискретна расподела вероватноће. Треба поређати чворове тако да, ако је Y потомак од Z, тада у редоследу Y стоји иза Z.[2] Нека је $X_1, X_2, ..., X_n$ резултат поретка променљивих. За дати скуп вредности $x_1, x_2, ..., x_n$, нека је rod_i подскуп свих вредности које садрже родитељи од X_i. Треба показати да кад год је $P(rod_i) \neq 0$ за $1 \leq i \leq n$ важи

$$P(x_n, x_{n-1}, ..., x_1) = P(x_n|rod_n)P(x_{n-1}|rod_{n-1}) \cdots P(x_1|rod_1).$$

Ова једнакост се показује помоћу индукције у зависности од броја променљивих у мрежи. Претпоставимо да, за неку комбинацију вредности x_i, важи $P(rod_i) \neq 0$ за $1 \leq i \leq n$.

ИНДУКЦИОНА БАЗА: Пошто је rod_1 празан,

$$P(x_1) = P(x_1|rod_1).$$

ИНДУКЦИОНА ХИПОТЕЗА: Претпоставимо да за ову комбинацију вредности x_i-јева важи

$$P(x_i, x_{i-1}, ..., x_1) = P(x_i|rod_i)P(x_{i-1}|rod_{i-1}) \cdots P(x_1|rod_1).$$

ИНДУКЦИОНИ КОРАК: Треба да покажемо да за ову комбинацију вредности x_i-јева важи

$$P(x_{i+1}, x_i, ..., x_1) = P(x_{i+1}|rod_{i+1})P(x_i|rod_i) \cdots P(x_1|rod_1). \tag{3.3}$$

Постоје два случаја:

СЛУЧАЈ 1: За ову комбинацију вредности,

$$P(x_i, x_{i-1}, ..., x_1) = 0. \tag{3.4}$$

Јасно, једнакост (3.4) имплицира

$$P(x_{i+1}, x_i, x_{i-1}, ..., x_1) = 0.$$

[2] Примери таквог поретка за директан ациклични граф су $[H, L, B, C, F]$ и $[H, B, L, F, C]$ на слици 3.4.

Осим тога, због једнакости (3.4) и индукционе хипотезе, постоји неко k, где је $1 \leq k \leq i$, тако да је $P(x_k|rod_k) = 0$. Дакле, важи једнакост (3.3).
СЛУЧАЈ 2: За ову комбинацију вредности,

$$P(x_i, x_{i-1}, ..., x_1) \neq 0.$$

У овом случају,

$$\begin{aligned} P(x_{i+1}, x_i, x_{i-1}, ..., x_1) &= P(x_{i+1}|x_i, ..., x_1)P(x_i, ..., x_1) \\ &= P(x_{i+1}|rod_{i+1})P(x_i, ..., x_1) \\ &= P(x_{i+1}|rod_{i+1})P(x_i|rod_i) \cdots P(x_1|rod_1). \end{aligned}$$

У првој једнакости је искоришћено правило за условну вероватноћу, у другој Марковљево својство и чињеница да $X_1, ..., X_i$ нису потомци од X_{i+1}, и у последњој једнакости је употребљена индукциона хипотеза.

$\square$

У пракси, обично се не наводе простор узорка и функција вероватноће из које се могу израчунати условне расподеле. Уместо тога, потребно је идентификовати случајне променљиве и вредности у условним расподелама. На пример, у моделу који испитује дијагнозу рака плућа, идентификујемо променљиве као што су *SmokingHistory*, *LungCancer* и *ChestXray* и вероватноће као што су *P(SmokingHistory =yes), P(LungCancer = present|SmokingHistory = yes),* и *P(ChestXray =positive| LungCancer = present).* Поставља се питање како можемо знати да је производ ових условних расподела уопште заједничка расподела вероватноће, а још мање она која задовољава Марковљево својство са неким директним ацикличним графом? У теореми 3.2.2 је показано да ако почнемо са заједничком расподелом која задовољава Марковљево својство са неким директним ацикличним графом, вредности у тој заједничкој расподели биће дате производом условних расподела. Међутим, потребно је решити проблем у обрнутом смеру. Уколико на почетку имамо условне расподеле, потребно је закључити да је производ ових расподела заједничка расподела вероватноће која задовољава Марковљево својство са неким директним ацикличним графом. Теорема која следи омогућава баш то.

Теорема 3.2.3 *Нека је дат директан ациклични граф $\mathbb{G}$ у којем је сваки чвор случајна променљива, и нека је дефинисана дискретна условна расподела вероватноће сваког чвора у односу на дате вредности њихових родитеља у $\mathbb{G}$.*

Тада производ ових условних расподела даје заједничку расподелу вероватноће P, а $(\mathbb{G}, P)$ задовољава Марковљево својство.

Поред условних независности дефинисаних у Марковљевом својству, условне независности се могу извести и на други, једноставан, начин, а то је провером својства d-*сепарације* (коју смо детаљније образложили у дефиницији 8) преко графа што је увек довољан услов за условне независности у P.

Алгоритам д - сепарације користан је приликом израчунавања међусобне зависности чворова у Бајесовој мрежи. Уколико је за два чвора познато да су независна, онда познавање исхода једног од њих неће утицати на исход другог чвора. Тада није потребно трошити време на његово рачунање.

Дефиниција 12 *Нека је дат граф $(V, E), A \subseteq V, X$ и Y су чворови унутар скупа V и ρ је ланац између X и Y. Пут ρ је **блокиран** ако важи један од следећих услова:*

- *Постоји чвор $Z \in A$ на путу ρ и гране чвора Z су усмерене по принципу **глава - реп**.*
- *Постоји чвор $Z \in A$ на путу ρ и гране чвора Z су усмерене по принципу **реп - реп**.*
- *Постоји чвор Z, тако да ни Z нити било који од његових потомака није у A и гране чвора Z су усмерене по принципу **глава - глава**.*

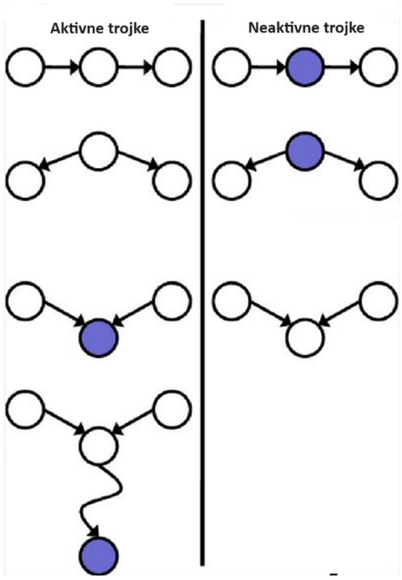
Може се десити да постоји један или више чворова који блокирају један ланац.

Дефиниција 13 *Нека је $\mathbb{G} = (V, E)$ ДАГ, $A \subseteq V$, и нека су X и Y различити чворови у V. Кажемо да су X и Y **d-сепарабилни** од стране A у $\mathbb{G}$ ако је сваки ланац између X и Y блокиран од стране A.*

Пример 4 Како би сликовито описали d-сепарацију и блокирање ланца, као и приближили ова два појма, погледајмо слику 3.5.

Кретање између нека два чвора X и Y може имати више путања, при чему се свака путања састоји од једне или више активних или неактивних тројки. На слици, обојени чворови представљају услов на основу којег се испитује условна независност.

- Путања је активна ако је свака тројка те путање активна, тј. ако је тројка нека од ова три случаја (на слици лево):



Слика 3.5: Провера условне независности са графа Бајесове мреже.[1]

- Тројка је серијска веза $A \to B \to C$ где B није услов за независност (важи за оба смера).
- Тројка је дивергентна веза $A \leftarrow B \to C$ где B није услов за независност.
- Тројка је конвергентна веза $A \to B \leftarrow C$ где је B или један од његових потомака услов за независност.

- A и C су условно независни у односу на B уколико су A и C d-сепарабилни у односу на B, тј. ако важи неки од три случаја када су тројке неактивне (на слици десно). Све што је потребно да се блокира путања, а самим тим и добије условна независност, је једна неактивна тројка у одређеној путањи.
- Уколико нема активних путања, добијамо независност између A и C.

Дакле, с обзиром на Марковљево својство, ако је чвор X d-сепарабилан од чвора Y датим чвором Z, X и Y су условно независни од Z. Тада се каже да су Бајесове мреже *мапа независности* у P.

Лема 1 *Нека је P расподела вероватноће променљивих у V и нека је $\mathbb{G} = (V, E)$ ДАГ. Тада $(\mathbb{G}, P)$ задовољава Марковљево својство ако и само ако за свака три међусобно дисјунктна подскупа $A, B, C \subseteq V$, кад год су A и B d-сепарабилни од C, A и B су условно независни у P датим C. То јест, $(\mathbb{G}, P)$*

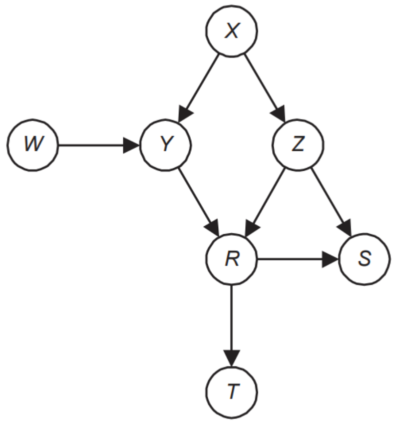
задовољава Марковљево својсӣво ако и само ако

$$I_{\mathbb{G}}(A, B|C) \Longrightarrow I_P(A, B|C).$$

Према претходној леми, ако су A и B d-сепарабилни датим C у $\mathbb{G}$, Марковљево својство подразумева $I_P(A, B|C)$. Из тог разлога, ако $(\mathbb{G}, P)$ задовољава Марковљево својство, кажемо да је $\mathbb{G}$ ***мапа независности*** од P. У наставку следи интуитивно објашњење зашто је свака d-сепарација условна независност.

Ако је $\mathbb{G} = (V, E)$ и $(\mathbb{G}, P)$ задовољава Марковљево својство, зависност у P између две промењиве у V постоји уколико ланац између ове две промењиве у $\mathbb{G}$ није блокиран на неки од наведена три начина. На пример, претпоставимо да P задовољава Марковљево својство са ДАГ-ом на слици 3.6. Постављамо питање да ли су X и T условно независни? Одговор ћемо дати базирајући се на примеру 4. Постоје три путање (ланца) од X до T, а то су $[X, Y, R, T]$, $[X, Z, R, T]$ и $[X, Z, S, R, T]$. У ланцу $[X, Y, R, T]$ имамо две активне тројке (обе имају серијску везу) $X \to Y \to R$ и $Y \to R \to T$, тако да је ова путања активна. Слично важи и за ланац $[X, Z, R, T]$, док ланац $[X, Z, S, R, T]$ није активан јер је блокиран од стране чвора S. У овом случају, прва тројка $X \to Z \to S$ је активна, међутим друга, која има конвергентну везу, $Z \to S \leftarrow R$ је неактивна. Већ смо напоменули да је довољно пронаћи једну неактивну тројку да би путања била неактивна. Међутим, иако смо пронашли једну путању која није активна, пронашли смо и две које су активне, а потребно је да нема активних путања, тако да не можемо гарантовати условну независност између X и T. Са друге стране, уколико посматрамо условну независност X и T у односу на $\{Y, Z\}$, X и T јесу условно независне зато што сада у ланцу $[X, Y, R, T]$ тројка $X \to Y \to R$ јесте неактивна јер је блокирана од стране Y, па је и путања неактивна. Слично важи и за ланац $[X, Z, R, T]$. Сада имамо све три путање које нису активне, јер у сваком ланцу постоји чвор који га блокира, што чини X и T независним. Дакле, према претходно дефинисаној леми 1, чињеница да важи $I_{\mathbb{G}}(\{X\}, \{T\}|\{Y, Z\})$ имплицира да важи $I_P(\{X\}, \{T\}|\{Y, Z\})$. Треба имати на уму да не можемо закључити да важи $I_P(\{X\}, \{T\}|\{Y, Z\})$ из Марковљевог својства, па самим тим не можемо знати да ли важи $I_{\mathbb{G}}(\{X\}, \{T\}|\{Y, Z\})$. У овом случају, све независности у расподели се читају директно са ДАГ-а.

Дефиниција 14 (Савршена мапа) *Ако важи да условна независносӣ ӣод-*



Слика 3.6: ДАГ за илустрацију блокирања ланца.[10]

разумева d-сепарацију (што није увек тачно за сваку расподелу), онда се каже да је P верна $\mathbb{G}$ или да је $\mathbb{G}$ савршена мапа за P.

Претпоставимо да имамо заједничку расподелу вероватноће P случајних променљивих у неком скупу V и ДАГ $\mathbb{G} = (V, E)$. Кажемо да је заједничка расподела вероватноће P **разложена на чиниоце** које чине условне вероватноће одређене на основу усмереног графа $\mathbb{G}$ ако

$$P(X_1, ..., X_n) = \prod_{i \in V} P(X_i | rod(X_i))$$

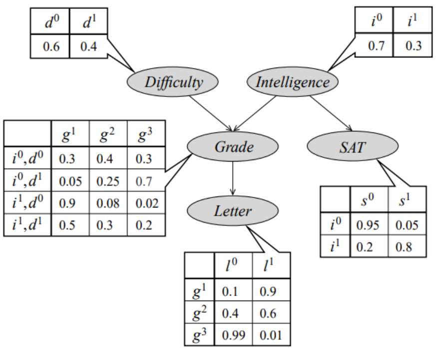
У том случају, P мора да задовољи све условне независности које имплицира граф.

Дефиниција 15 *Нека $I(\mathbb{G})$ означава скуп релација условних независности добијених применом својства d-сепарације на ДАГ $\mathbb{G}$, и нека $I(P)$ означава скуп релација условних независности које важе у расподели P. Кажемо да је $\mathbb{G}$ **I-мапа** за P ако $I(\mathbb{G}) \subseteq I(P)$. За I-мапу се каже да је **савршена** ако је $I(\mathbb{G}) = I(P)$.*

Такође, треба напоменути да, с обзиром на расподелу P, није увек могуће пронаћи ДАГ $\mathbb{G}$ такав да је $I(\mathbb{G}) = I(P)$.

Пример 5 На слици 3.7 разматрамо пример студента који на основу оцене (*Grade*) тражи препоруку од своје професорке (*Letter*). Оцена студента, у овом случају, зависи не само од његове интелигенције (*Intelligence)*, већ и од тежине курса (*Difficulty*), која је представљена као случајна променљива D

која узима вредности $Val(D)$ ={лако, тешко}. Дакле, наш студент тражи препоруку од своје професорке, при чему је професорка заборавна и никада се не сећа имена својих студената. Она може само да погледа његову оцену, и да да́ своју препоруку на основу тих информација. Квалитет њене препоруке је случајна промењива L која узима вредности $Val(L)$ ={јака, слаба}. Стварни квалитет препоруке зависи стохастички од оцене (може варирати у зависности од других фактора, као нпр. колико је професорка под стресом). Стога имамо пет случајних промењивих у овом: интелигенција студента (I), тежина курса (D), оцена (G), студентов резултат (S) и квалитет препоруке (L). Све промењиве осим G су бинарне вредности, а G је тернарна вредност. Дакле, заједничка расподела има 48 уноса.



d^0	d^1
0.6	0.4

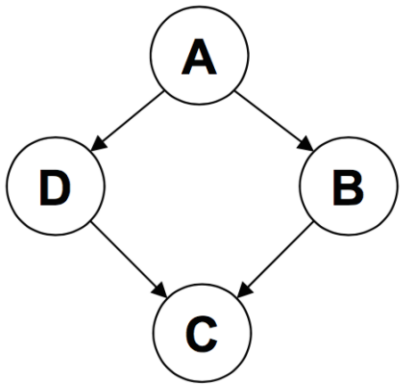
i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Слика 3.7: Пример где је $G_{student}$ савршена мапа за $P_{student}$.[6]



Слика 3.8: Пример где за P не постоји савршена мапа.[5]

На слици 3.7, све независности са графа тражимо у облику $\{X_i \perp NP(X_i) | rod(X_i)\}$, тј. тражимо све случајеве у којима је чвор X_i независан од свих својих непотомака $NP(X_i)$ с обзиром на скуп свих својих родитеља $rod(X_i)$. Све независности записујемо у $I(\mathbb{G})$, па ће бити

$$I(\mathbb{G}) = \{(D \perp I), (D \perp S), (G \perp S | I, D), (L \perp I, D, S | G), (S \perp D, G, L | I)\}.$$

Са друге стране, заједничка расподела је

$$P(I, D, G, S, L) = P(I) \underbrace{P(D|I)}_{P(D)} P(G|I, D) \underbrace{\overbrace{P(S|I, D, G)}^{S \perp D,G,L|I}}_{P(S|I)} \underbrace{\overbrace{P(L|I, D, G, S)}^{L \perp I,D,S|G}}_{P(L|G)}$$

Овде је употребљена чињеница да користимо правило ланца за Бајесове мреже $\prod_{X \in \{I,D,S,G,L\}} P(X|rod(X))$ (Теорема 3.2.1). Скуп $I(\mathbb{G})$ садржи све независности из расподеле $P(I,D,G,S,L)$, дакле G је савршена мапа за P.

Што се тиче слике 3.8, претпоставимо да имамо четири случајне променљиве A, B, C и D и скуп условних независности

$$I(P) = \{A \perp C|\{B,D\} \wedge B \perp D|\{A,C\}\}.$$

Међутим, овако дефинисане условне независности је немогуће приказати било којим ДАГ-ом. Као што је приказано на слици 3.8, у Бајесовој мрежи важи $A \perp C|\{B,D\}$ и $B \perp D|A$, али не важи $B \perp D|C$. У овом примеру, можемо видети да не постоји ниједан ДАГ за који важе ове условне независности, што доказује да нема свака расподела савршену мапу.

Јасно је да ако се инстанцирају сви родитељи променљиве X, сва деца од X и сви родитељи деце од X, онда је X d-сепарабилно од остатка мреже. Овај скуп променљивих је познат као *Марковљев покривач* променљиве X.

Дефиниција 16 (Марковљев покривач) *Марковљев покривач*[3] *променљиве X, $\boldsymbol{MB}(X)$, је скуп који се састоји од родитеља променљиве X, деце од X и променљивих које деле децу са X.*

Дакле, једини податак који је потребан како би се предвидело понашање X-а јесте $\mathbf{MB}(X)$.
Ако размотримо ДАГ дат на слици 3.6, Марковљев покривач променљиве Y је скуп чворова $\{X, W, R, Z\}$.

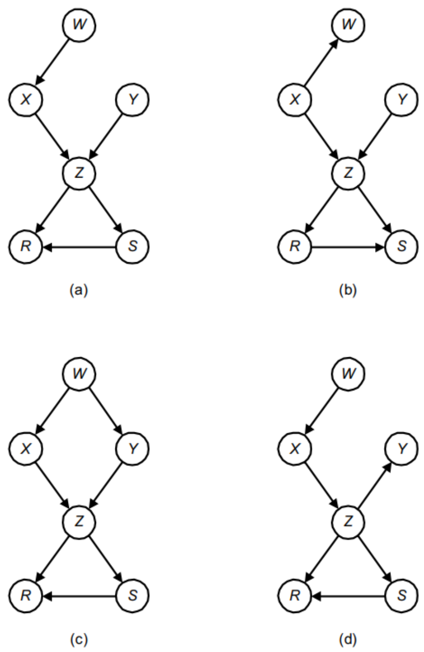
За дат скуп променљивих $V = \{X_1, ..., X_d\}$ може постојати неколико различитих ДАГ-ова од којих сваки представља исту структуру независности. За два ДАГ-а који представљају потпуно исту структуру независности се каже да су *еквивалентна у Марковљевом смислу.* На пример, БН са 3 чвора $X \to Y \to Z$, $X \leftarrow Y \to Z$ и $X \leftarrow Y \leftarrow Z$ су еквивалентна у Марковљевом смислу јер су наметнуте потпуно исте условне независности.

Теорема која следи омогућава да идентификујемо еквивалентност у Марковљевом смислу. Прво је потребно дефинисати следећи појам: **v-структура** је тројка чворова X, Y, Z таква да је $X \to Y \leftarrow Z$ конвергентна веза и не постоји лук који повезује X и Z.

[3]енг. Markov Blanket

Теорема 3.2.4 *Два ДАГ-а G_1 и G_2 су еквивалентна у Марковљевом смислу ако и само ако имају исте повезаности између чворова (без обзира на правац) и исти скуп в-структура.*

Пример 6 На слици 3.9, ДАГ-ови (a) и (b) су еквивалентна у Марковљевом смислу јер имају чворове који су исто повезани, као и исту v-структуру $X \to Z \leftarrow Y$. ДАГ (c) није еквивалентан у Марковљевом смислу са прва два зато што има повезану грану $W - Y$. Такође, ни ДАГ (d) није еквивалентан у Марковљевом смислу са прва два зато што, иако има чворове који су на исти начин повезани као код ДАГ-ова (a) и (b), нема исту v-структуру $X \to Z \leftarrow Y$.



Слика 3.9: ДАГ-ови (a) и (b) су Марков еквивалентни. ДАГ-ови (c) и (d) нису Марков еквивалентни са прва два ДАГ-а, као ни међусобно.[10]

Графички модел је пробабилистички модел где ДАГ приказује услов независности између случајних промењљивих. **Марковљева еквивалентна класа** је скуп ДАГ-ова који имају исти скуп условних независности. С друге стране, класу еквиваленције можемо дефинисати као скуп еквивалентних директних ацикличних графова. Наравно, сви графови у класи еквиваленције морају имати исти скуп чворова. С обзиром да се Марковљево својство дефинише помоћу графа и заједничке расподеле вероватноћа, графови у класи еквиваленције могу се посматрати као различите параметризације исте расподеле вероватноће. Корисна полазна тачка за проналажење свих ДАГ-ова који су еквивалентни у Марковљевом смислу са датим ДАГ-ом је да се пронађе *есенцијални граф*, дефинисан у наставку.

Дефиниција 17 (Есенцијални граф) *Есенцијални граф ДАГ-а D, означен као C, је граф који има исту повезаност између чворова, без обзира на правац, као D, и где је грана усмерена у C ако и само ако има исту оријентацију у сваком ДАГ-у који је еквивалентан са D.[3]*

Есенцијални граф C ДАГ-а D, поред горе наведене исте повезаности чворова, задржава и исту в-структуру као D.

Дакле, да сумирамо. Бајесове мреже се састоје од следећег:

1. Постоји скуп променљивих и скуп усмерених грана између променљивих.

2. Променљиве заједно са усмереним гранама чине ациклично усмерени граф (ДАГ).

3. За сваку променљиву A са родитељима $B_1, ..., B_n$ постоји табела условне вероватноће $P(A|B_1, ..., B_n)$. Ако је БН Бајесова мрежа над скупом $U = \{C_1, ..., C_n\}$, тада је БН одређена јединственом заједничком расподелом вероватноће $P(U)$ датом производом свих условних вероватноћа дефинисаних у БН: $P(U) = \prod_{i=1}^{n} P(C_i|rod(C_i))$.

3.3 Гаусове Бајесове мреже

Иако се већи део овог рада фокусира на дискретне променљиве, Бајесове мреже, као и факторизација расподеле, примењују се и на непрекидним променљивама. У овом делу рада, концентрисаћемо се на тип непрекидне расподеле која је посебно занимљива: класу вишедимензионих Гаусових расподела (илити заједничких нормалних расподела). Гаусове расподеле су посебно једноставна подкласа расподела који дају врло снажне претпоставке, попут експоненцијалног опадања расподеле удаљене од своје средње вредности, као и линеарне интеракције између променљивих. Иако су ове претпоставке често неисправне, Гаусове расподеле ипак изненађујуће добро апроксимирају многе расподеле у реалним ситуацијама.

3.3.1 Основна параметризација

На почетку рада је дефинисана једнодимензиона нормална (Гаусова) расподела (дефиниција 3). У наставку ће бити описана генерализација једнодимензионалне нормалне расподеле на више димензија.

Дакле, једнодимензиона нормална расподела је дефинисана помоћу два параметра: математичког очекивања и стандардне девијације. У својој најчешћој репрезентацији, вишедимензиону нормалну расподелу вектора $\mathbf{X} = (X_1, ..., X_n)$, у нотацији $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\Sigma})$, карактеришу n-димензиони вектор средње вредности $\boldsymbol{\mu}$ где је $\boldsymbol{\mu} = E(\mathbf{X}) = (E(X_1), E(X_2), ..., E(X_n))$ и симетрична $n \times n$ коваријациона матрица $\boldsymbol{\Sigma}$ где је $\Sigma_{i,j} := E[(X_i - \mu_i)(X_j - \mu_j)] = Cov[X_i, X_j]$ за $1 \leq i, j \leq n$; док је функција густине вероватноће најчешће дефинисана као:

$$f(x) = \frac{1}{(2\pi)^{n/2} det(\boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

где је $\boldsymbol{\Sigma}^{-1}$ инверзна матрица коваријациона матрице.

3.3.2 Трансформација Гаусове Бајесове мреже у вишедимензиону нормалну расподелу

Ако је rod_X скуп свих родитеља од X, онда се може записати:

$$X = w_X + \sum_{Z \in rod_X} b_{XZ} z, \tag{3.5}$$

где је w_X реализација случајне величине W_X која има функцију густине $\mathcal{N}(0; \sigma^2_{W_X})$, W_X је независно од сваког Z, а b_{XZ} је линеарни коефицијент регресије. Променљива W_X представља неизвесност вредности X у односу на вредности родитеља од X, а $\sigma^2_{W_X}$ дисперзију од X условљену вредностима њених родитеља. За сваки чвор X, његова функција густине је $\mathcal{N}(\mu_X; \sigma^2_X)$.

Размотрићемо опцију када не мора сваки W_X нужно да има средњу вредност 0 у једнакости (3.5), и у том случају W_X има густину $\mathcal{N}(E(W_X); \sigma^2_{W_X})$. Мрежа, у којој свака од ових променљивих има средњу вредност 0, може се добити из мреже која је одређена тако да се сваком чвору X додаје помоћни чвор, тј. родитељ Z, који има средњу вредност $E(W_X)$, дисперзију 0 и за који је $b_{XZ} = 1$. Ако променљива W_X у нашој новој мрежи има нормалну функцију густине са средњом вредношћу 0 и истом дисперзијом као одговарајућа променљива у нашој оригиналној мрежи, тада ће две мреже имати исту расподелу вероватноће.

Претпоставимо да имамо Гаусову Бајесову мрежу одређену као у једнакости (3.5), у којој не мора сваки W_X имати средњу вредност 0. Претпоставимо да смо поређали чворове у мрежи према тополошком уређењу (родитељи долазе испред своје деце у низу). Тада је сваки чвор линеарна функција која садржи вредности свих чворова који му претходе у поретку, где неки од коефицијената могу бити 0. Дакле, имамо

$$x_i = w_i + b_{i1}x_1 + b_{i2}x_2 + ... + b_{i,i-1}x_{i-1},$$

где W_i има густину нормалне расподеле $\mathcal{N}(E(W_i); \sigma_i^2)$ и $b_{ij} = 0$ ако X_j није родитељ од X_i, односно не постоји грана од X_j до X_i. Тада је условна функција густине за X_i:

$$f(x_i|rod_i) \sim \mathcal{N}(E(W_i) + \sum_{X_j \in rod_i} b_{ij}x_j; \sigma_i^2), \tag{3.6}$$

где је rod_i скуп свих родитеља од X_i. Како је

$$E(X_i) = E(W_i) + \sum_{X_j \in rod_i} b_{ij}E(X_j), \tag{3.7}$$

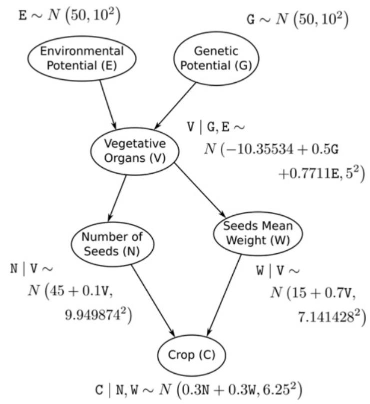
можемо одредити средњу вредност сваке променљиве X_i. Дакле, како би у потпуности одредили Гаусову Бајесову мрежу, користимо претпоставку да је за свако X_i средња вредност $\mu_i \equiv E(X_i)$ и условна дисперзија σ_i^2 (условљена скупом свих родитеља од X_i, rod_i). Захваљујући једнакости (3.7), имамо

$$E(W_i) = \mu_i - \sum_{X_j \in rod_i} b_{ij}\mu_j.$$

Када се замени овај израз за $E(W_i)$ у једначину 3.6, добија се условна функција густине за X_i:

$$f(x_i|rod_i) \sim \mathcal{N}(\mu_i + \sum_{X_j \in rod_i} b_{ij}(x_j - \mu_j); \sigma_i^2).$$

Пример 7 ДАГ на слици 3.10 представља мрежу усева (C) са чворовима који представљају еколошки потенцијал (E), генетски потенцијал (G), вегетативне органе (V), броја семена (N) и њихове просечне тежине (W). Локалне расподеле вероватноћа приказане су за сваки чвор. Слика преузета из књиге [14].



Слика 3.10

3.4 Динамичке Бајесове мреже

Досад разматрани модели Бајесових мрежа били су статични. У доменима који се развијају током времена (нпр. секвенцијална активација подручја мозга током когнитивног доношења одлука), потребне су нам *динамичке Бајесове мреже*. Уводи се дискретна временска ознака и исти локални модел се понавља за сваку јединицу времена. Тај локални модел представља део мреже назван *временски исечак* и приказује тренутну слику основног еволуирајућег временског процеса. Чворови унутар временског исечка t могу бити повезани са другим чворовима унутар истог исечка. Такође, временски исечци су међусобно повезани кроз временске или прелазне лукове који одређују како се променљиве мењају од једне временске тачке до друге. Временски лукови теку само унапред јер је стање променљиве у одређеном тренутку одређено стањима скупа променљивих у претходним тренуцима. У динамичким Бајесовим мрежама, структуре временских исечака су идентичне, а условне вероватноће такође остају идентичне током времена. Стога, динамичке Бајесове мреже су модели који су независни од времена, а динамичност заправо значи да могу моделирати динамичке системе.

Сада можемо дефинисати динамичке Бајесове мреже, које проширују Бајесове мреже за моделирање временских процеса. Претпостављамо да се промене дешавају између дискретних временских тачака, које су индексиране ненегативним целим бројевима, и да имамо неки коначан број тачака у времену T. Нека је $\{X_1, ..., X_n\}$ скуп променљивих чије се вредности мењају током времена, $X_i[t], 1 \leq i \leq n$, је случајна променљива која представља вредност X_i у тренутку t за $0 \leq t \leq T$, и нека је

$$\mathbf{X}[t] = \begin{pmatrix} X_1[t] \\ \vdots \\ X_n[t] \end{pmatrix}.$$

Динамичка Бајесова мрежа је Бајесова мрежа која садржи променљиве које чине случајан вектор дужине T, $\mathbf{X}[t]$, и која је одређена следећим спецификацијама:

1. Почетна Бајесова мрежа се састоји од почетног ДАГ-а $\mathbb{G}_0$ који садржи променљиве у $\mathbf{X}[0]$ и почетне расподеле вероватноће P_0 ових променљивих.

2. Транзициона Бајесова мрежа је модел који описује промену између два временска корака, обично означена као t и $t+1$. Овај модел се састоји из два главна елемента:
   а) ДАГ транзиције (усмерени ациклични граф транзиције), означен као $\mathbb{G}_{\rightarrow}$, који садржи све променљиве које су релевантне за временске кораке t и $t+1$. То укључује променљиве из $\mathbf{X}[t]$ и $\mathbf{X}[t+1]$ (вредности на временском кораку t и $t+1$). Ове променљиве су повезане усмереним стрелицама које представљају њихове међусобне зависности и како се једна вредност мења у другу током транзиције.
   б) Расподела вероватноће транзиције ($P_{\rightarrow}$), која додељује условне вероватноће свакој могућој вредности променљиве $\mathbf{X}[t+1]$ у односу на сваку могућу вредност променљиве $\mathbf{X}[t]$. Другим речима, за сваку комбинацију вредности $\mathbf{X}[t]$ и $\mathbf{X}[t+1]$, одређујемо вероватноћу да ће се појавити одређена вредност $\mathbf{X}[t+1]$ након транзиције, узимајући у обзир почетну вредност $\mathbf{X}[t]$. Ово омогућава моделу да предвиди како ће се променити стање система с обзиром на претходно стање. Дакле, за сваку вредност $x[t]$ и $x[t+1]$ наводимо

   $$P_{\rightarrow}(X[t+1] = x[t+1] | X[t] = x[t])$$

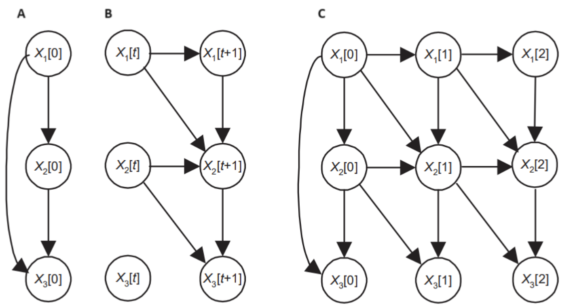
   .

3. Динамичка Бајесова мрежа која обухвата променљиве које чине случајан вектор дужине T састоји се од:

а) ДАГ-а који се састоји од ДАГ-а $\mathbb{G}_0$ и ДАГ-а $\mathbb{G}_{\rightarrow}$ процењеног у тренутку t, за $0 \leq t \leq T-1$; и

б) заједничке расподеле вероватноће

$$P(x[0], ..., x[T]) = P_0(x[0]) \prod_{t=0}^{T-1} P_{\rightarrow}(x[t+1]|x[t]). \tag{3.8}$$



Слика 3.11: Пример динамичке БН структуре са три променљиве X_1, X_2, X_3 и три временска исечка.**(A)** Претходна мрежа.**(B)** Транзициона мрежа са Марковљевом претпоставком првог реда.**(C)** Развој динамичке БН током времена у три временска исечка.[10]

Расподела вероватноће транзиције у мрежи на примеру 3.11 је

$$P_{\rightarrow}(x[t+1]|x[t]) = \prod_{i=0}^{n} P_{\rightarrow}(x_i[t+1]|rod_i[t+1]),$$

где $rod_i[t+1]$ означава вредности родитеља $X_i[t+1]$. Због једнакости 3.8, за свако t и за свако x:

$$P(x[t+1]|x[0], ..., x[t]) = P(x[t+1]|x[t]).$$

То значи да су све информације потребне за предвиђање стања у времену t садржане у опису стања у времену $t-1$. Није потребна никаква информација о ранијим тренуцима. Због ове особине, кажемо да процес има **Марковљево** својство. Осим тога, процес је **стационаран**[4]. То јесте, $P(x[t+1]|x[t])$ је исто за свако t. Генерално, није потребно да динамичка Бајесова мрежа има било који од ова два својстава, међутим, она смањују комплексност приказа и процене мреже. Процес не мора да се заустави у одређеном тренутку T, али се обично у пракси разматра неко коначно време.

[4] Стационаран процес је процес код којег статистичка својства остају константна с временом. То значи да стационаран процес не показује трендове, сезонске варијације или друге облике систематских промена с временом.

Треба напоменути да постављање усмерених лукова (грана) кроз време гарантује ацикличност графа, што је неопходно за БН. Динамички БН-ови су способни да моделирају рекурентне мреже, што је важно у неуронским системима, јер постоје цикличне функционалне мреже у мозгу, попут кортико-субкортикалних петљи[5].

Динамички БН-ови могу подразумевати потпуну или делимичну видљивост стања на чворовима. На пример, технике неуровизуализације пружају само индиректна посматрања неуронске активности одређеног региона интереса (РОИ, *eng. Region Of Interest*)[6], чије је стварно стање непознато. Скривена или латентна променљива може моделирати ову ситуацију.

3.4.1 Скривени Марковљеви модели

Скривени Марковљеви модели (*eng. Hidden Markov models*) су једноставни динамички БН-ови који се користе за моделирање Марковљевих процеса који се не могу директно посматрати, али се могу индиректно проценити на основу излазних података зависних од стања, односно, стање система није директно видљиво, али можемо посматрати излазне податке који су резултат тог стања. На пример, претпоставимо да имамо сензор који мери температуру у просторији, али немамо увид у тренутно стање да ли је укључено грејање или хлађење. Стање система (укључено или искључено грејање) је скривено, али излазни податак сензора (мерена температура) зависи од тог стања. Циљ је одредити оптимални низ стања који би могао да произведе посматрани низ излазних података.

Размотримо Бајесову мрежу описану графиконом на слици 3.12 и претпоставимо да променљиве $(X_1, X_2, X_3, ...)$ не могу бити посматране, али да вредности $(Y_1, Y_2, Y_3, ...)$ могу бити посматране. На пример, X_n може бити стање инфекције на дан n, а Y_n резултат теста на дан n.

[5]Кортико-субкортикалне петље су сложени неуронски путеви који повезују различите регије кортекса (спољашњи слој мозга одговоран за високе когнитивне функције) са субкортикалним структурама (структуре испод кортекса мозга). Ове петље играју кључну улогу у различитим аспектима неурофизиологије и понашања.

[6]РОИ је специфичан део мозга који истраживачи посебно проучавају у оквиру неуровизуализације студија. Овај регион је одређен анатомским структурама, функционалним областима или регијама које су повезане са одређеним когнитивним функцијама или поремећајима. Анализом активности у овим регијама, истраживачи могу добити увид у функционалне или структурне карактеристике мозга које су повезане са одређеним процесима или стањима.

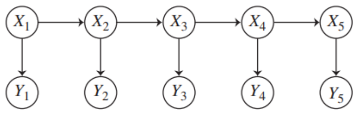
Модел претпоставља да су прошлост и будућност независне узимајући у обзир тренутно стање, тако да се секвенца променљивих $(X_j)_{j\geq 1}$ на примеру 3.12 формира као Марковљев ланац. Пошто се могу посматрати само променљиве $(Y_n)_{n\geq 1}$, секвенца $(X_n)_{n\geq 1}$ назива се *скривени Марковљев ланац*. Из графа,

$$P((X_1,Y_1),...,(X_5,Y_5)) = P(Y_5|X_5)P(X_5|X_4)P(Y_4|X_4)P(X_4|X_3)P(Y_3|X_3) \\ \times P(X_3|X_2)P(Y_2|X_2)P(X_2|X_1)P(Y_1|X_1)P(X_1).$$

Такође, из графа произилази следеће:

$$I_P(Y_n,Y_{n-1}|X_n), \qquad I_P(Y_{n+1},Y_n|X_n).$$

Другим речима, посматране променљиве су условно независне узимајући у



Слика 3.12: Илустрација скривеног Марковљевог модела.[7]

обзир секвенцу стања X_n.

Глава 4

Учење Бајесових мрежа из података

Структура и условне вероватноће неопходне за описивање Бајесове мреже могу бити обезбеђене или споља од стране стручњака, што је временски захтевно и подложно грешкама, или аутоматски, учењем из података. Учење Бајесових мрежа из података односи се на процес конструисања Бајесове мреже, односно вероватносног графичког модела, из посматраних података. Бајесове мреже су моћан алат за представљање и закључивање у условима неизвесности, што их чини широко коришћеним у различитим областима као што су машинско учење, вештачка интелигенција и системи за подршку о одлучивању. Процес учења Бајесових мрежа обично укључује два главна задатка која ћемо у наставку детаљније описати, а то су *структурно учење* и *параметарско учење*.

Учењем Бајесових мрежа из података можемо ухватити комплексне зависности и неизвесности у стварним системима, као и да вршимо задатке као што су вероватносно закључивање, предвиђање, откривање аномалија и доношење одлука. Ово чини Бајесове мреже вредним алатом за анализу података и подршку у одлучивању у широком спектру примена.

4.1 Параметарско учење

Претпоставимо да знамо структуру Бајесовог мрежног модела преко променљивих над простором $\mathcal{U}$[1], али без оцене за условне вероватноће. С друге

[1] Простор $\mathcal{U}$ представља скуп свих променљивих које су део Бајесове мреже.

стране, имамо приступ бази података са случајевима, што значи да за сваки случај у бази података имамо конкретне вредности за све (или већину) променљивих које се посматрају. Сада можемо користити ове случајеве за оцену параметара модела, односно условних вероватноћа. У овом поглављу прво разматрамо два приступа за решавање овог проблема када су подаци комплетни, а то су *Метод максималне веродостојности* и *Бајесова оцена*, а затим и један приступ када су подаци непотпуни, *ЕМ алгоритам*.

4.1.1 Метод максималне веродостојности

Овај метод (ММВ) у контексту Бајесових мрежа у неуронаучним истраживањима односи се на статистичку технику за оцену параметара модела који описују међусобне везе између различитих неуронских променљивих. Ова техника се фокусира на проналажење вредности параметара модела који максимизују веродостојност посматраних података.

У пракси, метод максималне веродостојности се користи за оцену параметара Бајесових мрежа на основу посматраних података, уз претпоставку да су подаци независни и једнако расподељени (*eng. iid*), што заправо значи да свако посматрање (инстанца) је независно од осталих посматрања. На пример, ако имамо податке о активностима неурона, активност у једном тренутку не зависи од активности у претходним или будућим тренуцима. Свако посматрање долази из исте расподеле вероватноће. То значи да сви подаци имају исте статистичке карактеристике, без промене током времена. Рецимо, претпоставимо да желимо да моделирамо како различити неурони у мозгу међусобно делују током одређеног задатка. Имамо скуп података који бележи активност неурона током времена, а свако посматрање представља тренутну активност свих неурона. Чворови представљају активности неурона, а гране представљају могуће условне зависности између њих. Затим прикупљамо *iid* податке о активностима неурона. Свако посматрање бележи тренутну активност свих неурона у одређеном тренутку. Користећи ММВ, оцењујемо параметре мреже (θ) који описују условне зависности између активности неурона, као што је оцењивање условних вероватноћа $P(X_i|rod(X_i))$ за сваки чвор.

Примена ММВ-а у неуронаучним истраживањима омогућава истраживачима да оцене вероватноће различитих параметара Бајесових мрежа на основу доступних података. Ова техника може бити корисна за моделирање и разумевање комплексних неуронских процеса, као и за анализу веза између неуронских промењљивих у мозгу.

Уопштено, ако имамо расподелу вероватноће са непознатим параметром θ, а подаци су генерисани према тој расподели, **оцена методом максималне веродостојности** (*eng. Maximum likelihood estimation*) параметра $\hat{\theta}$, на основу података, јесте вредност θ која чини податке највероватнијим. Другим речима, ако је D скуп наших података, тада је $\hat{\theta}$ заправо вредност θ која максимизује функцију веродостојности

$$L(\theta|D) = \prod_{d \in D} P(d|\theta).$$

Углавном се прелази на тражење максимума логаритма функције веродостојности која има облик:

$$l(\theta) = \log L(\theta|D) = \sum_{d \in D} \log P(d|\theta).$$

Оцена максималне веродостојности параметра θ је

$$\hat{\theta} = \underset{\theta}{argmax} L(\theta|D) = \underset{\theta}{argmax}\ \log l(\theta|D).$$

Ова оцена је она вредности параметра за коју је функција веродостојности највећа. Максимум даље налазимо одређивањем стационарне тачке $l(\theta)$, при чему $l(\theta)$ мора бити диференцијална функција, решавањем једначине

$$\frac{d}{d\theta} l(\theta) = 0.$$

Пример 8 Ако се шпенадла баци у ваздух, она ће пасти или на оштар врх (0) или на своју главу (1). Претпоставимо да се шпенадла баца n пута, при чему су, приликом сваког бацања појединачно, исходи независни и идентично расподељени (iid). Нека $\boldsymbol{X}_n$ означава низ резултата:

$$\boldsymbol{X}_n = (x_1, ..., x_n)^t.$$

Свако бацање има Бернулијеву расподелу са вероватноћом θ за успех (добијање 1), што означавамо са $X_i \sim Ber(\theta), i = 1, ..., n$. Наш задатак је да пронађемо добру вредност за параметар θ. Дефинишемо параметарски простор Θ — скуп могућности које разматрамо. У овом случају, наш параметарски

простор Θ је скуп свих параметара $\theta \in [0,1]$.
Функција веродостојности, $L(\theta)$, која даје вероватноћу посматраних података као функцију непознатих параметара модела је

$$L(\theta) \propto P(\boldsymbol{X}_n|\theta) = \prod_{l=1}^{n} \theta^{x_l}(1-\theta)^{1-x_l} = \theta^k(1-\theta)^{n-k},$$

где је $k = \sum\limits_{l=1}^{n} x_l$.
Да бисмо пронашли оцену методом максималне веродостојности за θ, постављамо једначину тако да је извод логаритма $P(\boldsymbol{X}_n|\theta)$ по θ једнак 0. Дакле,

$$\begin{aligned}\log(\boldsymbol{X}_n|\theta)) &= k \,\log\theta + (n-k)\log(1-\theta)\\ \tfrac{d\,\log(\boldsymbol{X}_n|\theta))}{d\theta} &= \tfrac{d(k\,\log\theta+(n-k)\log(1-\theta))}{d\theta}\\ &= \tfrac{k}{\theta} - \tfrac{n-k}{1-\theta}.\end{aligned}$$

Ако последњи израз изједначимо са 0, и једначину решимо по θ, добијамо

$$\theta = \frac{k}{n}.$$

Показивањем да је други извод логаритма у добијеној тачки негативан добијамо да је ово заиста максимум. Дакле, $\hat{\theta} = \frac{\sum_{l=1}^{n} X_l}{n}$, што смо и очекивали.

4.1.2 Бајесова оцена

Бајесова оцена у контексту Бајесових мрежа у неуронаучним истраживањима односи се на статистичку технику за процену параметара модела који описује међусобне везе између различитих неуронских промењљивих. Ова техника користи Бајесов приступ, који укључује претходна сазнања или уверења о параметрима у процес оцењивања, обично користећи Бајесове методе закључивања као што је узорковање Марковљевог ланца Монте Карло (*eng. MCMC*).

У практичној примени, Бајесова оцена омогућава истраживачима да процене вероватноће различитих параметара Бајесове мреже узимајући у обзир како доступне податке, тако и претходна сазнања или претпоставке о тим параметрима. Ова техника омогућава флексибилност и робустност у процени параметара, посебно у ситуацијама са малим скуповима података или када постоји неизвесност у вези са параметрима модела. Такође, ова техника може бити корисна за моделирање и разумевање сложених неуронских процеса и међусобних веза у мозгу.

Оцена параметара методом максималне веродостојности може имати недостатак када се примењује на ретке скупове података. У том случају, она може довести до проблема када појављивања одређених исхода буде 0. На пример, ако имате ретки скуп података који садржи само неколико случајева где су неки исходи незаступљени или се уопште не појављују (што је случај у Табели 4.1), ММВ би доделила тим исходима вероватноћу нула. Међутим, базирати закључке на оваквим ретким подацима може бити проблематично јер се претпоставља да су исходи са нултим бројем појављивања немогући, што је веома јака претпоставка заснована на малом броју случајева.

		Последња три слова							
		ааа	ааб	аба	абб	баа	бба	баб	ббб
Прва два слова	аа	2	2	2	2	5	7	5	7
	аб	3	4	4	4	1	2	0	2
	ба	0	1	0	0	3	5	3	5
	бб	5	6	6	6	2	2	2	2

Слика 4.1: Табела приказује број појављивања петословних речи ($T_1T_2T_3T_4T_5$) које су пренесене цртицом у следећи ред. Нпр, реч *abaab* се појавила 4 пута, а реч *bbabb* 6 пута.

Алтернатива за методу максималне веродостојности је Бајесова оцена. Овај процес се састоји из 2 корака:

- Почети са априорном расподелом. Пре него што се прикупе подаци, имамо неко почетно уверење о томе како су вероватноће различитих догађаја распоређене. Ово почетно уверење је априорна расподела. Она може бити заснована на претходном знању или претходним истраживањима. Међутим, она може бити и неинформативна или униформна ако немамо претходних информација.

- Користити искуства за ажурирање расподеле. Након што се прикупе подаци или стекне искуство, користимо их да ажурирамо своје почетно уверење. Ово се постиже применом Бајесове формуле која комбинује априорну расподелу са подацима како би се добила ажурирана расподела која боље описује стварне вероватноће.

Приступ се може илустровати Бајесовом мрежом, где је сваки параметар који се оцењује експлицитно дефинисан кроз чвор. Експеримент са бацањем шпенадле се поистовећује са моделом са речима, приказаног на слици 4.1, када бацамо спенадлу три пута где је нпр а=„пала је глава” и б=„пао је врх спенадле”, па би могући исходи бацања били као у наведеној табели. Означимо s_g као догађај да је приликом бацања шпенадла пала са главом окренутом ка горе, и слично, s_d догађај да је шпенадла пала са главом окренутом на доле. Условне вероватноће су $P(s_g|\theta) = \theta$, а априорна расподела $f(\theta)$ је, као и увек, на нама. Ако уопште немамо идеју за априорну расподелу, уобичајени приступ је коришћење неинформативне расподеле, а као најједноставнија се мозе узети униформна расподела са густином $f(\theta) = 1, 0 \leq \theta \leq 1$.

Претпоставимо да смо извршили један експеримент са резултатом „глава нагоре”. Користећи Бајесово правило, добијамо

$$f_p(\theta|s_g) = \frac{P(s_g|\theta)f(\theta)}{P(s_g)} = \frac{\theta f(\theta)}{P(s_g)}$$

за апостериорну густину f_p. Ако поставимо $f(\theta) = 1$, добијамо

$$f_p(\theta|s_g) = \frac{\theta}{P(s_g)}, \ \ P(s_g) = \int_0^1 \theta d\theta = \frac{1}{2}.$$

Тако да је

$$f_p(\theta|s_g) = 2\theta,$$

што је заправо апостериорна расподела за вредност параметра θ под условом да је глава пала на горе, па је најбоља оцена параметра θ, када је функција губитака квадратна, средња вредност ове расподеле:

$$\int_0^1 \theta \cdot 2\theta d\theta = \frac{2}{3}.$$

Затим, претпоставимо да приликом још једног бацања добијемо шпенадлу окренута главом надоле. Тада имамо нову апостериорну расподелу:

$$\begin{aligned} f_{p2}(\theta|s_d, s_g) &= \tfrac{P(s_d, s_g|\theta)f(\theta)}{P(s_g, s_d)} \\ &= \mu P(s_d|\theta)P(s_g|\theta)f(\theta) \\ &= \mu P(s_d|\theta)\theta \cdot 1 = \mu(1-\theta)\theta, \end{aligned}$$

где је μ нормирајућа константа

$$\frac{1}{\mu} = P(s_d, s_g) = \int_0^1 (1-\theta)\theta d\theta = \frac{1}{6}.$$

Апостериорна функција $f_{p2}(\theta|s_d, s_g)$ може бити записана као

$$f_{p2}(\theta|s_d, s_g) = 6(1-\theta)\theta,$$

и најбоља оцена параметра θ је

$$\int_0^1 6\theta(1-\theta)\theta d\theta = \frac{1}{2}.$$

4.1.3 EM алгоритам

Један од најпопуларнијих алгоритама за оцену параметара је EM алгоритам (*eng. Expectation-maximization algorithm*). EM алгоритам је општи алгоритам за проналажење оцена методом максималне веродостојности за скуп параметара $\boldsymbol{\theta}$ када се суочавамо са непотпуним скупом података. Алгоритам се извршава наизменично између такозваног *корака очекивања* (*eng. e-step*) и *корака максимизације*(*eng. m-step*) : у слободном преводу, у кораку очекивања „допуњујемо" скуп података коришћењем тренутних оцена параметара $\hat{\boldsymbol{\theta}}$ како бисмо израчунали очекивања за недостајуће вредности, а у кораку максимизације користимо „допуњени" скуп података да пронађемо нову оцену максималне веродостојности $\hat{\boldsymbol{\theta}}'$ за параметре. Ова оцена се затим користи за комплетирање скупа података у следећој итерацији алгоритма. Алгоритам се наставља или за унапред одређени број итерација или док алгоритам не конвергира.

Претпоставимо да имамо структуру модела B над променљивама $U = \{X_1, ..., X_n\}$, и нека θ_{ijk} означава параметар који одговара условној вероватноћи $P(X_i = k|rod(X_i) = j)$, тј. условној вероватноћи да је променљива X_i у својој k-тој вредности под условом да су родитељи од X_i у својој j-тој конфигурацији (нпр. $rod(X_i) = j$). Користећи ову нотацију, можемо пронаћи оцену максималне веродостојности, $\hat{\theta}_{ijk}$, за параметре θ_{ijk} датог скупа података $D = \{\boldsymbol{d}_1, ..., \boldsymbol{d}_m\}$ са m случајева на следећи начин:

1. *Изаберемо* $\epsilon > 0$ *да би регулисали критеријум заустављања.*

2. *Нека су* $\boldsymbol{\theta}^0 = \{\theta_{ijk}\}$, *где је* $1 \leq i \leq n, 1 \leq k \leq |sp(X_i)| - 1$, *и* $1 \leq j \leq |sp(rod(X_i))|$ *неке почетне оцене параметара (изабране произвољно), при чему* $|sp(X_i)|$ *представља број могућих стања за променљиву* X_i.

3. *Поставимо* $t := 0$.

4. *Понављамо:*

   e-step: За свако $1 \leq i \leq n$ израчунава се табела очекиваних фреквенција:

$$\mathbb{E}_{\boldsymbol{\theta}^t}[N(X_i, rod(X_i))|\mathcal{D}] = \sum_{\boldsymbol{d} \in \mathcal{D}} P(X_i, rod(X_i)|\boldsymbol{d}, \boldsymbol{\theta}^t).$$

   m-step: Користе се очекивани бројеви као да су стварни бројеви да би израчунали нову оцену максималне вероватноће за све θ_{ijk}:

$$\hat{\theta}_{ijk} = \frac{\mathbb{E}_{\boldsymbol{\theta}^t}[N(X_i = k, rod(X_i) = j)|\mathcal{D}]}{\sum\limits_{h=1}^{|sp(X_i)|} \mathbb{E}_{\boldsymbol{\theta}^t}[N(X_i = h, rod(X_i) = j)|\mathcal{D}]}.$$

   Поставимо: $\boldsymbol{\theta}^{t+1} := \hat{\boldsymbol{\theta}}$ *и* $t := t + 1$

   Све док је $|log_2 P(\mathcal{D}|\boldsymbol{\theta}_t) - log_2 P(\mathcal{D}|\boldsymbol{\theta}_{t-1})| \leq \epsilon$.

□

4.2 Структурно учење

У претходном поглављу смо анализирали параметарско учење Бајесових мрежа где смо имали јаку претпоставку да унапред знамо структуру мреже или да се бар одлучимо за неку структуру без обзира на то да ли је тачна или не. У овом поглављу разматрамо задатак учења у ситуацијама када унапред не знамо структуру Бајесове мреже. Кроз цело ово поглавље ћемо се водити (веома) јаком претпоставком да је наш скуп података у потпуности доступан, у смислу да не постоје недостајући подаци или празне вредности за било коју променљиву у скупу података. Постоје у основи две методе које се користе за учење структуре Бајесових мрежа, а то су **методе засноване на ограничењима** *(eng. constrained-based methods)* и **методе засноване на оцењивању** *(eng. score-based methods)*. Поред ове две основне методе, постоји и хибридни метод (*eng. hybrid method*), којег нећемо обрађивати у овом раду.

Нека $\mathbf{X}$ представља скуп свих променљивих које посматрамо. Сви ови алгоритми за учење структуре функционишу под скупом заједничких претпоставки:

- Мора постојати једнозначна кореспонденција између чворова у ДАГ-у и случајних променљивих у $\mathbf{X}$, тј. треба да постоји тачно једна веза између сваког чвора и променљиве, без преклапања или вишеструких веза. То посебно значи да не сме бити више чворова који су детерминистичке функције[2] једне променљиве.

- Сви односи између променљивих у $\mathbf{X}$ морају бити условне независности, јер су они по дефиницији једини тип односа који може бити изражен Бајесовом мрежом.

- Свака комбинација могућих вредности променљивих у $\mathbf{X}$ мора представљати валидан (чак и ако је веома мало вероватан) догађај. Ова претпоставка подразумева строго позитивну глобалну расподелу, што значи да свака комбинација могућих вредности променљивих има позитивну вероватноћу. Она је потребна за јединствено одређивање Марковљевих покривача и, самим тим, једнозначно препознатљив модел. Алгоритми засновани на ограничењима функционишу чак и када ово није тачно, јер је постојање савршене мапе такође довољан услов за јединственост Марковљевих покривача.

- Сваки податак или посматрање се у скупу података сматра независним од других посматрања. Ако је присутан неки облик временске или просторне зависности, то мора бити посебно узето у обзир приликом дефинисања мреже, као што је случај код динамичких Бајесових мрежа који су пример модела који експлицитно укључују временску зависност између посматрања.

4.2.1 Методе учења засноване на ограничењима

Најпре ћемо размотрити следећи проблем: Морамо да одредимо структуру Бајесове мреже, а једини извор информација су истинити одговори на питања попут „Да ли је променљива A d-сепарабилна од променљиве B у односу на скуп X?“. Нека $I(A, B, X)$ означава да је A d-сепарабилан од B у односу на X. Користимо $I(A, B)$ као скраћеницу за $I(A, B, \emptyset)$, а ако X садржи само један елемент C, пишемо $I(A, B, C)$.

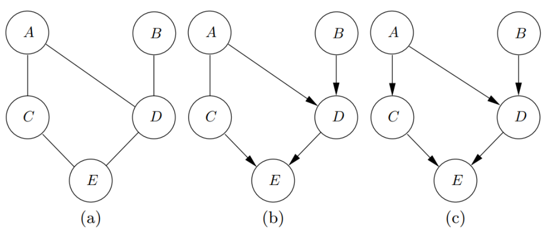
[2]Детерминистичка функција је функција која за сваки улаз даје тачно један излаз, без икакве несигурности или варијабилности. На пример, функција $f(x) = 2x + 3$ је детерминистичка јер за сваки улаз x даје тачно одређен излаз $2x + 3$.

Дефиниција 18 *Скелет Бајесове мреже N је неусмерени граф који се добија уклањањем стрелица са свих грана у N.*

Метод се састоји у томе да се прво одреди скелет мреже, а затим да се усмере везе:

1. Скелет се може прилично лако утврдити путем низа логичких закључака: ако постоји веза између A и B, онда они не могу бити d-сепарабилни. То јест, веза $A - B$ је део скелета акко $\neg I(A, B, X)$ за сваки скуп X који не садржи A или B. Као почетну тачку, претпоставимо да имамо скелет.

2. Усмерити везе:
   Правило 1 [увођење v-структура]: Ако имамо три чвора A, B, C таквих да $A - C$ и $B - C$, али не и $A - B$, онда се уводи v-структура $A \to C \leftarrow B$ ако постоји скуп X (могуће празан) такав да $I(A, B, X)$ и $C \notin X$.
   Пример 4.2: Размотримо скелет са слике 4.2 (a) са независностима $I(A,B), I(B,C), I(A,B,C), I(B,C,A), I(C,D,A), I(B,C,\{D,A\})$, $I(C,D,\{A,B\}), I(B,E,\{C,D\}), I(A,E,\{C,D\}), I(B,C,\{A,D,E\})$, $I(A,E,\{B,C,D\}), I(B,E,\{A,C,D\})$. Размотримо ланац $C - E - D$. С обзиром да E није члан скупа услова који чини C и D независним, уводимо v-структуру $C \to E \leftarrow D$. На исти начин уводимо v-структуру $A \to D \leftarrow B$ (Слика 4.2 (b)). Са ове две в-структуре, не може бити више њих. Ово такође потврђују условне независности, и пошто оне не пружају никакав траг о преосталој вези $A - C$, она може бити усмерена у било ком смеру (Слика 4.2 (c)).



Слика 4.2: (a) Скелет. (b) Две в-структуре уведене кроз правило 1. (c) Употпуњен ДАГ.[4]

Правило 2 [Избегавање нових v-структура]: Када смо применили правило 1 и више нема ситуација које одговарају том правилу, и имамо случај $A \to C - B$ (и нема веза између A и B) онда можемо усмерити $C \to B$.

Правило 3 [Избегавање циклуса]: Ако $A \to B$ резултира формирање усмереног циклуса у графу, тада правило налаже да се та веза постави у супротном смеру, $B \to A$.

Правило 4 [Насумично бирање]: Ако ниједно од правила 1–3 не може да се примени било где у графу, треба изабрати неусмерену везу и дати јој произвољан правац. На пример, након што смо пронашли в-структуре на слици 4.2 (b), можемо изабрати било који правац за $A - C$ (Слика 4.2 (c)).

Теорема 4.2.1 *Чворови A и B нису повезани у Бајесовој мрежи N акко $I(A, B, rod(A))$ или $I(A, B, rod(B))$.*

Теорема осигурава да је довољно знати информације о независностима облика $I(A, B, X)$ где је X подскуп суседа чвора A или B. Користи се у *PC алгоритму (енг. Peter-Clark algorithm)* који се фокусира на локална питања независности, тј. питања која се односе само на одређене чворове и њихове непосредне суседе.

Алгоритам 1 (PC алгоритам):

Проблем: Дат је скуп IND који садржи све d-сепарације и нека ADJ_X подразумева подскуп скупа V који се састоји од свих чворова који су суседни чвору X. Треба одредити, ако постоји, ДАГ који је веран скупу IND, односно ДАГ који треба тачно да одражава условне независности специфициране у скупу IND.

Улазни параметри: Скуп чворова V и скуп свих d-сепарација IND.

Излазни параметри: Уколико је скуп IND компатибилан са структуром ДАГ-а, ДАГ-а gp садржи d-сепарације из овог скупа.

```
PC <- function(skup_cvorova V, skup_dseparacija IND, graf gp){
  i<-0;
  repeat
    for(svaki X ∈ V)
      for(svaki Y ∈ ADJ_X){
        ispitati da li postoji skup S ⊆ ADJ_X − {Y}
        takav da |S| = i i I({X},{Y}|S) ∈ IND;
        if ako je takav skup S nadjen{
```

```
          S_XY = S;
          ukloniti granu X - Y iz gp; #Korak 1
         }
        }
        i<-i+1;
      until(|ADJ_X| < i za sve X ∈ V);
      for(za svaki nepovezani susret X - Z - Y)
        if(Z ∉ S_XY)
          orijentisati X - Z - Y kao X → Z ← Y; #Korak 2
      while(dokle god ima vise grana koje treba usmeriti){
        for(za svaki nepovezani susret X → Z - Y);
            orijentisati Z - Y kao Z → Y #Korak 3
        for(za svako X - Y takvo da postoji put od X do Y)
            orijentisati X - Y kao X → Y; #Korak 4
        for(za svaki nepovezani susret X - Z - Y
            takav da X → W,Y → W, i Z - W)
            orijentisati Z - W kao Z → W; #Korak 5
       }
}
```

4.2.2 Методе учења засноване на оцењивању

У претходном одељку смо видели како се изводи структурно учење засновано на ограничењима, а у овом одељку ћемо се фокусирати на други тип учења, који се назива учење засновано на оцењивању (*eng. score-based learning*). Овакво учење додељује број (оцену) свакој структури Бајесове мреже. Оцена описује „корисност” структуре, при чему термин „корисност” може, на пример, бити показатељ колико је вероватно да је структура могла бити коришћена за генерисање дате базе података.

Ако имамо функцију оцењивања која узима структуру Бајесове мреже као аргумент и враћа вредност, тада се задатак учења заснованог на оцењивању може сматрати проблемом претраге: једноставно тражимо структуру модела са највишом оценом. То такође значи да се алгоритам учења заснованог на оцењивању може у принципу потпуно описати помоћу две компоненте, а то су (1) *функција оцењивања (eng. score function)* и (2) *методе претраге простора* *(eng. search space methods).*

4.2.2.1 Методе претраге простора

У структурном учењу за Бајесове мреже, методе претраге простора су кључне за ефикасно проналажење најбоље структуре мреже која представља зависности међу променљивама на основу података. Простор претраге се састоји од свих могућих структура које мрежа може представљати, а различите методе претраживања простора помажу у управљању сложеношћу и величином овог простора. Свакој кандидатској Бајесовој мрежи се додељује оцена мреже која одражава њену прилагођеност, а алгоритам потом покушава да ту оцену максимизује. Неки примери таквих алгоритама су *greedy search algorithms, genetic algorithms, simulated annealing.* Најчешће коришћен алгоритам, представник *greedy search* алгоритма, јесте ***hill-climbing*** [14]. Овај алгоритам функционише тако што у свакој итерацији, локална претрага се врши све док се не постигне локални максимум. Затим се тренутна структура насумично поремети, и процес се понавља. На крају, користи се максимална вредност свих локалних максимума.

Алгоритам 2 (Hill-Climbing алгоритам):

1. Одаберати структуру мреже G која је обично (али не нужно) празна.
2. Израчунати оцену мреже G која се означава као $Score_G$.
3. Поставити $maxscore = Score_G$.
4. Понављати следеће кораке све док се $maxscore$ повећава:
   a) За свако могуће додавање, брисање или обртање гране које не резултира цикличном мрежом:
      i. Израчунати оцену модификоване мреже $G*$, $Score_{G*}$.
      ii. Ако је $Score_{G*} > Score_G$, поставити $G = G*$ и $Score_G = Score_{G*}$.
   b) Ажурирати $maxscore$ новом вредношћу $Score_G$.
5. Вратити ДАГ G.

Поред поменутих метода, када је у питању структурно учење, велику улогу игра, већ поменута, функција оцењивања, тако да ћемо се ми у овом раду више фокусирати на њеном детаљном опису.

4.2.2.2 Функција оцењивања

Уопштено, проблем избора модела је пронаћи концизан модел који, на основу случајног узорка посматрања из популације која одређује расподелу релативних фреквенција, укључује апроксимацију расподеле релативних фреквенција. Да бисмо извршили избор модела, развијамо **функцију оцењивања** *score* (критеријум оцењивања) која додељује вредност $score(d, \mathcal{M})$ сваком моделу који се разматра на основу података, при чему је d скуп вредности (података) узорка.

Дакле, тражимо функцију оцењивања која би требало да задовољава следећа два својства:

- Треба пронаћи оптималну равнотежу између тачности и сложености структуре.
- Треба да буде рачунски изводљиво за процену.

У наставку следи детаљнији опис неких од најчешће коришћених функција оцењивања.

Бајесов информативни критеријум

Пример функције оцењивања која задовољава наведене две особине је Бајесов информативни критеријум (*енīл. BIC*), који садржи податак који мери колико добро подаци одговарају моделу, као и податак који узима у обзир сложеност модела:

$$BIC(S|D) = log_2 P(D|\hat{\theta_S}, S) - \frac{size(S)}{2} log_2(N),$$

где је S скуп свих параметара у мрежи, $\hat{\theta_S}$ оцена параметара добијена методом максималне веродостојности за скуп S и N је величина узорка.

Напомена: Постоји још једна функција оцењивања, доста слична са BIC оценом, која се зове ***AIC***(*Akaike Information Criterion*) и њена формула је

$$AIC(S|D) = log_2 P(D|\hat{\theta_S}, S) - size(S).$$

Даље, ако додатно претпоставимо да су случајеви независни с обзиром на модел, тада

$$BIC(S|D) = \sum_{i=1}^{N} log_2 P(d_i|\hat{\theta_S}, S) - \frac{size(S)}{2} log_2(N).$$

Да бисмо оценили модел користећи BIC, почињемо од оцене параметара методом максималне веродостојности за модел. С обзиром да је база података потпуна, онда је ово заправо пребројавање фреквенција. На основу ових оцена израчунава се вероватноћа за сваки случај у бази података. Ово се може урадити једноставним уносом вредности променљивих из одређеног случаја (податка) у Бајесову мрежу и извођењем пропагације[3]. Ако су сви случајеви потпуни, онда је задатак још једноставнији, потребно је, да би израчунали укупну вероватноћу комплетног случаја, помножити вероватноће из табеле условних вероватноћа за сваку променљиву. Ово заузврат даје број фреквенција добијених из базе података. То су заправо статистике које се добијају једноставним бројањем појављивања различитих вредности променљивих и њихових комбинација у скупу података. Ово такође значи да је израчунавање BIC-а сведено на проблем пребројавања: нека r_i означава број стања за променљиву X_i, и нека $q_i = \prod_{X_l \in \prod_i} r_l$ означава број конфигурација[4] над родитељима за X_i у S (ако X_i нема родитеље, тада нека буде $q_i = 1$). Са овом нотацијом сада имамо

$$BIC(S|D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} log_2 \left(\frac{N_{ijk}}{N_{ij}} \right) - \frac{log_2 N}{2} \sum_{i=1}^{n} q_i(r_i - 1),$$

где је N_{ijk} број случајева у бази података који задовољавају следећа два услова: X_i има вредност која одговара њеној k-тој конфигурацији и родитељске променљиве $rod(X_i)$ имају вредности које одговарају њиховој j-тој конфигурацији.

Модел са нижим BIC резултатом се сматра бољим јер представља добар компромис између прецизности и једноставности.

Бајесово оцењивање

BIC је пример функције оцењивања који комбинује члан који мери прилагођеност модела (максимум функције веродостојности модела) и члан који мери комплексност. Други приступ за мерење подобности структуре модела Бајесове мреже $\mathbb{G}$ је Бајесов приступ који покушава да пронађе струк-

[3]Пропагација је процес рачунања вероватноћа у Бајесовој мрежи након уноса вредности променљивих. У овом контексту, пропагација омогућава да се ажурирају вероватноће свих других променљивих у мрежи с обзиром на унете вредности.

[4]Број конфигурација се односи на број свих могућих комбинација вредности које родитељске променљиве неке променљиве могу имати.

туру са максималном апостериорном вероватноћом датих података, то јест, $arg\, max P(\mathbb{G}|\mathcal{D})$. Користећи Бајесову формулу, имамо:

$$P(\mathbb{G}|\mathcal{D}) = \frac{P(\mathcal{D},\mathbb{G})}{P(\mathcal{D})} = \frac{P(\mathbb{G})P(\mathcal{D}|\mathbb{G})}{P(\mathcal{D})} = \mu P(\mathbb{G})P(\mathcal{D}|\mathbb{G}), \tag{4.1}$$

где је $\mu = \frac{1}{P(\mathcal{D})}$ нормирајућа константа. Ова константа не зависи од $\mathbb{G}$, стога није потребно рачунати је када упоређујемо две структуре мреже. Заправо, ако би хтели да израчунамо $P(\mathcal{D})$, суочили бисмо се са рачунским проблемом, јер израчунавање ове константе подразумева сумирање преко свих могућих структура модела, тј. $P(\mathcal{D}) = \sum_B P(B)P(\mathcal{D}|B)$.

Из једначине (4.1) можемо видети како бисмо оценили структуру на основу њене апостериорне вероватноће с обзиром на податке, међутим потребне су нам две вредности, односно априорна расподела структуре $P(\mathbb{G})$ и *маргинална вероватноћа* података $P(\mathcal{D}|\mathbb{G})$ коју треба максимизирати. Обично се одабере априорна расподела вероватноће за структуре која је релативно лака за израчунавање, док је главни рачунски проблем израчунавање маргиналне вероватноће, дефинисане као:

$$P(\mathcal{D}|\mathbb{G}) = \int_{\boldsymbol{\theta}_{\mathbb{G}}} P(\mathcal{D}|\mathbb{G}, \boldsymbol{\theta}_{\mathbb{G}}) f(\boldsymbol{\theta}_{\mathbb{G}}|\mathbb{G}) d\boldsymbol{\theta}_{\mathbb{G}}, \tag{4.2}$$

где је $f(\boldsymbol{\theta}_{\mathbb{G}}|\mathbb{G})$ априорна расподела вероватноће над параметрима (условним вероватноћама) за $\mathbb{G}$. Интеграл у горњој једначини обухвата све параметре, и, у суштини, све могуће Бајесове мреже са истом структуром, али са различитим условним расподелама вероватноће.

K2 алгоритам

Дакле, за израчунавање вероватноће $P(\mathbb{G}|\mathcal{D})$ тежи део представља оцена интеграла из једначине (4.2). Срећом, показано је да се уместо оцене комплексног интеграла, проблем може решити пребројавањем одређених конфигурација или појава у подацима, и то на основу следећих шест претпоставки:

1. Подаци у бази $\mathcal{D}$ тачно одражавају зависности и независности између променљивих како су представљене у одређеној Бајесовој мрежи.

2. Случајеви у бази $\mathcal{D}$ су независни према БН моделу.

3. База података је комплетна.

4. Априорна расподела параметара у свакој Бајесовој мрежи је униформна.

5. [Локална независност] За било које две конфигурације родитеља за променљиву X_i, параметри за условне расподеле вероватноћа повезане са X_i су независни.

6. [Глобална независност] Густине параметара за условне расподеле вероватноћа за X_i и X_j су независне за $i \neq j$.

На основу ових претпоставки, доказана је следећа теорема:

Теорема 4.2.2 (К2 оцењивање) *Нека је D база података над променљивама $X_1, X_2, ..., X_n$, и размотримо структуру Бајесове мреже B_s над истим скупом променљивих. Са N_{ijk} означавамо број случајева у бази података који укључују конфигурацију $(X_i = k, rod(X_i) = j)$. Узимајући у обзир претходних шест претпоставки, важи формула за К2 оцењивање:*

$$P(\mathcal{D}|\mathbb{G}) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$$

где је $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

На сличан начин, претпостављајући униформну расподелу за $P(\mathbb{G})$ и Дирихлеову расподелу са параметрима a_{ijk} за $f(\boldsymbol{\theta}_{\mathbb{G}}|\mathbb{G})$, добијамо следећу теорему када постоји n променљивих из мултиномне расподеле[5]:

Теорема 4.2.3 *Претпоставимо да имамо ДАГ $\mathbb{G} = (V, E)$ где је V скуп мултиномијалних случајних променљивих. Користимо Дирихлеову расподелу да представимо наше претходно уверење за сваку условну расподелу вероватноће за сваку променљиву у скупу V. Такође, претпоставимо да имамо податке D који се састоје од скупа података тако да је сваки податак вектор вредности свих променљивих у V. Тада је*

$$score(\mathbb{G} : D) = P(D|\mathbb{G}) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})}$$

где је

[5] Променљиве које могу имати више од две могуће категоријске вредности, и те вредности су из мултиномне расподеле.

1. *n је број променљивих;*
2. *q_i је број различитих могућих комбинација вредности родитељских променљивих од X_i;*
3. *a_{ijk} је наше утврђено претходно веровање о броју пута када је X_i узела своју k-ту вредност и када су родитељи од X_i имали своју j-ту инстанцу;*
4. *s_{ijk} представља број пута у подацима када је X_i узела своју k-ту вредност и када су родитељи од X_i имали своју j-ту инстанцу;*
5. *N_{ij} и M_{ij} су заправо следеће: $N_{ij} = \sum_k a_{ijk}$ и $M_{ij} = \sum_k s_{ijk}$.*

Вероватноћу $P(D|\mathbb{G})$, добијену користећи претпоставке из претходне теореме, називамо **Бајесово оцењивање уз претпоставку Дирихлеове априорне расподеле**.

Приметимо у теореми 4.2.3 да је оцена за цео ДАГ производ локалних оцена за сваки чвор. Нека ROD_i означава скуп родитеља од X_i у $\mathbb{G}$. За сваки чвор X_i, вредност

$$\prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})}$$

зависи само од вредности X_i и чворова у ROD_i, као и од вредности параметара локално смештених у X_i, тј. параметара који укључују вероватноће да X_i узме одређене вредности условљене вредностима његових родитеља. Означимо

$$score(X_i, ROD_i, D) = \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})}.$$

Да би приближно максимизовали $score(\mathbb{G} : D)$, што је и циљ, за свако i, тј. за сваки чвор, алгоритам тражи скуп родитеља ROD_i који ће максимизовати локалну оцену $score(X_i, ROD_i, D)$ за тај чвор. Једина операција у овом алгоритму претраге јесте додавање родитеља чвору. Алгоритам K2 функционише на следећи начин: Претпостављамо да је редослед чворова такав да ако X_i претходи X_j у редоследу, смер од X_j до X_i није дозвољен. Нека је $Pred(X_i)$ скуп чворова који претходе X_i у редоследу. Ови чворови су потенцијални кандидати за родитеље чвора X_i. Иницијално постављамо скуп родитеља ROD_i чвора X_i на празан скуп и израчунавамо $score(X_i, ROD_i, D)$, затим итеративно пролазимо кроз чворове редом према редоследу. Кад стигнемо до чвора X_i, одређујемо чвор из скупа $Pred(X_i)$ који највише повећава локалну оцену $score(X_i, ROD_i, D)$ и онда тај чвор додајемо скупу ROD_i. Процес се понавља све док додавање новог чвора не престане да повећава оцену $score(X_i, ROD_i, D)$. Када више ниједан чвор не може повећати оцену, алгоритам престаје да додаје нове родитеље.

Алгоритам 2 (K2 алгоритам):
Проблем: Пронаћи ДАГ $\mathbb{G}'$ који приближно максимизује $score(\mathbb{G} : D)$.
Улазни параметри: Бајесова мрежа за учење структуре BL која садржи n променљивих; горња граница u за број родитеља који чвор може имати; подаци d.
Излазни параметри: n скупова родитеља ROD_i, где је $1 \leq i \leq n$, у ДАГ-у који приближно максимизује $score(\mathbb{G} : D)$.

```
K2 <- function(Bayes_net_struct_learn BL, int u,
data d, for 1 ≤ i ≤ n parent_set ROD_i){
     for(i = 1; i ≤ n; i++){ #n je broj cvorova
       ROD_i = ∅;
       P_old = score(X_i, ROD_i, D);
       findmore = true;
       while(findmore && |ROD_i| < u){ #| | vraca velicinu skupa
            Z = cvor iz Pred(X_i) - ROD_i koji maksimizira
            score(X_i, ROD_i ∪ {Z}, D);
            P_new = score(X_i, ROD_i ∪ {Z}, D);
            if(P_new > P_old){
              P_old = P_new;
              ROD_i = ROD_i ∪ {Z};
             }
             else
               findmore = false;
        }
     }
}
```

Глава 5

Примена Бајесових мрежа користећи *R*

У овом поглављу ћемо представити основне идеје Бајесових мрежа и њихово тумачење приликом коришћења хипотетичког истраживања о утицају неких фактора на бронхијалну астму. База података над којом ћемо вршити истраживање се зове *data_astma*[1]. Неуронаука о астми указује на то како је мозак укључен у постојање астме код неке особе и како мозак контролише наше дисање (нпр. пречник дисајних путева). Мозак је такође укључен у начин на који свако од нас доживљава дисање. Верује се да је центар интеракције између ових сензорских и когнитивних механизама мали део мозга познат као *периакведуктална сива маса*. Бољим разумевањем како је мозак укључен у астму, могли би се осмислити нови третмани за астму. Ови третмани би посебно могли помоћи људима чија астма није у потпуности контролисана тренутно доступним лековима. Више о овој теми може се наћи на сајту [11].

База података *data_astma* садржи следеће податке за 2755 испитаника: `sex`, `age` (променљива типа *string* која има 3 могуће вредности: `adult, old, young`), `urbanization` (са вредностима `low, high, medium`), `education` (`low, high`), `geographic_area (south/islands, centre, north)`, `allergy, smoke, sedentary`, `asthma (yes, no` вредности) и изгледа на следећи начин:

```
> head(data_astma)
     SEX   AGE    URB  EDU            GEO ALG SMK SED ASTHMA
1   male adult    low  low south/islands yes yes yes    yes
```

[1]База *data_astma* је преузета са сајта Kaggle, [16]

```
2 female   old    low  low south/islands yes  no yes    yes
3 female adult   high high        centre  no  no yes    yes
4   male adult medium  low south/islands yes  no  no     no
5 female adult    low high         north  no  no  no     no
6 female adult medium high         north  no yes  no    yes
```

5.1 Графички приказ

Да бисмо креирали и манипулисали ДАГ-овима у контексту Бајесових мрежа, углавном ћемо користити пакет `bnlearn` (енг. **B**ayesian **n**etwork **learn**ing). Као први корак, креирамо ДАГ са једним чвором за сваку променљиву и без усмерених стрелица. Такав ДАГ се обично назива *празан граф*, јер има празан скуп усмерених грана. ДАГ је сачуван у објекту класе `bn`, па се приликом позива функције `empty.graph` добија генерисана Бајесова мрежа:

```
> dag <- empty.graph(nodes = c("Sex", "Age", "Urb", "Edu", "Geo",
"Alg", "Smk","Sed", "Asthma"))
> dag

  Random/Generated Bayesian network

  model:
   [Sex][Age][Urb][Edu][Geo][Alg][Smk][Sed][Asthma]
  nodes:                                 9
  arcs:                                  0
    undirected arcs:                     0
    directed arcs:                       0
  average markov blanket size:           0.00
  average neighbourhood size:            0.00
  average branching factor:              0.00

  generation algorithm:                  Empty
```

Сада додајемо стрелице које представљају директне зависности између променљивих помоћу функције `set.arc`. На старост, пол и географско подручје не утиче ниједна од других променљивих, док са друге стране, старост, пол и географско подручје имају директан утицај на образовање. На пример, познато је да се број људи који похађају универзитете повећао током година.

Као последица тога, већа је вероватноћа да млађи људи имају универзитетску диплому него старији. Слично, пол такође утиче на образовање. Родни јаз у пријавама на универзитете се годинама повећава, при чему жене надмашују и бројем и успехом мушкарце.

```
> dag<-set.arc(dag, from = "Sex", to = "Edu")
> dag<-set.arc(dag, from = "Age", to = "Edu")
> dag<-set.arc(dag, from = "Geo", to = "Edu")
> dag<-set.arc(dag, from = "Edu", to = "Urb")
> dag<-set.arc(dag, from = "Urb", to = "Alg")
> dag<-set.arc(dag, from = "Urb", to = "Sed")
> dag<-set.arc(dag, from = "Alg", to = "Asthma")
> dag<-set.arc(dag, from = "Sed", to = "Asthma")
> dag<-set.arc(dag, from = "Smk", to = "Asthma")
> dag

  Random/Generated Bayesian network

  model:
   [Sex][Age][Geo][Smk][Edu|Sex:Age:Geo][Urb|Edu][Alg|Urb][Sed|Urb]
   [Asthma|Alg:Smk:Sed]
  nodes:                                 9
  arcs:                                  9
    undirected arcs:                     0
    directed arcs:                       9
  average markov blanket size:           3.33
  average neighbourhood size:            2.00
  average branching factor:              1.00

  generation algorithm:                  Empty
```

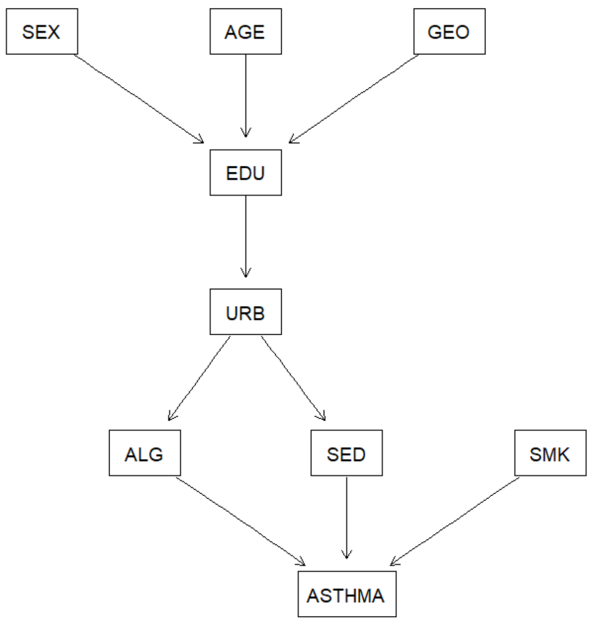
Овај приступ гарантује да ће ДАГ заиста бити ацикличан, па ће покушај увођења циклуса у ДАГ вратити грешку.

```
> dag<-set.arc(dag, from = "SED", to = "EDU")
Error in arc.operations(x = x, from = from, to = to, op = "set",
check.cycles = check.cycles,  :the resulting graph contains cycles.
```

Директне зависности су наведене за сваку промењиву, означене вертикалном цртом (|) и са две тачке (:). На пример, $[Edu|Sex : Age : Geo]$ значи да

Sex → *Edu*, *Age* → *Edu* и *Geo* → *Edu*, док [*Age*] значи да нема стрелица које показују према *Age*. Ова репрезентација структуре графа је осмишљена да подсећа на производ условних вероватноћа, и може се произвести помоћу функције `modelstring`. Графичка репрезентација графа се добија помоћу функције `graphviz.plot`.

```
> modelstring(dag)
[1] "[AGE][GEO][SEX][SMK][EDU|AGE:GEO:SEX][URB|EDU][ALG|URB]
[SED|URB][ASTHMA|ALG:SED:SMK]"
> dag<-model2network("[SEX][AGE][GEO][SMK][EDU|SEX:AGE:GEO][URB|EDU]
[ALG|URB][SED|URB][ASTHMA|ALG:SMK:SED]")
> graphviz.plot(dag)
```



Слика 5.1: Бајесова мрежа креирана на основу базе података *data_astma* у *R*-у.

5.2 Вероватносни приказ

Комбинујемо ДАГ који смо сачували у промењивој *dag* и листу која садржи локалне расподеле, што су уствари табеле условних вероватноћа, коју ћемо назвати *cpt*, и ту комбинацију смештамо у објекат класе `bn.fit` под називом bn, при чему тај објекат заправо представља параметре Бајесове мреже (нећемо их све приказати).

```
> cpt<-list(SEX=bn.bayes$SEX$prob,AGE=bn.bayes$AGE$prob,
            URB=bn.bayes$URB$prob,EDU=bn.bayes$EDU$prob,
            GEO=bn.bayes$GEO$prob,ALG=bn.bayes$ALG$prob,
            SMK=bn.bayes$SMK$prob,SED=bn.bayes$SED$prob,
            ASTHMA=bn.bayes$ASTHMA$prob)
> bn<-custom.fit(dag,cpt)
> bn

  Bayesian network parameters

  Parameters of node SEX (multinomial distribution)

Conditional probability table:
   female      male
0.5331162 0.4668838

  Parameters of node AGE (multinomial distribution)

Conditional probability table:
    adult       old     young
0.5083846 0.3770057 0.1146097

  Parameters of node URB (multinomial distribution)

Conditional probability table:

        EDU
URB           high       low
  high   0.3725026 0.2664437
  low    0.2179285 0.3286511
  medium 0.4095689 0.4049052
```

Број параметара Бајесове мреже може се израчунати помоћу функције `nparams` и заиста износи 42, што се и очекивало на основу дефиниције 11.

```
> nparams(bn)
[1] 42
```

У оквиру пакета *bnlearn* који се користи за рад са Бајесовим мрежама налази се, већ поменута, класа објеката *bn.fit*. Она служи за представљање и описивање Бајесових мрежа (као што је представљање родитеља и деце сваког чвора, као и локалних расподела, тј. њихових параметара). Дакле, неке од функција су `nodes`, `arcs`, `parents`, `children`, итд.

```
> nodes(bn)
[1] "SEX"    "AGE"    "URB"    "EDU"    "GEO"    "ALG"    "SMK"
"SED"    "ASTHMA"
> arcs(bn)
      from  to
 [1,] "SEX" "EDU"
 [2,] "AGE" "EDU"
 [3,] "URB" "ALG"
 [4,] "URB" "SED"
 [5,] "EDU" "URB"
 [6,] "GEO" "EDU"
 [7,] "ALG" "ASTHMA"
 [8,] "SMK" "ASTHMA"
 [9,] "SED" "ASTHMA"
```

Можемо испитати да ли су два чвора у објекту *bn* d-сепарабилни користећи функцију `dsep`. Она узима три аргумента x, y и z који одговарају чворовима X, Y и Z при чему прва два аргумента морају бити имена два чвора која се тестирају за d-сепарацију, док је последњи опциони скуп који је услов за d-сепарацију. На пример, ако посматрамо редну везу $AGE \to EDU \to URB$, AGE и URB су д-сепарабилни.

```
> dsep(dag,"AGE","URB","EDU")
[1] TRUE
```

Дакле, AGE и URB су условно независни.
Ако посматрамо дивергентну везу $ALG \leftarrow URB \to SED$, ALG и SED су d-сепарабилни.

```
> dsep(dag,"ALG","SED","URB")
[1] TRUE
```

Такође, ALG и SED су условно независни.
Ако посматрамо конвергентну везу $ALG \to ASTHMA \leftarrow SMK$ и немамо евиденцију о чвору $ASTHMA$, онда су ALG и SMK d-сепарабилни.

```
> dsep(dag,"ALG","SMK")
[1] TRUE
```

Међутим, уколико имамо евиденцију о чвору $ASTHMA$, онда ALG и SMK нису d-сепарабилни, тј. условно зависни су.

```
> dsep(dag,"ALG","SMK","ASTHMA")
[1] FALSE
```

5.3 Оцене мреже

Оцене мреже су статистике које мере колико добро ДАГ одражава зависну структуру података. Најчешће се користи *Бајесов информативни критеријум (BIC)*, који за нашу Бајесову мрежу има облик

$$\begin{aligned} BIC = &\log_2 \hat{P}(SEX, AGE, GEO, EDU, URB, ALG, SED, SMK, ASTHMA) \\ &-\frac{d}{2}\log_2 n = (\log_2 \hat{P}(GEO) - \frac{d_{GEO}}{2}\log_2 n) + (\log_2 \hat{P}(AGE) - \frac{d_{AGE}}{2}\log_2 n) \\ &+ \cdots + (\log_2 \hat{P}(ASTHMA|ALG, SED, SMK) - \frac{d_{ASTHMA}}{2}\log_2 n) \end{aligned}$$

где је n величина узорка, d је број параметара целе мреже (тј. 42), а $d_{GEO}, d_{AGE}, ..., d_{ASTHMA}$ су бројеви параметара сваког чвора.

Наредна оцена која се такође у пракси користи је AIC оцена, доста слична BIC оцени, која у нашем случају има облик

$$\begin{aligned} AIC = &\log_2 \hat{P}(SEX, AGE, GEO, EDU, URB, ALG, SED, SMK, ASTHMA) \\ &-d. \end{aligned}$$

Још једна оцена која се често користи је *BDeu (Bayesian Dirichlet equivalent uniform)*, чији је општи облик дат у теореми 4.2.3, која је повезана са априорном униформним расподелом над простором ДАГ-а и параметара. То значи да се параметри N_{ij} и a_{ijk} постављају на униформну вредност (нпр. 1) за све могуће вредности стања k. Ова униформност указује на то да се сви могући параметри третирају као једнако вероватни пре него што се подаци употребе за њихову процену. Дакле, овим добијамо неутралност и једноставност. Често се ова оцена једноставно означава као *BDe*. И *BIC* и *BDe* додељују више оцене моделима (ДАГ-овима) који боље одговарају подацима. У суштини, ако ДАГ боље уклапа податке, његова оцена ће имати већу вредност.

Све три оцене могу се израчунати у `bnlearn` пакету користећи функцију *score*. *BIC* се рачуна када је *type* = "*bic*", *AIC* када је *type* = "*aic*", а *BDe* када је *type* = "*bde*".

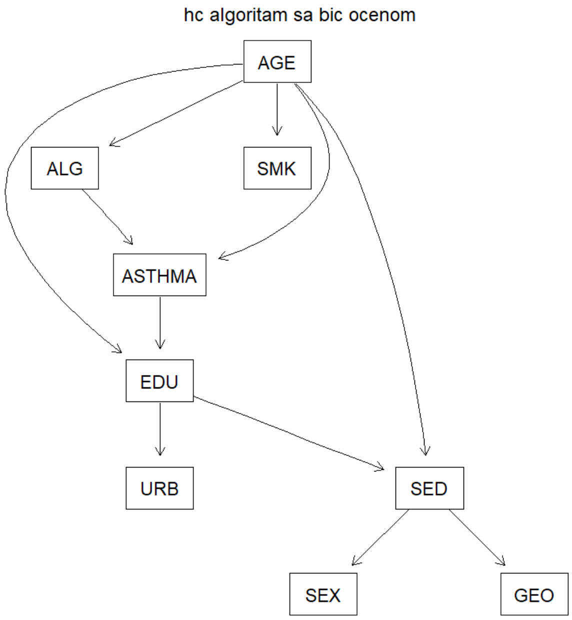
```
> score(dag,data=data_astma,type="bic")
[1] -18230.42
> score(dag,data=data_astma,type="aic")
[1] -18106.08
> score(dag,data=data_astma,type="bde",iss=8)
[1] -18197.26
```

Опциони аргумент `iss`, односно *имагинарна величина узорка* (енг. *imaginary sample size*), одређује колико априорна расподела утиче на коначне процене у Бајесовој мрежи. Већа вредност `iss` значи да априорна расподела има већу тежину, док мања вредност значи да се више ослањамо на посматране податке, тј. нове информације или опсервације. За мале вредности `iss` или велике посматране узорке, *BDe* и *BIC* оцене дају сличне вредности.

За одабир мрежне структуре, постоји неколико агоритама који траже ДАГ који максимизира дату оцену мреже. Један једноставан, који смо помињали у поглављу 4.2.2.1, је *hill-climbing*. Почевши од ДАГ-а без грана, алгоритам додаје, уклања или обрће једну грану и бира промену која највише повећава оцену мреже. То је имплементирано у `hc` функцији, која у свом најједноставнијем облику узима податке (*data_astma*) као једини аргумент и подразумевано користи *BIC* оцену.

```
> learned<-hc(data_astma,score="bic")
> graphviz.plot(learned,main="hc␣algoritam␣sa␣bic␣ocenom")
> modelstring(learned)
[1] "[AGE][ALG|AGE][SMK|AGE][ASTHMA|AGE:ALG][EDU|AGE:ASTHMA]
[URB|EDU][SED|AGE:EDU][SEX|SED][GEO|SED]"
> learned

  Bayesian network learned via Score-based methods

  model:
   [AGE][ALG|AGE][SMK|AGE][ASTHMA|AGE:ALG][EDU|AGE:ASTHMA]
   [URB|EDU][SED|AGE:EDU][SEX|SED][GEO|SED]
  nodes:                                 9
  arcs:                                  11
```

```
    undirected arcs:                       0
    directed arcs:                         11
  average markov blanket size:             2.44
  average neighbourhood size:              2.44
  average branching factor:                1.22

  learning algorithm:                      Hill-Climbing
  score:                                   BIC (disc.)
  penalization coefficient:                3.960586
  tests used in the learning procedure:    124
  optimized:                               TRUE

> score(learned,data=data_astma,type="bic")
[1] -17939.44
> nodes(learned)
[1] "SEX"    "AGE"    "URB"    "EDU"    "GEO"    "ALG"    "SMK"
"SED"    "ASTHMA"
> arcs(learned)
      from     to
 [1,] "ALG"    "ASTHMA"
 [2,] "AGE"    "EDU"
 [3,] "AGE"    "SED"
 [4,] "AGE"    "ASTHMA"
 [5,] "SED"    "GEO"
 [6,] "EDU"    "SED"
 [7,] "EDU"    "URB"
 [8,] "AGE"    "SMK"
 [9,] "AGE"    "ALG"
[10,] "SED"    "SEX"
[11,] "ASTHMA" "EDU"
```



Слика 5.2: Структура Бајесове мреже научене помоћу алгоритма заснованог на оценама (hill-climbing).

На сличан начин можемо видети како изгледа графичка репрезентација модела, као и њене карактеристике када се користи `hc` алгоритам у комбинацији са `bde` или `aic` оценом.

Ради поређења оцена, из приложеног можемо видети да је бољи графички модел *dag* него *learned* који је добијен `hc` алгоритмом, с обзиром да нижу вредност оцена има први модел.

```
> #standard system
> a<-score(dag,data=data_astma,type="bic")
> #hc algorithm
> b<-score(learned,data=data_astma,type="bic")
> bic<-c(a,b)
> names(bic)<-c("standard␣system","hc␣algorithm")
> bic
standard system    hc algorithm
      -18230.42       -17939.44
```

Тачне процедуре закључивања

У претходном одељку смо видели како можемо одговорити на упите о условној независности, помоћу функције `dsep`, ослањајући се искључиво на информацијама са графичког приказа мреже. Међутим, сложенији упити захтевају коришћење локалних расподела. ДАГ се и даље користи индиректно јер одређује састав локалних расподела и зато што може смањити сложеност проблема закључивања у Бајесовим мрежама. На пример, ако у мрежи постоје променљиве које су условно независне од других под одређеним условима, можемо игнорисати непотребне променљиве у том контексту, што смањује број операција које су потребне за израчунавање тачних вероватноћа. То чини проблем закључивања математички и рачунарски једноставнијим и ефикаснијим за решавање.

Две најчешће врсте закључивања су *упити о условним вероватноћама* (енг. *conditional probability queries*), који истражују расподелу једне или више променљивих под нетривијалним условљавањем, и *упити о највероватнијем објашњењу* (енг. *most likely explanation queries*), који траже највероватнији исход једне или више променљивих. У оба контекста, променљиве које су условљене представљају нове *доказе* који приморавају да се вероватноћа догађаја од интереса поново процени. Ови упити могу бити решени коришћењем *тачног закључивања*.

Тачно закључивање, које је имплементирано у пакету `gRain` (***gRa***phical model ***in***ference), ослања се на трансформисање Бајесове мреже у посебно креирано стабло како би се убрзало израчунавање условних вероватноћа. Такво стабло се назива *заједничко стабло* (енг. *junction tree*) и може се конструисати из објекта Бајесове мреже *bn* на следећи начин:

```
> junction<-compile(as.grain(bn))
```

Када је заједничко стабло направљено (помоћу `as.grain`) и његове табеле вероватноћа израчунате (помоћу `compile`), можемо унети доказе[2] у *junction* користећи функцију `setEvidence`. Локалне расподеле чворова на које се односе докази се тада ажурирају, а промене се пропагирају кроз заједничко стабло. Стварни упит се извршава функцијом `querygrain` која извлачи расподелу чворова од интереса из *junction*.

[2]На пример, у контексту медицинске дијагнозе, ако пацијент има одређени симптом, као што је грозница, тај симптом би био „доказ”. Овај доказ се користи да би се ажурирале вероватноће различитих болести које би могле бити узрок тог симптома.

Аргумент `type` у функцији `querygrain` одређује коју од могућих расподела које укључују чворове `nodes` треба вратити. Подразумевана вредност је „`marginal`”, за маргиналну расподелу сваког чвора.

```
> querygrain(junction,nodes="ASTHMA",type="marginal")
```

Друга могућа опција је „`joint`”, за заједничку расподелу чворова. Ако посматрамо однос између $ASTHMA$ и ALG, њихова заједничка расподела вероватноћа $P(ASTHMA, ALG)$ може се израчунати користећи `querygrain` на сличан начин:

```
> querygrain(junction,nodes=c("ASTHMA","ALG"),type="joint")
     ASTHMA
ALG            no       yes
  no   0.50141386 0.2217066
  yes  0.06971911 0.2071605
```

Последња валидна опција је „`conditional`”. У том случају, функција `querygrain` враћа расподелу првог чвора из `nodes` условљену другим чворовима из `nodes`. Условна расподела чвора $ASTHMA$ у односу на чворове ALG и SMK је

```
> querygrain(junction,nodes=c("ASTHMA","ALG","SED"),type=
             "conditional")
, , SED = no

      ALG
ASTHMA         no       yes
   no   0.7535039 0.2755079
   yes  0.2464961 0.7244921

, , SED = yes

      ALG
ASTHMA         no       yes
   no   0.6751085 0.2445218
   yes  0.3248915 0.7554782
```

Постављамо питање: да ли пушачи и особе које пате од алергија чешће имају астму?

```
> j1<-setEvidence(junction,nodes=c("ALG","SMK"),states=c("yes",
                  "yes"))
```

```
> j1$isPropagated
[1] TRUE
```

Када се Бајесова мрежа трансформише у заједничко стабло, потребно је пропагирати информације кроз то стабло да би се ажурирале вероватноће свих чворова. `isPropagated` функција омогућава да проверимо да ли су те информације успешно пропагиране кроз стабло, што је неопходно пре него што се изврши било какав упит или закључивање на моделу.

```
> querygrain(junction,nodes="ASTHMA",type="marginal")
$ASTHMA
ASTHMA
      no      yes
0.571133 0.428867

> querygrain(j1,nodes="ASTHMA",type="marginal")
$ASTHMA
ASTHMA
       no       yes
0.2435954 0.7564046
```

Приметимо да су се вероватноће значајно промениле. Слично је и код условних расподела:

```
> querygrain(junction,nodes=c("ASTHMA","SED"),type="conditional")
      SED
ASTHMA        no       yes
   no   0.6204933 0.5560702
   yes 0.3795067 0.4439298
> querygrain(j1,nodes=c("ASTHMA","SED"),type="conditional")
      SED
ASTHMA       no      yes
   no   0.171875 0.265625
   yes 0.828125 0.734375
```

Закључак за ово испитивање коришћењем Бајесових мрежа је да се вероватноће за појаву астме мењају када се постави доказ да је особа и пушач и има алергију. Након постављања овог доказа, маргиналне и условне вероватноће за развој астме се повећавају, што указује на то да су пушачи и особе са алергијом заиста под већим ризиком да имају астму. Функција querygrain показује да је вероватноћа за астму већа када су присутни и пушење и алергија.

Дакле, `bnlearn` је дизајниран да пружи свестраност потребну за анализу експерименталних података. Рукује како дискретним, тако и непрекидним подацима, и подржава било коју комбинацију имплементираних алгоритама за учење, као и мрежне оцене (за алгоритме засноване на оценама) или тестове условне независности (за алгоритме засноване на ограничењима). Штавише, поједностављује анализу научених мрежа пружајући једну класу објеката (`bn`) за све алгоритме и скуп помоћних функција за извођење дескриптивних статистика и основних процедура закључивања.[13]

Кључна снага Бајесових мрежа (постојећих или направљених), као и графичких модела уопште, јесте могућност њиховог проучавања кроз графичке приказе. Стога је могућност ефикасног приказивања Бајесове мреже кључни алат у закључивању у Бајесовим мрежама.

Да закључимо, након што се Бајесова мрежа направи, постоји неколико корака и активности које се обично предузимају да би се искористила за различите анализе и одлуке:

- Постављање вероватноћа: Ово укључује дефинисање условних вероватноћа за сваки чвор у мрежи у односу на његове родитеље.

- Проучавање структуре: Истражују се односи између променљивих у мрежи и анализирају се условне зависности и независности које мрежа приказује.

- Закључивање: Користе се методе за извођење закључивања како би се израчунале вероватноће непознатих променљивих на основу познатих информација. Ово може укључивати упите о условним вероватноћама или упите о највероватнијим објашњењима. Такође, може се користити тачно закључивање (као што су елиминација променљивих и алгоритам за заједничко стабло).

- Ажурирање мреже новим подацима (докази): Када се добију нови подаци, мрежа се ажурира како би рефлектовала ове нове информације. Ово се може урадити уношењем нових доказа и прилагођавањем вероватноћа у мрежи.

- Евалуација перформанси: Подразумева евалуацију мреже у смислу тачности њених предикција и одлука. Ово може укључивати упоређивање са стварним подацима или симулацијом различитих сценарија.

- Одлуке и предикције: Мрежа се користи за доношење одлука у стварним ситуацијама, као што су нпр. медицинске дијагнозе. Такође, на основу тренутних и претходних информација, мрежа може предвидети будуће догађаје или резултате.

- Визуализација: Визуализација структуре Бајесове мреже може помоћи у бољем разумевању односа између променљивих и у комуникацији са другим корисницима или доносиоцима одлука.

Глава 6

Закључак

У овом раду смо истражили примену Бајесових мрежа у неуронаукама, приказујући њихов потенцијал као моћан алат за анализу сложених односа између променљивих у оквиру мождaних функција и структура. Бајесове мреже пружају научни метод заснован на принципима вероватноће, омогућавајући прецизно моделирање и предикцију различитих неуролошких феномена. Иако су досадашње примене у овој области биле ограничене, показано је да Бајесове мреже могу значајно допринети како у класификацији и процесу идентификације и анализе когнитивних стања, тако и у откривању сложених образаца, тј. не очигледних, вишедимензионалних веза између различитих променљивих у неуролошким подацима. На основу ових налаза, може се закључити да постоји широк простор за даљи развој и интеграцију Бајесових мрежа у неуронауци, посебно кроз прилагођавање постојећих модела и развој нових алгоритама који ће омогућити још прецизније анализе и увиде у функционалне и структуралне аспекте мозга.

У овом раду смо представили основне појмове Бајесових мрежа, као графичке структуре и у смислу вероватноћа. Затим, показали смо њихову могућност резоновања и навели које врсте Бајесових мрежа постоје. Од врста смо дефинисали Гаусове Бајесове мреже, као и динамичке Бајесове мреже. Због комплексности рачунања вероватноћа, нисмо их рачунали ручно у примерима, већ у програмском језику *R* кроз функције *custom.fit* и *querygrain*. Показали смо широку примену Бајесових мрежа, а посветили смо се и практичном примеру на конкретној бази података која испитује утицаје одређених фактора на астму.

Библиографија

[1] Bayes nets. https://www.cs.cmu.edu/ ./15281/coursenotes/bayesnets/index.html.

[2] C. Bielza and P. Larrañaga. Bayesian networks in neuroscience: a survey. *Frontiers in computational neuroscience*, 8:131, 2014.

[3] Y. He, J. Jia, and B. Yu. Counting and exploring sizes of markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 16(1):2589–2609, 2015.

[4] F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs*, volume 2. Springer, 2007.

[5] M. Kastrati and M. Biba. Statistical relational learning: A state-of-the-art review. *Journal of Engineering Technology and Applied Sciences*, 4(3):141–156, 2019.

[6] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[7] T. Koski and J. Noble. *Bayesian networks: an introduction*. John Wiley & Sons, 2011.

[8] M. Mijic. Bejzove mreze i njihova primena.

[9] R. E. Neapolitan. *Probabilistic methods for bioinformatics: with an introduction to Bayesian networks*. Morgan Kaufmann, 2009.

[10] R. E. Neapolitan et al. *Learning bayesian networks*, volume 38. Pearson Prentice Hall Upper Saddle River, 2004.

[11] U. o. O. Nuffield Department of Clinical Neurosciences. Neuroscience of asthma, 2019. https://www.ndcn.ox.ac.uk/research/breathe-oxford/research-projects/neuroscience-of-asthma.

[12] M. Rudman. *Kompleksnost skrivenih Markovljevih modela.* PhD thesis, University of Zagreb. Faculty of Science. Department of Mathematics, 2014.

[13] M. Scutari. Learning bayesian networks with the bnlearn r package. *arXiv preprint arXiv:0908.3817*, 2009.

[14] M. Scutari and J.-B. Denis. *Bayesian networks: with examples in R.* Chapman and Hall/CRC, 2021.

[15] T. J. Sejnowski, C. Koch, and P. S. Churchland. Computational neuroscience. *Science*, 241(4871):1299–1306, 1988.

[16] M. Tucci. Bayesian network dataset, 2021. https://www.kaggle.com/datasets/marcotucci/bayesian-network-dataset/data.

Биографија аутора

Катарина Јелесијевић је рођена 14. јануара 1994. године у Горњем Милановцу. Завршила је основну школу „Момчило Настасијевић” у Горњем Милановцу, а након тога гимназију „Таковски устанак” (друштвено-језички смер). 2013. године уписује Математички факултет Универзитета у Београду, на смеру Математика и модулу Статистика, актуарска и финансијска математика. Основне студије завршава у септембру 2020. године. Након тога 2021. уписује мастер студије на истом модулу.

Каријеру започиње у мају 2020. године у ОТП банци Србије. Затим од јануара 2023. године започиње рад у ИТ индустрији као *Junior Software Developer* у фирми *SEE Digital* где и данас ради.