Міністерство освіти і науки України

Дніпровський національний університет імені Олеся Гончара

Я. С. Бондаренко, Д. О. Рачко

ПОСІБНИК ДО ВИВЧЕННЯ ДИСЦИПЛІНИ “ІМОВІРНІСНІ ГРАФІЧНІ МОДЕЛІ” ЧАСТИНА 3. ТОЧНИЙ ІМОВІРНІСНИЙ ВИСНОВОК

Дніпро

2021

УДК 519.21:004.032.26
Б81

Рецензенти: канд. фіз.-мат. наук, доц. М.Є. Ткаченко,
канд. фіз.-мат. наук, доц. А.М. Пасько

Б81 Бондаренко Я. С. Посібник до вивчення дисципліни “Імовірнісні графічні моделі”. Частина 3. Точний імовірнісний висновок [Текст] / Я.С. Бондаренко, Д.О. Рачко. – Дніпро: Ліра, 2021. – 36 с.

Викладено теоретичні положення щодо формування точного ймовірнісного висновку на основі байєсівської мережі.

Для студентів механіко-математичного факультету ДНУ спеціальності “Статистика”.

Робота нагороджена дипломом II ступеня
на Всеукраїнському конкурсі студентських наукових робіт
зі спеціальності «Комп’ютерні науки» у 2020/2021 н.р.

Рекомендовано до друку вченою радою
механіко-математичного факультету
Дніпровського національного університету імені Олеся Гончара
протокол №9 від 20.04.2021 року

Навчальне видання

Яна Сергіївна Бондаренко
Деніс Олексійович Рачко

Посібник до вивчення дисципліни
“Імовірнісні графічні моделі”
Частина 3. Точний імовірнісний висновок

Друкується за авторською редакцією

Підписано до друку 26.04.2021. Формат 60×84/16. Папір друкарський. Друк плоский. Ум. друк. арк. 2,25. Тираж 20 пр. Зам. № 98.

Друкарня «Ліра», вул. Наукова, 5, м. Дніпро, 49107. Свідоцтво про внесення до Державного реєстру серія ДК №6042 від 26.02.2018 р.

© Бондаренко Я.С., Рачко Д.О., 2021

ВСТУП

Використання байєсівської мережі в системах підтримки прийняття рішень дозволяє враховувати та використовувати різні дані: експертні оцінки, апріорну інформацію, а також статистичні дані, які надходять в режимі реального часу. Структура байєсівської мережі візуалізується за допомогою зображення взаємодії між змінними процесу у вигляді причино-наслідкових зв'язків, і це істотно виділяє байєсівські мережі серед інших методів машинного навчання, в яких застосовуються нейронні мережі.

В першому розділі навчального посібника розглянуто *подання байєсівської мережі*, наведена структура та експертні оцінки параметрів умовних імовірнісних розподілів вершин мережі Insurance (додаток А).

В другому розділі наведено метод максимальної правдоподібності та байєсівський метод статистичного оцінювання для *навчання байєсівської мережі*: змодельовано навчальні вибірки даних з експертними оцінками параметрів умовних імовірнісних розподілів вершин мережі Insurance; застосовано метод максимальної правдоподібності та байєсівський метод для знаходження невідомих оцінок параметрів (додаток А); обчислено дивергенцію Кульбака-Лейблера між експертними та емпіричними умовними ймовірнісними розподілами вершин мережі; порівняно експертні оцінки, оцінки максимальної правдоподібності та байєсівські оцінки параметрів умовних імовірнісних розподілів вершин мережі.

В третьому розділі вивчено клас алгоритмів точного імовірнісного висновку на основі байєсівської мережі, заснованих на виключенні змінних, зокрема, проведено *точний імовірнісний висновок* за допомогою класичного алгоритму виключення змінних, алгоритму виключення змінних для оцінювання максимуму апостеріорної ймовірності, алгоритму виключення змінних для оцінювання максимуму маргінальної апостеріорної ймовірності. Моделі міркувань, побудовані за допомогою алгоритмів точного імовірнісного висновку, представлено графічно та описово (додаток Б).

1. Подання байєсівської мережі

1.1. Формула множення для байєсівської мережі

Означення. *Байєсівською мережею* називатимемо пару $\mathcal{B} = (G, P)$, де G – орієнтований ациклічний граф, P – спільний розподіл, що факторизується на графі G.

Означення. Граф G *задає множину умовних незалежностей, пов'язаних зі спільним розподілом* P, якщо кожна вершина X_i умовно не залежить від вершин $NonDescendants_{X_i}$, які не є її нащадками при заданих батьківських вершинах $Parents_{X_i}$ для вершини X_i:

$$P(X_i|X_1, X_2, \dots, X_{i-1}) = P\left(X_i\middle|Parents_{X_i}\right), \qquad i = 1, \dots, n,$$

де вершини $X_1, \dots, X_n$ розташовані в топологічному порядку відносно графа G.

Формула множення для байєсівської мережі. Нехай G – байєсівська мережа, побудована на $X_1, \dots, X_n$. Спільний розподіл $P(X_1, \dots, X_n)$ факторизується на графі G, якщо його можна подати як добуток умовних імовірнісних розподілів $P\left(X_i|Parents_{X_i}\right)$:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i|Parents_{X_i}\right).$$

Теорема [1]. Нехай G – байєсівська мережа, побудована на $X_1, \dots, X_n$, P – спільний розподіл випадкових величин $X_1, \dots, X_n$. Якщо граф G задає множину умовних незалежностей, пов'язаних зі спільним розподілом P, то P факторизується на графі G.

Доведення. Згідно з формулою множення спільний розподіл $X_1, \dots, X_n$ дорівнює

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$$
$$\dots P(X_i|X_1, \dots, X_{i-1}) \dots\ P(X_n|X_1, \dots, X_{n-1}).$$

Розглянемо один множник $P(X_i|X_1,\ldots,X_{i-1})$. Оскільки граф G задає множину умовних незалежностей, пов'язаних зі спільним розподілом P, то для кожної вершини X_i:

$$\left(X_i \perp NonDescendants_{X_i}|Parents_{X_i}\right), i = 1,\ldots,n.$$

Всі батьківські вершини для X_i містяться в множині $\{X_1,\ldots,X_{i-1}\}$, а жоден з нащадків вершини X_i не належить множині $\{X_1,\ldots,X_{i-1}\}$. Отже,

$$\{X_1,\ldots,X_{i-1}\} = Parents_{X_i} \cup \boldsymbol{Z},$$

де $\boldsymbol{Z} \subseteq NonDescendants_{X_i}$. Тоді

$$\left(X_i \perp \boldsymbol{Z}|Parents_{X_i}\right).$$

Отже,

$$P(X_i|X_1,\ldots,X_{i-1}) = P\left(X_i\middle|Parents_{X_i} \cup \boldsymbol{Z}\right) = P\left(X_i\middle|Parents_{X_i}\right).$$

Тоді спільний розподіл дорівнює

$$P(X_1,X_2,\ldots,X_n) = P\left(X_1\middle|Parents_{X_1}\right)P\left(X_2\middle|Parents_{X_2}\right)\ldots P\left(X_n\middle|Parents_{X_n}\right),$$

тобто P факторизується на графі G. Теорема доведена.

Теорема [1]. Нехай G – байєсівська мережа, побудована на $X_1,\ldots,X_n$, P – спільний розподіл випадкових величин $X_1,\ldots,X_n$. Якщо P факторизується на графі G, то граф G задає множину умовних незалежностей, пов'язаних зі спільним розподілом P.

Доведення. Спільний розподіл P факторизується на графі G, тобто справедлива формула множення для байєсівської мережі

$$P(X_1,X_2,\ldots,X_n) = P\left(X_1\middle|Parents_{X_1}\right)P\left(X_2\middle|Parents_{X_2}\right)\ldots P\left(X_n\middle|Parents_{X_n}\right).$$

Покажемо, що $\left(X_i \perp \boldsymbol{Z}\middle|Parents_{X_i}\right), \boldsymbol{Z} \subseteq NonDescendants_{X_i}$, зокрема,

$$P(X_i|X_1,X_2,\ldots,X_{i-1}) = P\left(X_i\middle|Parents_{X_i}\right).$$

Дійсно,

$$P(X_i|X_1,X_2,\ldots,X_{i-1}) = \frac{P(X_1,X_2,\ldots,X_i)}{P(X_1,X_2,\ldots,X_{i-1})} =$$

$$= \frac{P\left(X_1\middle|Parents_{X_1}\right)\ldots P\left(X_i\middle|Parents_{X_i}\right)}{P\left(X_1\middle|Parents_{X_1}\right)\ldots P\left(X_{i-1}\middle|Parents_{X_{i-1}}\right)} = P\left(X_i\middle|Parents_{X_i}\right).$$

Теорема доведена.

1.2. Байєсівська мережа Insurance

Байєсівська мережа Insurance представлена орієнтованим ациклічним графом, в вершинах якого знаходяться характеристики клієнтів та їхніх автомобілів, а орієнтовані ребра представляють вплив однієї характеристики на іншу. Структура мережі та експертні оцінки параметрів умовних імовірнісних розподілів вершин [2] подані на рис. А1 (додаток А). Так концентрація уваги клієнта $Focused$ та його вік Age впливають на успішність клієнта в автошколі $Good\ Student$; концентрація уваги клієнта $Focused$, його вік Age та проходження курсу екстремального водіння $Extra\ Training$ впливають на кваліфікованість водія $Driver\ Quality$; кваліфікованість водія $Driver\ Quality$ впливає на кількість аварій $Driving\ History$; концентрація уваги клієнта $Focused$ впливає на розмір автомобіля $Vehicle\ Size$; кваліфікованість водія $Driver\ Quality$, розмір автомобіля $Vehicle\ Size$, рік випуску автомобіля $Vehicle\ Year$, наявність подушки безпеки в автомобілі $Airbag$ та наявність антиблокувальної системи в автомобілі $Antilock\ Brakes$ впливає на ступінь тяжкості аварії $Accident$; рік випуску автомобіля $Vehicle\ Year$ та ступінь тяжкості аварії $Accident$ впливають на витрати страхової компанії $Cost$.

Спільний розподіл вершин $X_1, X_2, \dots, X_{12}$ задається за допомогою формули множення й з урахуванням множини умовних незалежностей перепишеться у вигляді:

$$P(X_1, X_2, \dots, X_{12}) = P(X_{12}|X_{11}, X_8)P(X_{11}|X_{10}, X_9, X_8, X_7, X_5)P(X_{10}|X_8) \cdot$$
$$\cdot P(X_9|X_8)P(X_8)P(X_7|X_1)P(X_6|X_5)P(X_5|X_1, X_2, X_4)P(X_4) \cdot$$
$$\cdot P(X_3|X_2, X_1)P(X_2)P(X_1).$$

Перевага подання спільного розподілу $P(X_1, X_2, \dots, X_{12})$ як добутку умовних імовірнісних розподілів $P\left(X_i|Parents_{X_i}\right)$ полягає в тому, що нам необхідно знати в $(2^{10}3^2 - 1)/2^5 3$ разів менше оцінок імовірностей можливих значень станів вершин, ніж при безпосередньому заданні спільного розподілу.

2. Навчання байєсівської мережі

2.1. Метод максимальної правдоподібності для байєсівської мережі

Розглянемо байєсівську мережу $B = (G, P)$ з відомою структурою G та невідомими параметрами $\boldsymbol{\theta}$ умовних імовірнісних розподілів вершин мережі. Необхідно оцінити невідомі параметри згідно з методом максимальної правдоподібності за вибіркою $\xi[1], \dots, \xi[M]$, де $\xi \in \mathbb{R}^n$. Функція максимальної правдоподібності як функція від параметрів $\boldsymbol{\theta}$ запишеться:

$$L(\boldsymbol{\theta}, D) = P(\,\xi[1], \dots, \xi[M]; \boldsymbol{\theta}) = \prod_{m=1}^{M} P(\xi[m]; \boldsymbol{\theta}).$$

Згідно з формулою множення для байєсівської мережі спільний розподіл вершин $X_1, \dots, X_n$ можна подати як добуток умовних імовірнісних розподілів вершин мережі:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| Parents_{X_i}\right).$$

Тоді функція максимальної правдоподібності перепишеться у вигляді добутку локальних функцій максимальної правдоподібності :

$$L(\boldsymbol{\theta}, D) = \prod_{m=1}^{M} \prod_{i=1}^{n} P\left(x_i[m] \middle| Pa_{X_i}[m]; \boldsymbol{\theta}\right) =$$

$$= \prod_{i=1}^{n} \prod_{m=1}^{M} P\left(x_i[m] \middle| Pa_{X_i}[m]; \boldsymbol{\theta}\right) = \prod_{i=1}^{n} L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right). \qquad (2.1.1)$$

Оскільки параметри $\boldsymbol{\theta}_{X_i|Pa_{X_i}}$ та $\boldsymbol{\theta}_{X_j|Pa_{X_j}}$, $i \neq j$, різних вершин не пов'язані між собою, то оцінки максимальної правдоподібності знаходяться для кожної локальної функції максимальної правдоподібності окремо.

Розглянемо одну локальну функцію максимальної правдоподібності

$$L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right) = P\left(x_i[m] \middle| Pa_{X_i}[m]; \boldsymbol{\theta}_{X_i|Pa_{X_i}}\right)$$

та знайдемо оцінки максимальної правдоподібності.

Локальна функція запишеться як добуток добутків умовних імовірностей по всіх різних можливих значеннях, які набувають батьківські вершини Pa_{X_i} для вершини X_i:

$$L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right) = \prod_{m:Pa_{X_i}[m]=pa^1_{X_i}} P\left(x_i[m]\middle|Pa_{X_i}[m];\boldsymbol{\theta}_{X_i|pa^1_{X_i}}\right)\cdot\ldots$$

$$\cdot\prod_{m:Pa_{X_i}[m]=pa^K_{X_i}} P\left(x_i[m]\middle|Pa_{X_i}[m];\boldsymbol{\theta}_{X_i|pa^K_{X_i}}\right)$$

Далі локальна функція перепишеться як добуток добутків умовних імовірностей по всіх комбінаціях різних можливих значень, які набувають батьківські вершини Pa_{X_i} для вершини X_i та сама вершина X_i:

$$L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right) = \prod_{m:Pa_{X_i}[m]=pa^1_{X_i},\ \ x_i[m]=x^1} P\left(x_i[m]\middle|Pa_{X_i}[m];\theta_{x^1|pa^1_{X_i}}\right)\cdot\ldots$$

$$\cdot\prod_{m:Pa_{X_i}[m]=pa^K_{X_i},\ \ x_i[m]=x^s} P\left(x_i[m]\middle|Pa_{X_i}[m];\theta_{x^s|pa^K_{X_i}}\right).$$

Підставимо умовні ймовірності в останню формулу (це і є наші невідомі параметри, для яких необхідно здобути оцінки):

$$L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right) = \left(\prod_{\substack{m:Pa_{X_i}[m]=pa^1_{X_i},\\ x_i[m]=x^1}} \theta_{x^1|pa^1_{X_i}}\cdot\ldots\cdot\prod_{\substack{m:Pa_{X_i}[m]=pa^1_{X_i},\\ x_i[m]=x^s}} \theta_{x^s|pa^1_{X_i}}\cdot\right)\cdot\ldots$$

$$\cdot\left(\prod_{\substack{m:Pa_{X_i}[m]=pa^K_{X_i},\\ x_i[m]=x^1}} \theta_{x^1|pa^K_{X_i}}\cdot\ldots\cdot\prod_{\substack{m:Pa_{X_i}[m]=pa^K_{X_i},\\ x_i[m]=x^s}} \theta_{x^s|pa^K_{X_i}}\cdot\right)$$

Підраховуємо число вибіркових значень, які задовольняють умовам по яких рахуються добутки і запишемо ці числа як ступені відповідних умовних ймовірностей:

$$L_i\left(\boldsymbol{\theta}_{X_i|Pa_{X_i}}\right) = \left(\theta_{x^1|pa^1_{X_i}}^{M\left[x^1,pa^1_{X_i}\right]}\cdot\ldots\cdot\theta_{x^s|pa^1_{X_i}}^{M\left[x^s,pa^1_{X_i}\right]}\right)\cdot\ldots$$

$$\cdot \left(\theta_{x^1|pa_{X_i}^K}^{M\left[x^1, pa_{X_i}^K\right]} \cdot \ldots \cdot \theta_{x^s|pa_{X_i}^K}^{M\left[x^s, pa_{X_i}^K\right]} \right)$$

Отже, здобули добуток функцій максимальної правдоподібності мультиноміальних розподілів. Оцінки максимальної правдоподібності мультиноміальних розподілів мають вигляд:

$$\hat{\theta}_{x_i|pa_{X_i}} = \frac{M\left[x_i,\ pa_{X_i}\right]}{M\left[pa_{X_i}\right]}, \qquad M\left[pa_{X_i}\right] = \sum_{x_i} M\left[x_i, pa_{X_i}\right]. \tag{2.1.2}$$

де $M\left[x_i,\ pa_{X_i}\right]$ – число спостережень таких, що $x_i[m] = x_i,\ Pa_{X_i}[m] = pa_{X_i}$.

Для навчання мережі нам необхідно обчислити величини $M\left[x_i,\ pa_{X_i}\right]$ для кожної комбінації станів вершини X_i та станів її батьківських вершин Pa_{X_i}, а також здобути суми $M\left[pa_{X_i}\right]$ по всіх можливих станах вершини X_i [1, 4].

2.2. Байєсівське оцінювання параметрів байєсівської мережі

Розглянемо байєсівську мережу $B = (G, P)$ з відомою структурою G та невідомим вектором параметрів $\boldsymbol{\theta}$ умовних імовірнісних розподілів вершин мережі. Необхідно оцінити невідомі параметри $\boldsymbol{\theta}$ згідно з байєсівським методом статистичного оцінювання.

Згідно з теоремою Байєса апостеріорна щільність розподілу вектору параметрів $\boldsymbol{\theta}$ за умови заданої вибірки D дорівнює

$$p\ (\boldsymbol{\theta} \mid D)\ = \frac{p\ (D\ |\boldsymbol{\theta})p(\boldsymbol{\theta})}{p\ (D)}, \tag{2.2.1}$$

де $p(D|\boldsymbol{\theta})$ – функція максимальної правдоподібності, $p(\boldsymbol{\theta})$ – апріорна щільність розподілу вектору параметрів $\boldsymbol{\theta}$, $p(D)$ – маргінальна правдоподібність.

Функція максимальної правдоподібності дорівнює спільному розподілу вибіркових значень $D\ = (\xi[1], \cdots, \xi[M])$:

$$p(D,\boldsymbol{\theta}) = P(\xi[1],\cdots,\xi[M],\boldsymbol{\theta}) = \prod_{m=1}^{M} P(\xi[m],\boldsymbol{\theta}). \qquad (2.2.2)$$

Згідно з формулою множення для байєсівської мережі спільний розподіл вершин $X_1, \dots, X_n$ можна подати як добуток умовних імовірнісних розподілів:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i|\ Parents_{X_i}\right). \qquad (2.2.3)$$

Тоді функція максимальної правдоподібності перепишеться у вигляді добутку локальних функцій максимальної правдоподібності (2.2.3):

$$p(D,\boldsymbol{\theta}) = \prod_{m=1}^{M}\prod_{i=1}^{n} P\left(x_i[m]|\ Pa_{X_i}[m];\ \boldsymbol{\theta}\right) = \prod_{i=1}^{n}\prod_{m=1}^{M} P\left(x_i[m]|\ Pa_{X_i}[m];\ \boldsymbol{\theta}\right) =$$

$$= \prod_{i=1}^{n} L_i\left(\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}\right). \qquad (2.2.4)$$

Апріорна щільність розподілу вектору параметрів $\boldsymbol{\theta}$ набуває вигляду добутку апріорних щільностей векторів параметрів $\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}$ умовних імовірнісних розподілів вершин мережі:

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{n} p\left(\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}\right). \qquad (2.2.5)$$

Отже, апостеріорна щільність розподілу вектору параметрів $\boldsymbol{\theta}$ запишеться у вигляді:

$$p(\boldsymbol{\theta}|D) = \frac{1}{p(D)}\prod_{i=1}^{n} L_i\left(\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}\right) p\left(\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}\right). \qquad (2.2.6)$$

Оскільки множини параметрів $\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}$ та $\boldsymbol{\theta}_{X_j|\ Pa_{X_j}}$, $i \neq j$, незалежні, то апостеріорна щільність $p(\boldsymbol{\theta}|D)$ запишеться у вигляді добутку апостеріорних щільностей параметрів $\boldsymbol{\theta}_{X_i|\ Pa_{X_i}}$ умовних імовірнісних розподілів вершин мережі:

$$p(\boldsymbol{\theta}|D) = \prod_{i=1}^{n} p\left(\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}|D\right). \qquad (2.2.7)$$

Спрогнозувати ймовірність того, що наступне вибіркове значення $(X_1[M+1], \dots, X_n[M+1])$ набуватиме певних значень можна за формулою:

$$P(X_1[M+1], \dots, X_n[M+1]|D) =$$

$$= \prod_{i=1}^{n} \int_{\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}} P\left(X_i[M+1]|\, Pa_{X_i}[M+1], \boldsymbol{\theta}_{X_i|\,Pa_{X_i}}\right) P\left(\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}|D\right) d\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}$$

(2.2.8)

Згідно з (2.2.8) ми можемо знайти умовні ймовірності для кожної вершини мережі X_i окремо, а потім перемножити здобуті результати.

Про задання апріорного розподілу вектору параметрів $\boldsymbol{\theta}$ в байєсівській мережі. Кожна вершина X_i мережі має мультиноміальний розподіл з вектором параметрів $\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}$, який має розподіл Діріхле з гіперпараметрами $\boldsymbol{\alpha}_{X_i|\,pa_{X_i}} = (\alpha_{x_i^1|\,pa_{X_i}}, \dots, \alpha_{x_i^{K_i}|\,pa_{X_i}})$, де K_i – число станів вершини X_i. Ми можемо скористатися результатами пілотного експерименту, а саме обчислити

$$\alpha_{x_i|\,Pa_{X_i}} = \alpha\left[x_i, pa_{X_i}\right],$$

$\alpha\left[x_i, pa_{X_i}\right]$ –число спостережень, для яких $X_i = x_i$, $Pa_{X_i} = pa_{X_i}$ у вибірці D'. В цьому випадку байєсівське оцінювання параметрів $\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}$ еквівалентно оцінювання параметрів $\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}$ згідно з методом максимальної правдоподібності за об'єднанням вибірок $D' \cup D$.

Отже, байєсівські точкові оцінки параметрів дорівнюють

$$P\left(x_i|pa_{X_i}, D\right) = \frac{\alpha_{x_i|\,pa_{X_i}} + M\left[x_i, pa_{X_i}\right]}{\alpha_{pa_{X_i}} + M\left[pa_{X_i}\right]}, \quad \alpha_{pa_{X_i}} = \sum_{x_i} \alpha_{x_i|\,pa_{X_i}},$$

де величини $M\left[x_i,\ pa_{X_i}\right]$ обчислюються для кожної комбінації значень вершини X_i та значень її батьківських вершин Pa_{X_i}, а суми $M\left[pa_{X_i}\right] = \sum_{x_i} M\left[x_i, pa_{X_i}\right]$ знаходяться по всіх можливих значеннях вершини X_i [1, 4].

2.3. Дивергенція Кульбака-Лейблера

Проведемо дослідження байєсівських оцінок параметрів умовних імовірнісних розподілів вершин байєсівської мережі порівняно з експертними оцінками.

1. Змоделюємо навчальну вибірку даних з експертними оцінками параметрів умовних імовірнісних розподілів вершин мережі та застосуємо байєсівський метод статистичного оцінювання для знаходження байєсівських оцінок. Кожна вершина X_i має мультиноміальний розподіл з вектором параметрів $\boldsymbol{\theta}_{X_i|\,Pa_{X_i}}$, який розподілений згідно з розподілом Діріхле з гіперпараметрами $\boldsymbol{\alpha}_{X_i|\,pa_{X_i}}$. Байєсівські оцінки обчислюються так:

$$P\left(x_i|pa_{X_i}, D\right) = \frac{\alpha_{x_i|\,pa_{X_i}} + M\left[x_i, pa_{X_i}\right]}{\alpha_{pa_{X_i}} + M\left[pa_{X_i}\right]},$$

при цьому гіперпараметри $\alpha_{x_i|\,pa_{X_i}}$ знаходяться за апріорною інформацією, здобутою за результатами пілотного експерименту так:

$$\alpha_{x_i|\,pa_{X_i}} = \alpha \cdot P'\left(x_i, pa_{X_i}\right),$$

де α – обсяг вибірки в експерименті, $P'\left(x_i, pa_{X_i}\right)$ – імовірність набуття вершиною X_i і її батьками Pa_{X_i} станів x_i та pa_{X_i} відповідно.

2. Обчислимо дивергенцію Кульбака-Лейблера між експертними та емпіричними умовними ймовірнісними розподілами вершин мережі.

3. Порівняємо експертні оцінки та байєсівські оцінки параметрів умовних імовірнісних розподілів вершин.

Залежність відстані Кульбака-Лейблера $D\left(P \parallel \hat{P}\right)$ від обсягу M навчальної вибірки наведена на рис. 2.3.1. При збільшенні обсягу вибірки відстань Кульбака-Лейблера між експертними та емпіричними умовними ймовірнісними розподілами вершин зменшується, при цьому суттєве скорочення відстані спостерігається для малих обсягів вибірки і повільніше – для великих обсягів.

Відстань Кульбака-Лейблера між експертним та емпіричним умовними ймовірнісними розподілами з параметрами, здобутими *методом максимальної правдоподібності* зображено неперервною лінією синього кольору на рис. 2.3.1. Відстань Кульбака-Лейблера між експертним та емпіричним умовними ймовірнісними розподілами з параметрами, здобутими байєсівським методом за умови наявної апріорної інформації для всіх вершин: *Focused*, *Age*, *Good Student*, *Extra Training*, *Vehicle Size*, *Driver Quality*, *Driving History*, *Vehicle Year*, *Airbag*, *Antilock Brakes*, *Accident*, *Cost*, зображено неперервною лінією червоного кольору на рис. 2.3.1.

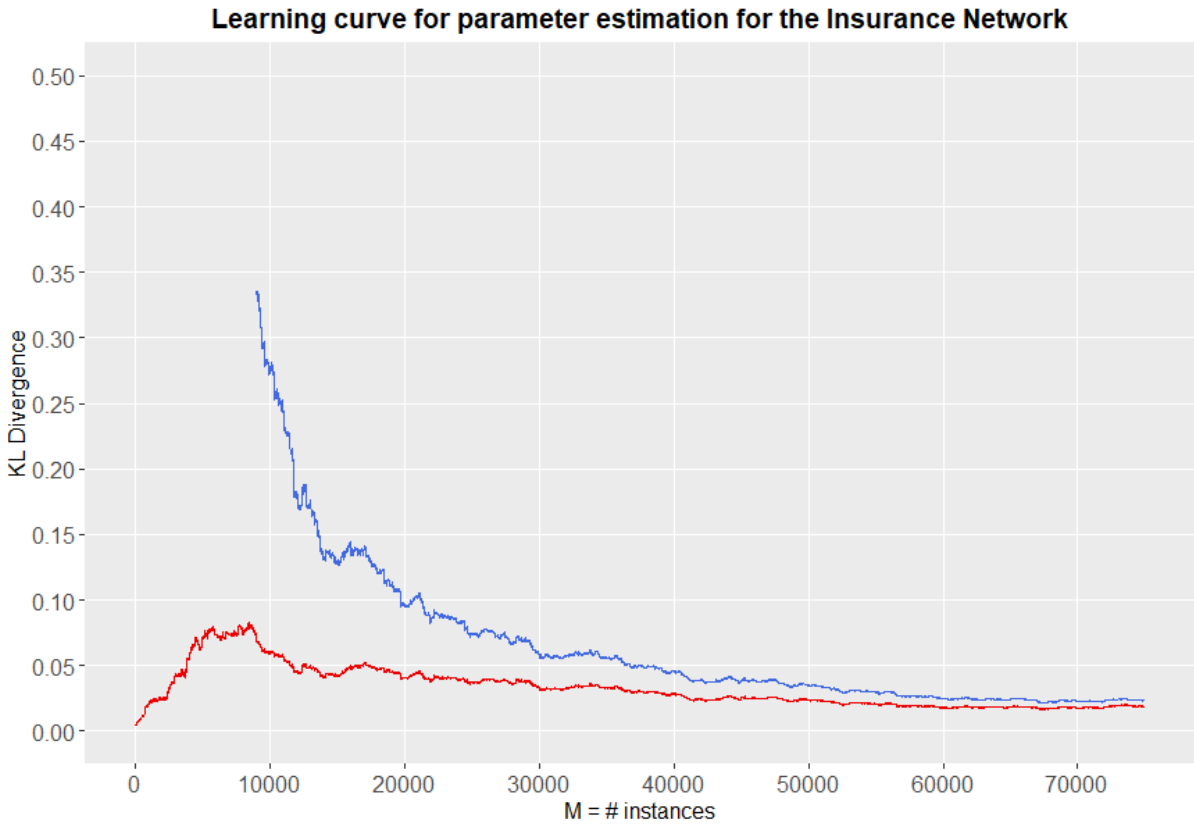


Рис. 2.3.1. Крива навчання байєсівської мережі Insurance

Байєсівські оцінки – спроможні та асимптотично ефективні оцінки параметрів умовних імовірнісних розподілів вершин байєсівської мережі.

Наявність апріорної інформації дозволяє набагато швидше здобути байєсівські оцінки параметрів близькі до експертних оцінок і розпочати застосовувати алгоритми формування ймовірнісного висновку для побудови моделей міркувань на основі байєсівської мережі.

3. Точний імовірнісний висновок

3.1. Апріорний та апостеріорний розподіли

Побудувавши структуру мережі самостійно або залучивши експертів для її побудови та навчивши мережу, дослідник може моделювати різноманітні ситуації, тобто задавати вершинам мережі деякі значення станів і отримувати найбільш імовірні значення станів для інших вершин мережі. Процес обчислення апостеріорних імовірностей станів вершини X_i при заданих станах спостережуваних вершин називатимемо **ймовірнісним висновком**.

Апріорний розподіл будь-якої вершини X_i обчислюється за спільним розподілом $P(X_1, X_2, \dots, X_n)$ вершин $X_1, X_2, \dots, X_n$:

$$P(X_i) = \sum_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n} P(X_1, \dots, X_n),$$

де спільний розподіл $P(X_1, \dots, X_n)$ обчислюється за формулою множення для байєсівської мережі:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i | Parents_{X_i}\right).$$

Апостеріорний розподіл вершини X_i за умови заданого стану вершини X_j обчислюється за формулою умовної ймовірності:

$$P\left(X_i | X_j\right) = \frac{P\left(X_i, X_j\right)}{P\left(X_j\right)},$$

$$P\left(X_i, X_j\right) = \sum_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_{j-1},X_{j+1},\dots,X_n} P(X_1, \dots, X_n),$$

$$P\left(X_j\right) = \sum_{X_1,\dots,X_{j-1},X_{j+1},\dots,X_n} P(X_1, \dots, X_n).$$

де спільний розподіл $P(X_1, \dots, X_n)$ обчислюється за формулою множення для байєсівської мережі:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i | Parents_{X_i}\right).$$

3.2. Алгоритм Sum-Product Variable Elimination

Алгоритм точного ймовірнісного висновку Sum-Product Variable Elimination застосовано для обчислення розподілів станів вершин байєсівської мережі Insurance з метою здобуття портретів дорожньо-транспортних пригод.

Ідея алгоритму полягає в послідовному обчисленні сум добутків умовних імовірнісних розподілів за всіма станами вершин, які виключаються [1]. Ми застосуємо цей підхід до обчислення маргінального розподілу вершини *Cost*, виключаючи всі інші вершини мережі:

$$P(X_{12}) = \sum_{X_1,\dots,X_{11}} P(X_1,\dots,X_{12}) =$$

$$= \sum_{X_1,\dots,X_{11}} P(X_{12}|X_{11},X_8)P(X_{11}|X_{10},X_9,X_8,X_7,X_5)P(X_{10}|X_8)P(X_9|X_8)\cdot$$

$$\cdot P(X_8)P(X_7|X_1)P(X_6|X_5)P(X_5|X_1,X_2,X_4)P(X_4)P(X_3|X_2,X_1)P(X_2)P(X_1).$$

Розглянемо виключення вершин в наступному порядку

$$Z = \{X_4,X_6,X_2,X_3,X_1,X_5,X_7,X_9,X_{10},X_8,X_{11}\},$$

який знайдений згідно з жадібним алгоритмом [1]. На вхід алгоритму подається неорієнтований граф з вершинами, які перебираються та видаляються в порядку зростання числа сусідів, кожній вершині присвоюється свій порядковий номер виключення (ранг виключення).

Ми можемо переписати вищенаведену суму у такому вигляді:

$$P(X_{12}) = \sum_{X_{11}}\sum_{X_8} P(X_{12}|X_{11},X_8)P(X_8)\sum_{X_{10}} P(X_{10}|X_8)\sum_{X_9} P(X_9|X_8)\cdot$$

$$\cdot\sum_{X_7}\sum_{X_5} P(X_{11}|X_{10},X_9,X_8,X_7,X_5)\sum_{X_1} P(X_1)P(X_7\,|X_1)\cdot$$

$$\cdot\sum_{X_3}\sum_{X_2} P(X_3|X_2,X_1)P(X_2)\sum_{X_6} P(X_6|X_5)\sum_{X_4} P(X_5|X_4,X_2,X_1)P(X_4).$$

Обчислення зручно подати у вигляді таблиці 3.2.1.

Таблиця 3.2.1. Алгоритм Sum-Product Variable Elimination

Крок	Виключена вершина	Умовні ймовірнісні розподіли та множники	Новий множник
1	X_4	$P(X_5 \mid X_1, X_2, X_4), P(X_4)$	$\tau_1(X_1, X_2, X_5)$
2	X_6	$P(X_6 \mid X_5)$	$\tau_2(X_5)$
3	X_2	$P(X_3 \mid X_1, X_2), P(X_2),$ $\tau_1(X_1, X_2, X_5)$	$\tau_3(X_5, X_3, X_1)$
4	X_3	$\tau_3(X_5, X_3, X_1)$	$\tau_4(X_5, X_1)$
5	X_1	$P(X_1), P(X_7 \mid X_1), \tau_4(X_5, X_1)$	$\tau_5(X_7, X_5)$
6	X_5	$P(X_{11} \mid X_{10}, X_9, X_8, X_7, X_5),$ $\tau_2(X_5), \tau_5(X_7, X_5)$	$\tau_6(X_{11}, X_{10}, X_9, X_8, X_7)$
7	X_7	$\tau_6(X_{11}, X_{10}, X_9, X_8, X_7)$	$\tau_7(X_{11}, X_{10}, X_9, X_8)$
8	X_9	$\tau_7(X_{11}, X_{10}, X_9, X_8),$ $P(X_9 \mid X_8)$	$\tau_8(X_{11}, X_{10}, X_8)$
9	X_{10}	$\tau_8(X_{11}, X_{10}, X_8), P(X_{10} \mid X_8)$	$\tau_9(X_{11}, X_8)$
10	X_8	$\tau_9(X_{11}, X_8), P(X_{12} \mid X_{11}, X_8),$ $P(X_8)$	$\tau_{10}(X_{12}, X_{11})$
11	X_{11}	$\tau_{10}(X_{12}, X_{11})$	$\tau_{11}(X_{12})$

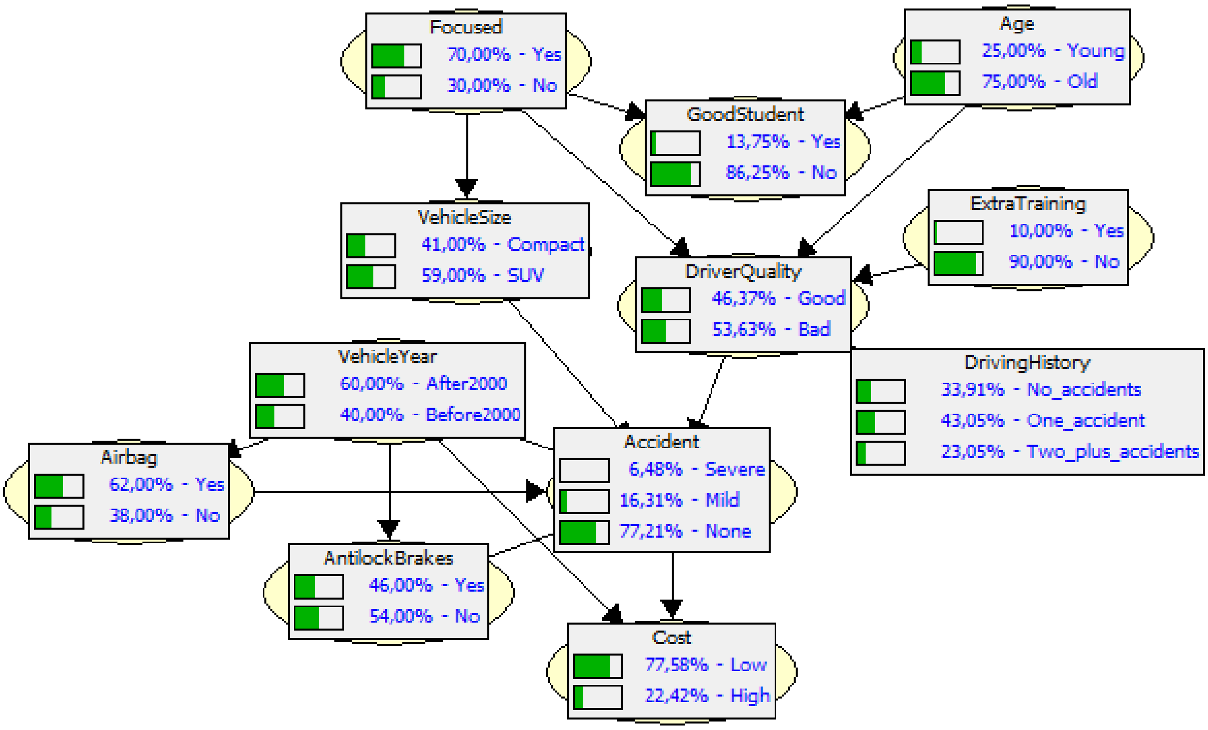


Рис. 3.2.1. Маргінальний розподіл всіх вершин мережі Insurance

3.3. Алгоритм Conditioning

Альтернативний підхід точного ймовірнісного висновку заснований на спостереження певних станів вершин мережі. Ідея Conditioning алгоритму полягає в тому, що ми будуємо низку мереж для кожного стану вершини X_j, що спостерігається. Ці мережі мають однакову структуру, але різні параметри умовних імовірнісних розподілів вершин. Ми застосовуємо Sum-Product Variable Elimination алгоритм для кожної мережі для обчислення спільного розподілу $P(X_i, X_j = x_j)$ і підсумовуємо ці ймовірності для знаходження розподілу $P(X_i)$ вершини X_i:

$$P(X_i) = \sum_{x_j} P(X_i, X_j = x_j) = \sum_{x_j} P(X_i \middle| X_j = x_j) P(X_j = x_j).$$

Застосуємо цей підхід до обчислення маргінального розподілу вершини *Cost*, при цьому спільний розподіл $P(X_1, \dots, X_{12})$ подається як добуток умовних імовірнісних розподілів:

$$P(X_{12}) = \sum_{X_{11}} \sum_{X_1, \dots, X_{10}} P(X_1, \dots, X_{12}) =$$

$$\sum_{X_{11}} \sum_{X_8} P(X_{12}|X_{11}, X_8) P(X_8) \sum_{X_{10}} P(X_{10}|X_8) \sum_{X_9} P(X_9|X_8) \cdot$$

$$\cdot \sum_{X_7} \sum_{X_5} P(X_{11}|X_{10}, X_9, X_8, X_7, X_5) \sum_{X_1} P(X_1) P(X_7\,|X_1) \cdot$$

$$\cdot \sum_{X_3} \sum_{X_2} P(X_3|X_2, X_1) P(X_2) \sum_{X_6} P(X_6|X_5) \sum_{X_4} P(X_5|X_4, X_2, X_1) P(X_4),$$

де для кожного стану вершини X_{11} маємо:

$$P\big(X_{12}, X_{11} = x_j\big)$$

$$= \sum_{X_8} P(X_{12} \middle| X_{11} = x_j, X_8) P(X_8) \sum_{X_{10}} P(X_{10}|X_8) \sum_{X_9} P(X_9|X_8) \cdot$$

$$\cdot \sum_{X_7} \sum_{X_5} P(X_{11} = x_j|X_{10}, X_9, X_8, X_7, X_5) \sum_{X_1} P(X_1) P(X_7\,|X_1) \cdot$$

$$\cdot \sum_{X_3} \sum_{X_2} P(X_3|X_2, X_1) P(X_2) \sum_{X_6} P(X_6|X_5) \sum_{X_4} P(X_5|X_4, X_2, X_1) P(X_4).$$

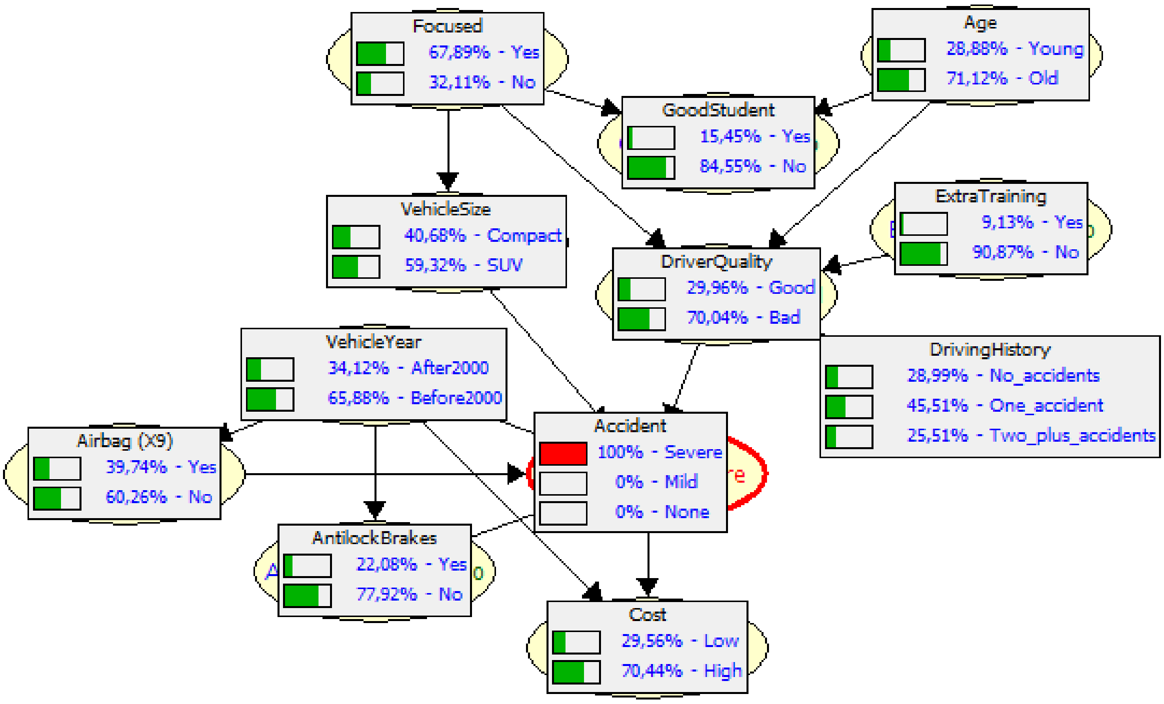


Рис. 3.3.1. Апостеріорний розподіл $P(X_i|X_j = x_j)$ вершин мережі при заданому стані *Severe* спостережуваної вершини *Accident*

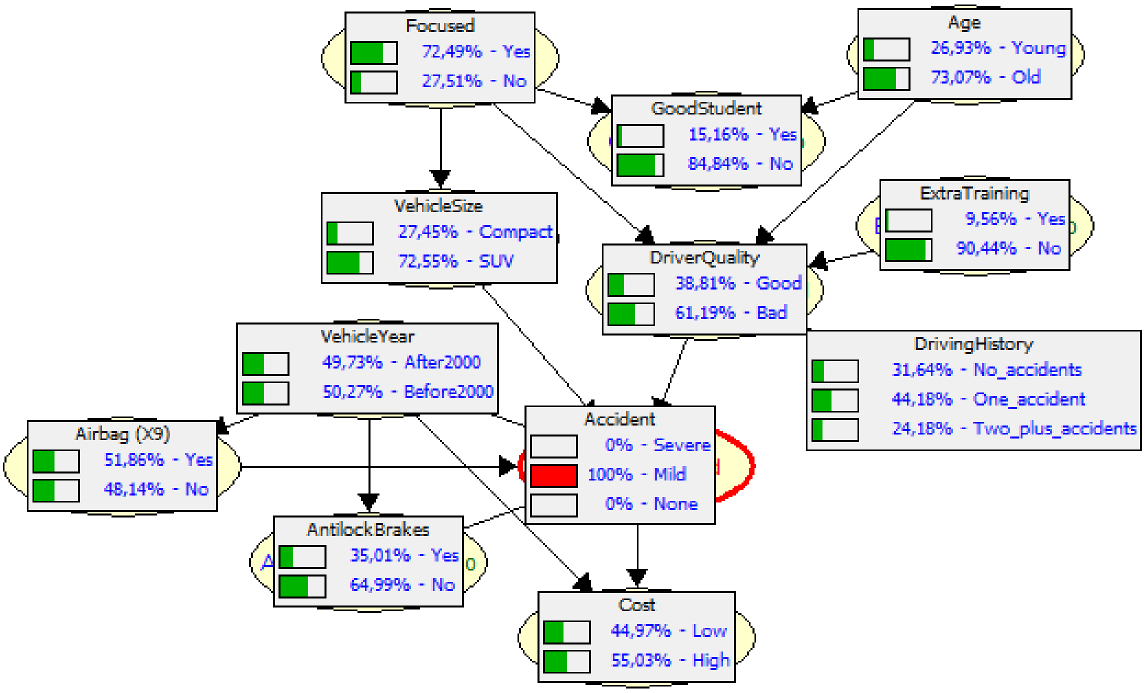


Рис. 3.3.2. Апостеріорний розподіл $P(X_i|X_j = x_j)$ вершин мережі при заданому стані *Mild* спостережуваної вершини *Accident*

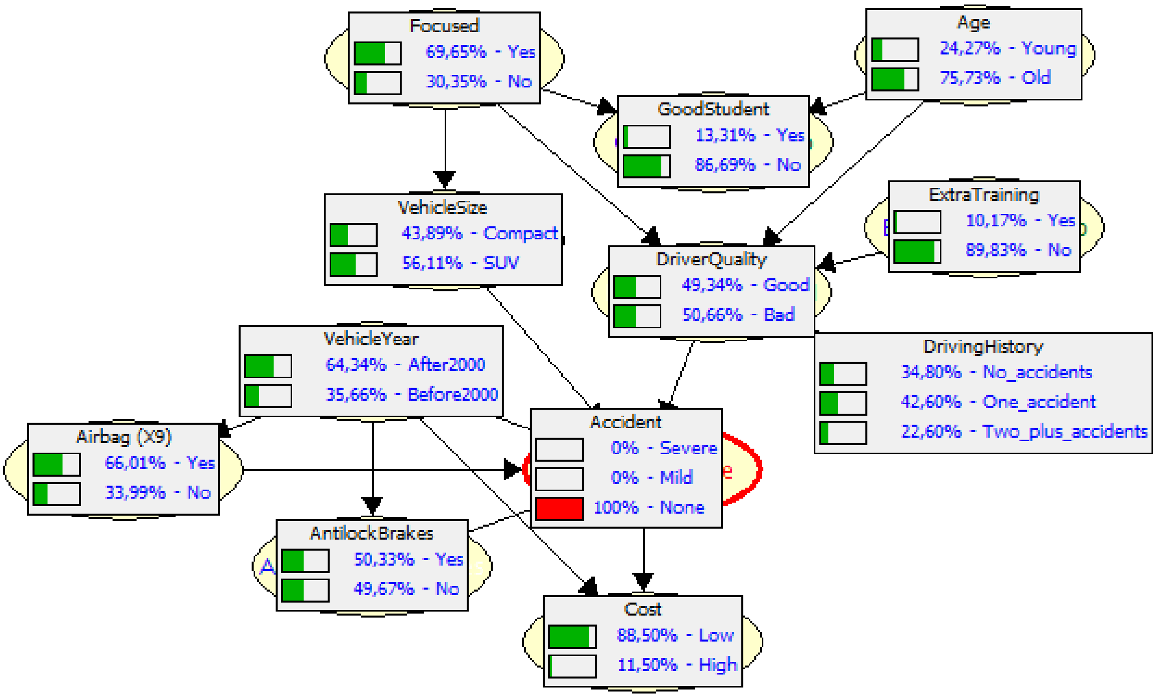


Рис. 3.3.3. Апостеріорний розподіл $P(X_i|X_j = x_j)$ вершин мережі при заданому стані *None* спостережуваної вершини *Accident*

3.4. Алгоритм Max-Product Variable Elimination

Алгоритм точного ймовірнісного висновку Max-Product Variable Elimination застосовано для обчислення найбільш імовірних станів вершин байєсівської мережі Insurance з метою здобуття портретів дорожньо-транспортних пригод.

Алгоритм точного ймовірнісного висновку Max-Product Variable Elimination призначений для обчислення максимуму апостеріорної ймовірності станів множини вершин W при заданих станах спостережуваних вершин E. Ідея алгоритму полягає в послідовному обчисленні максимумів добутків умовних імовірнісних розподілів за всіма станами вершин, які виключаються [1].

Ми застосуємо цей підхід до обчислення: найбільш імовірних станів всіх вершин мережі; найбільш імовірних станів певної множини вершин W при заданих станах спостережуваних вершин E. Розглянемо

$$\max_{X_1,\dots,X_{12}} P(X_1,\dots,X_{12}) =$$
$$= \max_{X_1,\dots,X_{12}} P(X_{12}|X_{11},X_8)P(X_{11}|X_{10},X_9,X_8,X_7,X_5)P(X_{10}|X_8)P(X_9|X_8)\cdot$$
$$\cdot P(X_8)P(X_7|X_1)P(X_6|X_5)P(X_5|X_1,X_2,X_4)P(X_4)P(X_3|X_2,X_1)P(X_2)P(X_1).$$

Розглянемо виключення вершин в наступному порядку

$$Z = \{X_4,X_6,X_2,X_3,X_1,X_5,X_7,X_9,X_{10},X_8,X_{11},X_{12}\},$$

який знайдений згідно з жадібним алгоритмом [1].

Ми можемо переписати вищенаведений максимум у такому вигляді:

$$\max_{X_1,\dots,X_{12}} P(X_1,\dots,X_{12}) =$$
$$= \max_{X_{12},X_{11},X_8} P(X_{12}|X_{11},X_8)P(X_8)\max_{X_{10}} P(X_{10}|X_8)\max_{X_9} P(X_9|X_8)$$
$$\max_{X_7}\max_{X_5} P(X_{11}|X_{10},X_9,X_8,X_7,X_5)\max_{X_1} P(X_1)P(X_7|X_1)$$
$$\max_{X_3}\max_{X_2} P(X_3|X_2,X_1)P(X_2)\max_{X_6} P(X_6|X_5)\max_{X_4} P(X_5|X_4,X_2,X_1)P(X_4).$$

Обчислення зручно подати в таблиці 3.4.1.

Тепер ми розглянемо Traceback-MAP алгоритм, який дозволяє ідентифікувати найбільш імовірні стани всіх вершин мережі одразу як процес виключення змінних буде завершений:

$$X_1^*,\dots,X_{12}^* = \underset{X_1,\dots,X_{12}}{argmax} P(X_1,\dots,X_{12}).$$

Ми починаємо з обчислення $X_{12}^* = \underset{X_{12}}{argmax}\,\psi_{12}(X_{12})$. Стан X_{12}^* є станом вершини X_{12} серед найбільш імовірних станів всіх вершин $X_1,\dots,X_{12}$ мережі. Знаючи X_{12}^*, ми можемо обчислити $X_{11}^* = \underset{X_{11}}{argmax}\,\psi_{11}(X_{12}^*,X_{11})$. Стан X_{11}^* є станом вершини X_{11} серед найбільш імовірних станів всіх вершин $X_1,\dots,X_{12}$ мережі. Далі ми скористаємося описаною процедурою для всіх вершин, які залишилися. Отже,

$$X_8^* = \underset{X_8}{argmax}\,\psi_{10}(X_{12}^*,X_{11}^*,X_8),\ X_{10}^* = \underset{X_{10}}{argmax}\,\psi_9(X_{11}^*,X_{10},X_8^*),$$

$$X_9^* = \underset{X_9}{argmax}\psi_8(X_{11}^*, X_{10}^*, X_9, X_8^*),\ X_7^* = \underset{X_7}{argmax}\psi_7(X_{11}^*, X_{10}^*, X_9^*, X_8^*, X_7),$$

$$X_5^* = \underset{X_5}{argmax}\psi_6(X_{11}^*, X_{10}^*, X_9^*, X_8^*, X_7^*, X_5),\ X_1^* = \underset{X_1}{argmax}\psi_5(X_7^*, X_5^*, X_1),$$

$$X_3^* = \underset{X_3}{argmax}\psi_4(X_5^*, X_3, X_1^*),\ X_2^* = \underset{X_2}{argmax}\psi_3(X_5^*, X_3^*, X_2, X_1^*),$$

$$X_6^* = \underset{X_6}{argmax}\psi_2(X_6, X_5^*),\ X_4^* = \underset{X_4}{argmax}\psi_1(X_1^*, X_2^*, X_4, X_5^*).$$

Найбільш імовірні стани $X_1^*, \ldots, X_{12}^*$ всіх вершин байєсівської мережі наведені на рисунку 3.4.1.

Якщо W множина всіх вершин мережі, а $E = \emptyset$, тоді $MAP(W|E = e)$ поданий на рисунку 3.4.1.

Якщо $W = \{X_1, X_2, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{11}\}$ шукана множина вершин мережі, а $E = \{X_3, X_6, X_{12}\}$ множина спостережуваних вершин, тоді $MAP(W|E = e)$ поданий на рисунку 3.4.2.

Таблиця 3.4.1. Алгоритм Max-Product Variable Elimination

Крок	Виключена вершина	Умовні ймовірнісні розподіли та множники	Новий множник
1	X_4	$P(X_5\|X_1, X_2, X_4), P(X_4)$	$\tau_1(X_1, X_2, X_5)$
2	X_6	$P(X_6\|X_5)$	$\tau_2(X_5)$
3	X_2	$P(X_3\|X_1, X_2), P(X_2),$ $\tau_1(X_1, X_2, X_5)$	$\tau_3(X_5, X_3, X_1)$
4	X_3	$\tau_3(X_5, X_3, X_1)$	$\tau_4(X_5, X_1)$
5	X_1	$P(X_1), P(X_7\|X_1), \tau_4(X_5, X_1)$	$\tau_5(X_7, X_5)$
6	X_5	$P(X_{11}\|X_{10}, X_9, X_8, X_7, X_5),$ $\tau_2(X_5), \tau_5(X_7, X_5)$	$\tau_6(X_{11}, X_{10}, X_9, X_8, X_7$
7	X_7	$\tau_6(X_{11}, X_{10}, X_9, X_8, X_7)$	$\tau_7(X_{11}, X_{10}, X_9, X_8)$
8	X_9	$\tau_7(X_{11}, X_{10}, X_9, X_8), P(X_9\|X_8$	$\tau_8(X_{11}, X_{10}, X_8)$
9	X_{10}	$\tau_8(X_{11}, X_{10}, X_8), P(X_{10}\|X_8)$	$\tau_9(X_{11}, X_8)$
10	X_8	$\tau_9(X_{11}, X_8), P(X_{12}\|X_{11}, X_8),$ $P(X_8)$	$\tau_{10}(X_{12}, X_{11})$
11	X_{11}	$\tau_{10}(X_{12}, X_{11})$	$\tau_{11}(X_{12})$
12	X_{12}	$\tau_{11}(X_{12})$	$\underset{X_1,\ldots,X_{12}}{max} P(X_1, \ldots, X_{12})$

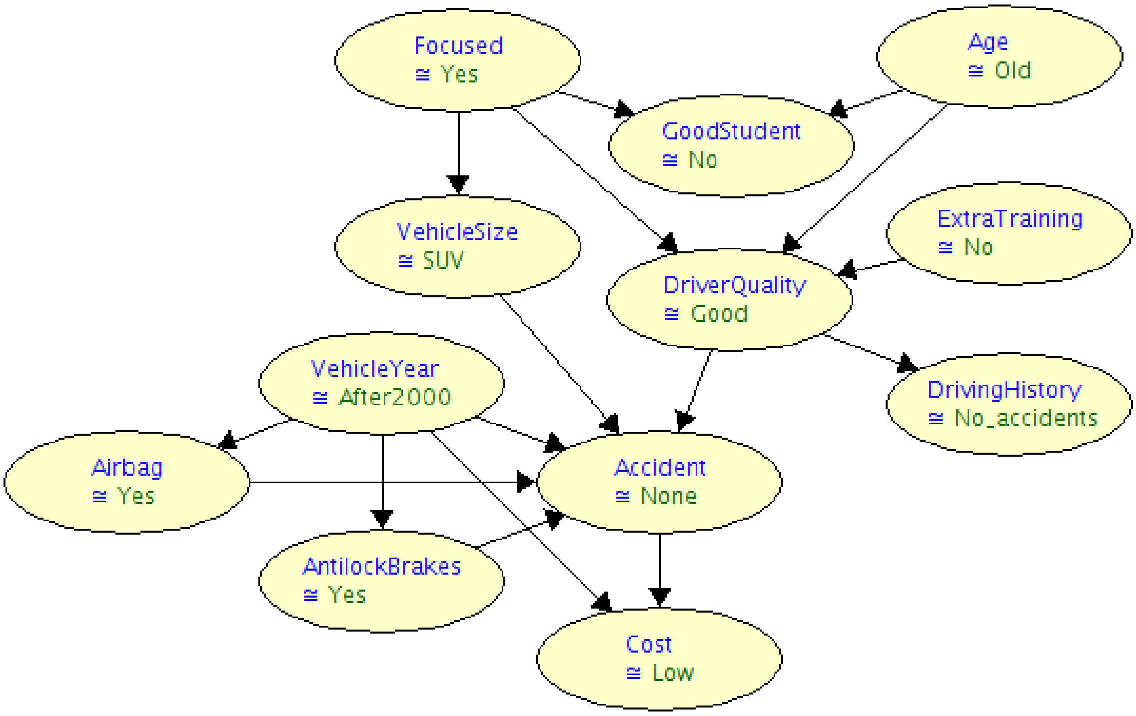


Рис. 3.4.1. Найбільш імовірні стани всіх вершин мережі Insurance

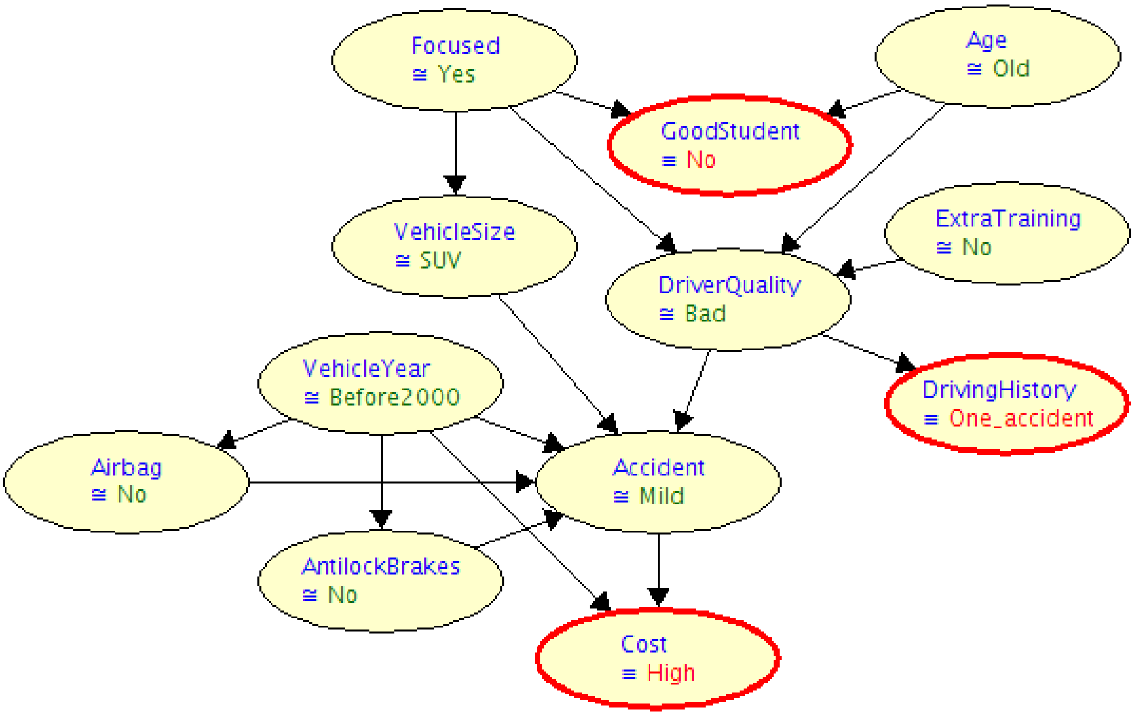


Рис. 3.4.2. Найбільш імовірні стани вершин *Focused*, *Age*, *Vehicle Size*, *Driving Quality*, *Extra Training*, *Vehicle Year*, *Airbag*, *Antilock Brakes*, *Accident* при заданих станах спостережуваних вершин *Good Student*, *Driving History*, *Cost*

3.5. Алгоритм Max-Sum-Product Variable Elimination

Алгоритм точного висновку Max-Sum-Product Variable Elimination застосовано для обчислення найбільш імовірних станів вершин байєсівської мережі Insurance, за умови відсутньої інформації про стани інших вершин, з метою здобуття портретів дорожньо-транспортних пригод.

Розглянемо алгоритм точного імовірнісного висновку Max-Sum-Product Variable Elimination для обчислення максимуму маргінальної апостеріорної ймовірності станів множини вершин Y, маргіналізованих за множиною станів вершин Z, при заданих станах спостережуваних вершин E. Ідея алгоритму полягає в послідовному обчисленні сум добутків умовних імовірнісних розподілів за всіма станами вершин Z, які виключаються, та подальшій максимізації суми за множиною станів вершин Y [1].

Ми застосуємо цей підхід до обчислення:

1) найбільш імовірних станів вершин мережі Y, маргіналізованих за множиною станів вершин Z;
2) найбільш імовірних станів вершин мережі Y, маргіналізованих за множиною станів вершин Z, при заданих станах спостережуваних вершин E:

$$MAP(Y|E=e) = \underset{Y}{argmax} \sum_{Z} P(Y,Z|E=e).$$

Ми застосуємо цей підхід до обчислення найбільш імовірних станів вершин *Focused*, *Age*, *Extra Training*, *Vehicle Year*:

$$\max_{X_1,X_2,X_4,X_8} \sum_{X_3,X_5,X_6,X_7,X_9,X_{10},X_{11},X_{12}} P(X_1,\dots,X_{12}).$$

Розглянемо виключення вершин в наступному порядку

$$Z = \{X_3, X_6, X_{12}, X_{11}, X_{10}, X_9, X_7, X_5, X_8, X_4, X_2, X_1\}.$$

Ми можемо переписати вищенаведений максимум у такому вигляді:

$$\max_{X_1,X_2,X_4,X_8} \sum_{X_5} P(X_5|X_4,X_2,X_1) \sum_{X_7} P(X_7|X_1) \sum_{X_9} P(X_9|X_8) \sum_{X_{10}} P(X_{10}|X_8) \cdot$$

$$\cdot \sum_{X_{11}} P(X_{11}|X_{10}, X_9, X_8, X_7, X_5) \sum_{X_{12}} P(X_{12}|X_{11}, X_8) \sum_{X_6} P(X_6|X_5) \sum_{X_3} P(X_3|X_1, X_2).$$

Обчислення зручно подати в таблиці 3.5.1. Тепер ми розглянемо Traceback-MAP алгоритм, який дозволяє ідентифікувати найбільш імовірні стани вершин мережі одразу як процес виключення змінних буде завершений:

$$X_1^* = \underset{X_1}{argmax} \psi_{12}(X_1),\ X_2^* = \underset{X_2}{argmax} \psi_{11}(X_1^*, X_2),$$

$$X_4^* = \underset{X_4}{argmax} \psi_{10}(X_1^*, X_2^*, X_4),\ X_8^* = \underset{X_8}{argmax} \psi_9(X_1^*, X_2^*, X_4^*, X_8).$$

Якщо $Y = \{X_1, X_2, X_4, X_8\}$ шукана множина вершин мережі, а $Z = \{X_5, X_7, X_9, X_{10}, X_{11}\}$ множина вершин мережі, інформація за якими відсутня, $E = \{X_3, X_6, X_{12}\}$ множина спостережуваних вершин мережі, тоді $MAP(Y|E = e) = \underset{Y}{argmax} \sum_Z P(Y, Z|E = e)$. поданий на рисунку 3.5.1.

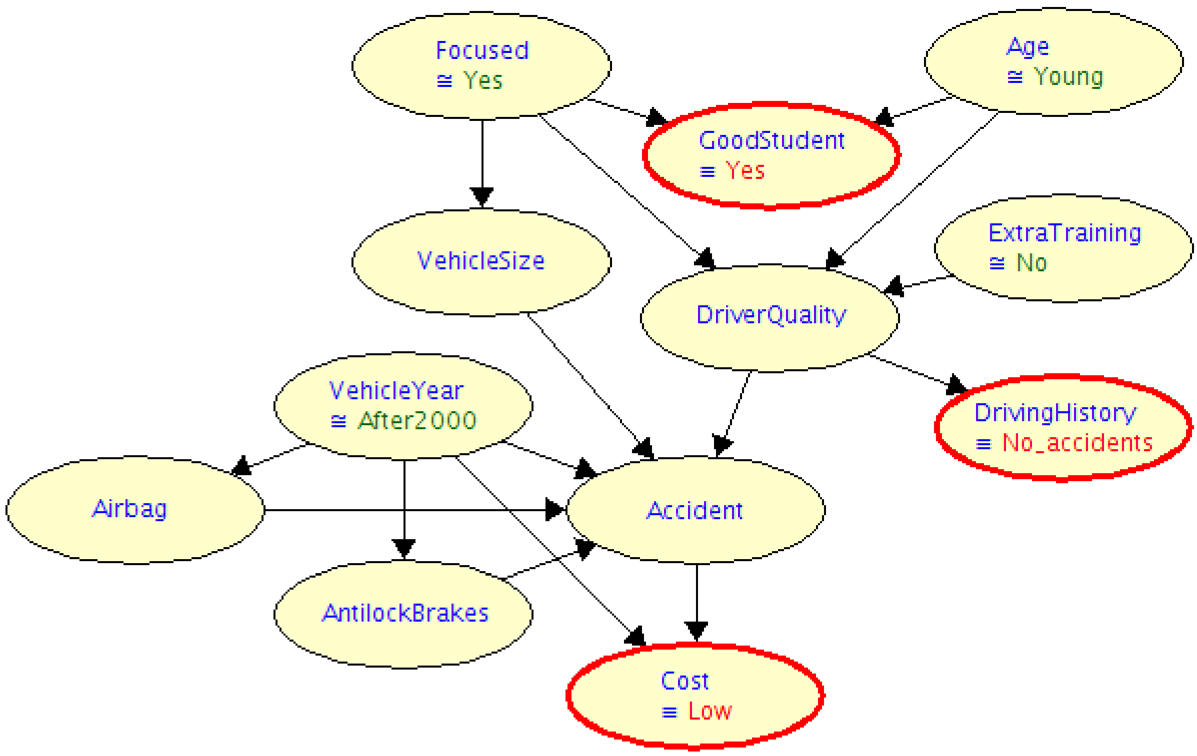


Рис. 3.5.1. Найбільш імовірні стани вершин *Focused, Age, Extra Training, Vehicle Size* при заданих станах спостережуваних вершин *Good Student, Driving History, Cost*

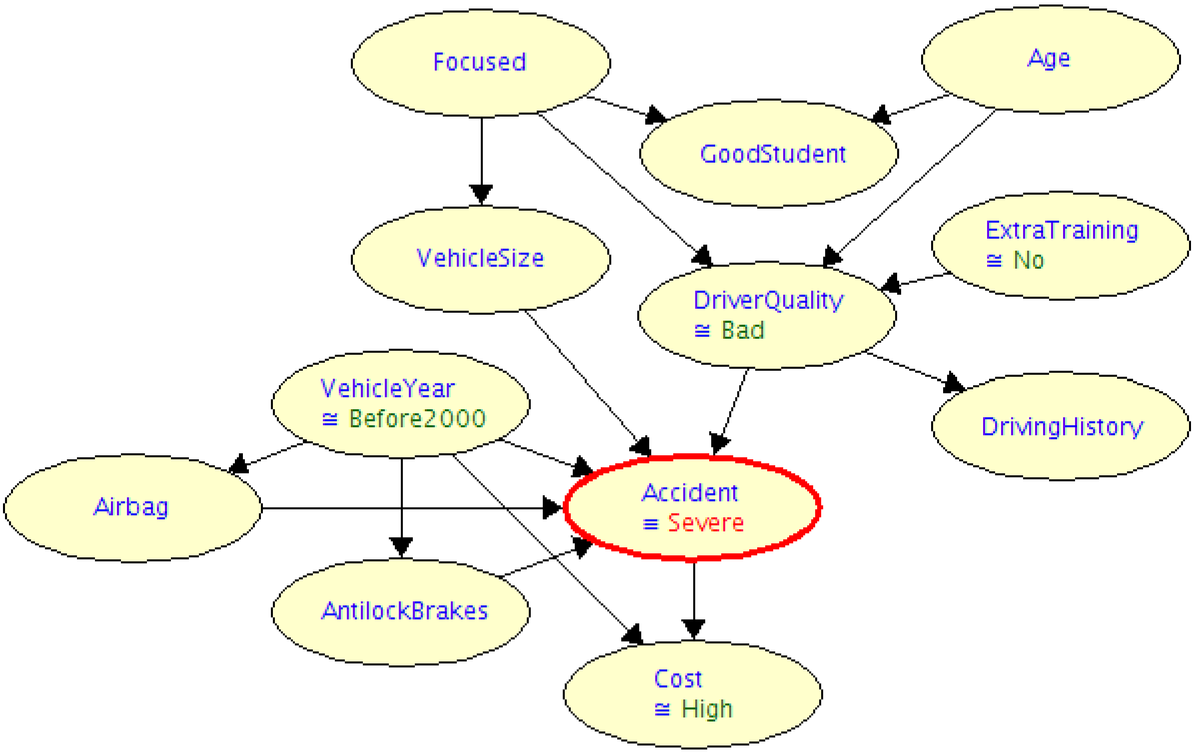


Рис. 3.5.2. Найбільш імовірні стани вершин *Extra Training*, *Driving Quality*, *Vehicle Year*, *Cost* при заданому стані спостережуваної вершини *Accident*

Таблиця 3.5.1. Алгоритм Max-Sum-Product Variable Elimination

Крок	Виключена вершина	Умовні ймовірнісні розподіли та множники	Новий множник
1	X_3	$P(X_3 \mid X_1, X_2)$	$\tau_1(X_1, X_2)$
2	X_6	$P(X_6 \mid X_5)$	$\tau_2(X_5)$
3	X_{12}	$P(X_{12} \mid X_{11}, X_8)$	$\tau_3(X_8, X_{11})$
4	X_{11}	$P(X_{11} \mid X_{10}, X_9, X_8, X_7, X_5)$, $\tau_3(X_8, X_{11})$	$\tau_4(X_5, X_7, X_8, X_9, X_{10})$
5	X_{10}	$\tau_4(X_5, X_7, X_8, X_9, X_{10}), P(X_{10} \mid X_8)$	$\tau_5(X_5, X_7, X_8, X_9)$
6	X_9	$\tau_5(X_5, X_7, X_8, X_9), P(X_9 \mid X_8)$	$\tau_6(X_5, X_7, X_8)$
7	X_7	$\tau_6(X_5, X_7, X_8), P(X_7 \mid X_1)$	$\tau_7(X_1, X_5, X_8)$
8	X_5	$\tau_2(X_5), \tau_7(X_1, X_5, X_8)$, $P(X_5 \mid X_4, X_2, X_1)$	$\tau_8(X_1, X_2, X_4, X_8)$
9	X_8	$\tau_8(X_1, X_2, X_4, X_8), P(X_8)$	$\tau_9(X_1, X_2, X_4)$
10	X_4	$\tau_9(X_1, X_2, X_4), P(X_4)$	$\tau_{10}(X_1, X_2)$
11	X_2	$\tau_1(X_1, X_2), \tau_{10}(X_1, X_2), P(X_2)$	$\tau_{11}(X_1)$
12	X_1	$\tau_{11}(X_1), P(X_1)$	τ_{12}

Висновки

В навчальному посібнику наведено розв’язання задачі прогнозування страхового відшкодування майнової шкоди завданої внаслідок дорожньо-транспортної пригоди за допомогою методу машинного навчання – байєсівської нейронної мережі. Створено програмне забезпечення “learning Bayesian network”, основною задачею якого є навчання байєсівської нейронної мережі за допомогою методу максимальної правдоподібності та байєсівського методу статистичного оцінювання.

Під час навчання байєсівської нейронної мережі було встановлено, що байєсівський метод статистичного оцінювання дозволяє усунути недоліки, властиві методу максимальної правдоподібності, а саме:

1) оцінки максимальної правдоподібності не враховують апріорну інформацію про розподіл навчальної вибірки, а байєсівські оцінки дозволяють цю інформацію використати під час навчання;
2) оцінки максимальної правдоподібності погіршуються при фрагментації даних, а байєсівський метод починає працювати при обсязі навчальної вибірки, який дорівнює одиниці;
3) оцінки максимальної правдоподібності використовуються лише для вибірок великих обсягів (наприклад, для онлайн-навчання мережі), а байєсівський метод добре працює незалежно від обсягу навчальної вибірки.

Цікавою частиною роботи є побудова та візуалізація портретів дорожньо-транспортних пригод для визначення страхового відшкодування майнової шкоди (див. додаток Б). Багаточисельні експерименти з різними станами вершин байєсівської нейронної мережі Insurance дозволили знайти стани вершин, при яких імовірність високих витрат страхової компанії зменшувалась, а імовірність низьких страхових витрат збільшувалась.

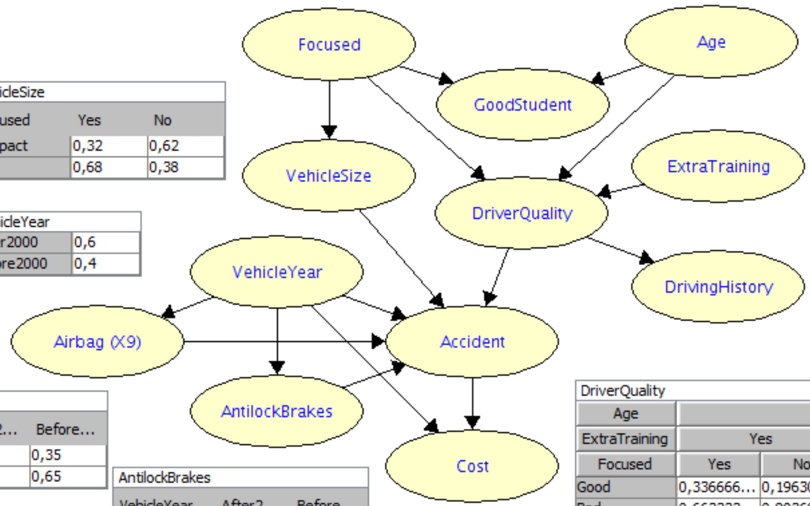
Фінансовий ризик і пов’язана з ним небезпека банкрутства – характерні особливості роботи кожної страхової компанії, тому розроблене програмне забезпечення “learning Bayesian network” дозволить забезпечити фінансову стійкість страхової компанії і кількісно оцінювати ризики в її діяльності.

Список рекомендованої літератури

1. D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques. The MIT Press, 2009.
2. D. Koller, Probabilistic Graphical Models 1: Representation // Stanford University. URL: https://www.coursera.org/learn/probabilistic-graphical-models. Online. Accessed 1-Sep-2020.
3. D. Koller, Probabilistic Graphical Models 2: Inference // Stanford University. URL: https://www.coursera.org/learn/probabilistic-graphical-models. Online. Accessed 1-Sep-2020.
4. D. Koller, Probabilistic Graphical Models 3: Learning // Stanford University. URL: https://www.coursera.org/learn/probabilistic-graphical-models. Online. Accessed 1-Sep-2020.
5. Турчин В.Н. Теория вероятностей и математическая статистика. Учебник для студентов высших учебных заведений. – Днепр, Издательство «Лира». – 2018. – 752 с.
6. Бондаренко Я.С. Посібник до вивчення дисципліни "Байєсівський аналіз даних" [Текст]/ Бондаренко Я.С., Кравченко С.В., Сологуб К.М. – Дніпро: Ліра, 2018. – 40 с.
7. Айвазян С.А. Байесовский подход в эконометрическом анализе // Прикладная Эконометрика. – 2008. – №1. – С. 93–130.
8. J. Pearl, Probabilistic Inference in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmenn, 1988.
9. J. Pearl, Causality: Models, Reasoning and Inference. Cambridge University Press, 2000.
10. K.B. Korb, A. E. Nicholson, Bayesian artificial intelligence. Chapman & Hall/CRC Press LLC, 2004.
11. C.M. Bishop, Pattern Recognition and Machine Learning. Springer, 2007.
12. A. Darwiche, Modeling and Reasoning with Bayesian networks. Cambridge University Press, 2009.

13. D. Barber, Bayesian Reasoning and Machine Learning. Cambridge University Press, 2012.
14. K.P. Murphy, Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
15. SamIam: Sensitivity Analysis, Modeling, Inference and More – програмний продукт для моделювання та аналізу байєсівських мереж, розроблений Automated Reasoning Group prof. Adnan Darwiche in University of California, Los Angeles (UCLA). URL: http://reasoning.cs.ucla.edu/samiam/
16. Бондаренко Я.С. Посібник до вивчення дисципліни «Імовірнісні графічні моделі». Частина 1. Подання байєсівської мережі [Текст]/ Бондаренко Я.С., Рачко Д.О., Розливан А.О. – Дніпро: Ліра, 2019.
17. Бондаренко Я.С. Посібник до вивчення дисципліни «Імовірнісні графічні моделі». Частина 2. Навчання байєсівської мережі [Текст]/ Бондаренко Я.С., Рачко Д.О., Розливан А.О. – Дніпро: Ліра, 2020
18. Ya.S. Bondarenko, D.O. Rachko, A.O. Rozlyvan Probabilistic Inference in Bayesian Insurance Network // Питання прикладної математики та математичного моделювання. - Д.: Вид-во ДНУ, 2020. - с. 3-20.
19. J. Binder, D. Koller, S. Russell, K. Kanazawa, Adaptive Probabilistic Networks with Hidden Variables, Machine Learning, vol. 29, no. 2-3, pp. 213-244, 1997.
20. M. Qazi, G.M. Fung, K.J. Meissner, E.R. Fontes, An Insurance Recommendation System Using Bayesian Networks, Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 274–278, 2017.
21. Xin Zou, Wen Long Yue, A Bayesian Network Approach to Causation Analysis of Road Accidents Using Netica, Journal of Advanced Transportation, Vol. 2017, pp. 1-18, 2017.
22. Zhu Xu, Yi Jiang, Fan Lin, Long Dai, The Analysis and Prevent in Traffic Accidents Based on Bayesian Network, Advanced Engineering Forum, Vol. 1, pp. 21-25, 2011.

Додаток А



Focused	
Yes	0,7
No	0,3

GoodStudent				
Focused	Yes		No	
Age	Young	Old	Young	Old
Yes	0,7	0,01	0,1	0,01
No	0,3	0,99	0,9	0,99

Age	
Young	0,25
Old	0,75

VehicleSize		
Focused	Yes	No
Compact	0,32	0,62
SUV	0,68	0,38

ExtraTraining	
Yes	0,1
No	0,9

VehicleYear	
After2000	0,6
Before2000	0,4

DrivingHistory		
DriverQuality	Good	Bad
No_accidents	0,5	0,2
One_accident	0,35	0,5
Two_plus_a...	0,15	0,3

Airbag (X9)		
VehicleYear	After2...	Before...
Yes	0,8	0,35
No	0,2	0,65

AntilockBrakes		
VehicleYear	After2...	Before...
Yes	0,68	0,13
No	0,32	0,87

DriverQuality								
Age	Young				Old			
ExtraTraining	Yes		No		Yes		No	
Focused	Yes	No	Yes	No	Yes	No	Yes	No
Good	0,336666...	0,196309...	0,267	0,116666...	0,763333...	0,54	0,576666...	0,403333...
Bad	0,663333...	0,803690...	0,733	0,883333...	0,236666...	0,46	0,423333...	0,596666...

Cost						
Accident	Severe		Mild		None	
VehicleYear	After2000	Before2000	After2000	Before2...	After2000	Before2...
Low	0,21	0,34	0,5	0,4	0,91	0,84
High	0,79	0,66	0,5	0,6	0,09	0,16

Accident																
VehicleYear	After2000															
VehicleSize	Compact								SUV							
DriverQuality	Good				Bad				Good				Bad			
AntilockBrakes	Yes		No		Yes		No		Yes		No		Yes		No	
Airbag (X9)	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Severe	0,01	0,0...	0,02	0,04	0,03	0,05	0,06	0,1	0,01	0,03	0,04	0,08	0,04	0,06	0,07	0,1
Mild	0,04	0,0...	0,06	0,1	0,06	0,1	0,12	0,27	0,14	0,17	0,19	0,2	0,17	0,2	0,21	0,29
None	0,95	0,9...	0,92	0,86	0,91	0,85	0,82	0,63	0,85	0,8	0,77	0,72	0,79	0,74	0,72	0,61

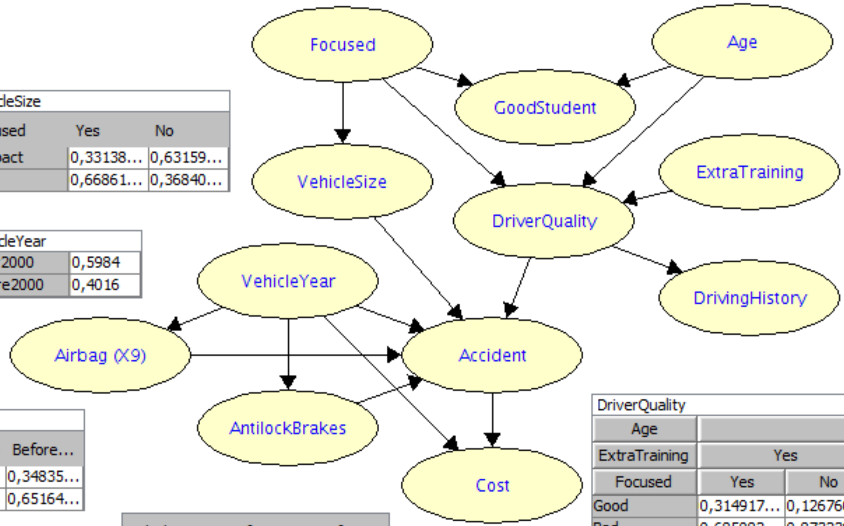
Accident																
VehicleYear	Before2000															
VehicleSize	Compact								SUV							
DriverQuality	Good				Bad				Good				Bad			
AntilockBrakes	Yes		No		Yes		No		Yes		No		Yes		No	
Airbag (X9)	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Severe	0,04	0,06	0,06	0,09	0,06	0,1	0,11	0,17	0,02	0,04	0,05	0,09	0,05	0,08	0,0...	0,17
Mild	0,1	0,15	0,12	0,15	0,13	0,15	0,15	0,2	0,14	0,18	0,15	0,2	0,19	0,22	0,2...	0,33
None	0,86	0,79	0,82	0,76	0,81	0,75	0,74	0,63	0,84	0,78	0,8	0,71	0,76	0,7	0,6...	0,5

Рис. А1. Експертні оцінки параметрів умовних імовірнісних розподілів вершин мережі Insurance

Додаток А

Focused	
Yes	0,7025
No	0,2975

GoodStudent				
Focused	Yes		No	
Age	Young	Old	Young	Old
Yes	0,710005…	0,00894…	0,089385…	0,01150…
No	0,289994…	0,99105…	0,910614…	0,98849…



Age	
Young	0,2485
Old	0,7515

VehicleSize		
Focused	Yes	No
Compact	0,33138…	0,63159…
SUV	0,66861…	0,36840…

ExtraTraining	
Yes	0,1028
No	0,8972

VehicleYear	
After2000	0,5984
Before2000	0,4016

DrivingHistory		
DriverQuality	Good	Bad
No_accidents	0,50150…	0,19742…
One_accident	0,35440…	0,50428…
Two_plus_a…	0,14408…	0,29828…

Airbag (X9)		
VehicleYear	After2000	Before…
Yes	0,79177…	0,34835…
No	0,20822…	0,65164…

DriverQuality								
Age	Young				Old			
ExtraTraining	Yes		No		Yes		No	
Focused	Yes	No	Yes	No	Yes	No	Yes	No
Good	0,314917…	0,126760…	0,272670…	0,114728…	0,736086…	0,547945…	0,579910…	0,396078…
Bad	0,685082…	0,873239…	0,727329…	0,885271…	0,263913…	0,452054…	0,420089…	0,603921…

VehicleYear	After2000	Before…
Yes	0,68432…	0,1319…
No	0,31567…	0,8680…

Cost						
Accident	Severe		Mild		None	
VehicleYear	After2000	Before2000	After2000	Before2…	After2000	Before2…
Low	0,180952…	0,315662…	0,504305…	0,395216…	0,915138…	0,850532…
High	0,819047…	0,684337…	0,495694…	0,604783…	0,084861…	0,149467…

Accident																
VehicleYear	After2000															
VehicleSize	Compact								SUV							
DriverQuality	Good				Bad				Good				Bad			
AntilockBrakes	Yes		No		Yes		No		Yes		No		Yes		No	
Airbag (X9)	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Severe	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,1…
Mild	0,0…	0,075	0,0…	0,1…	0,0…	0,1…	0,1…	0,2…	0,1…	0,1…	0,2…	0,1…	0,1…	0,1…	0,2…	0,3…
None	0,9…	0,9…	0,9…	0,8…	0,9…	0,8…	0,8…	0,6…	0,8…	0,8…	0,7…	0,8…	0,7…	0,7…	0,6…	0,5…

Accident																
VehicleYear	Before2000															
VehicleSize	Compact								SUV							
DriverQuality	Good				Bad				Good				Bad			
AntilockBrakes	Yes		No		Yes		No		Yes		No		Yes		No	
Airbag (X9)	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Severe	0,0…	0,0…	0,0…	0,0…	0,125	0,1…	0,1…	0,1…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,0…	0,1…
Mild	0,1…	0,1…	0,1…	0,1…	0,125	0,1…	0,1…	0,2…	0,1…	0,2…	0,1…	0,2…	0,1…	0,2…	0,2…	0,3…
None	0,7…	0,8…	0,8…	0,7…	0,75	0,7…	0,7…	0,6…	0,8…	0,7…	0,7…	0,6…	0,8…	0,6…	0,6…	0,5…

Рис. А2. Байєсівські оцінки параметрів умовних імовірнісних розподілів вершин мережі Insurance (обсяг навчальної вибірки 10 000)

Додаток Б. Портрети дорожньо-транспортних пригод для визначення страхового відшкодування майнової шкоди

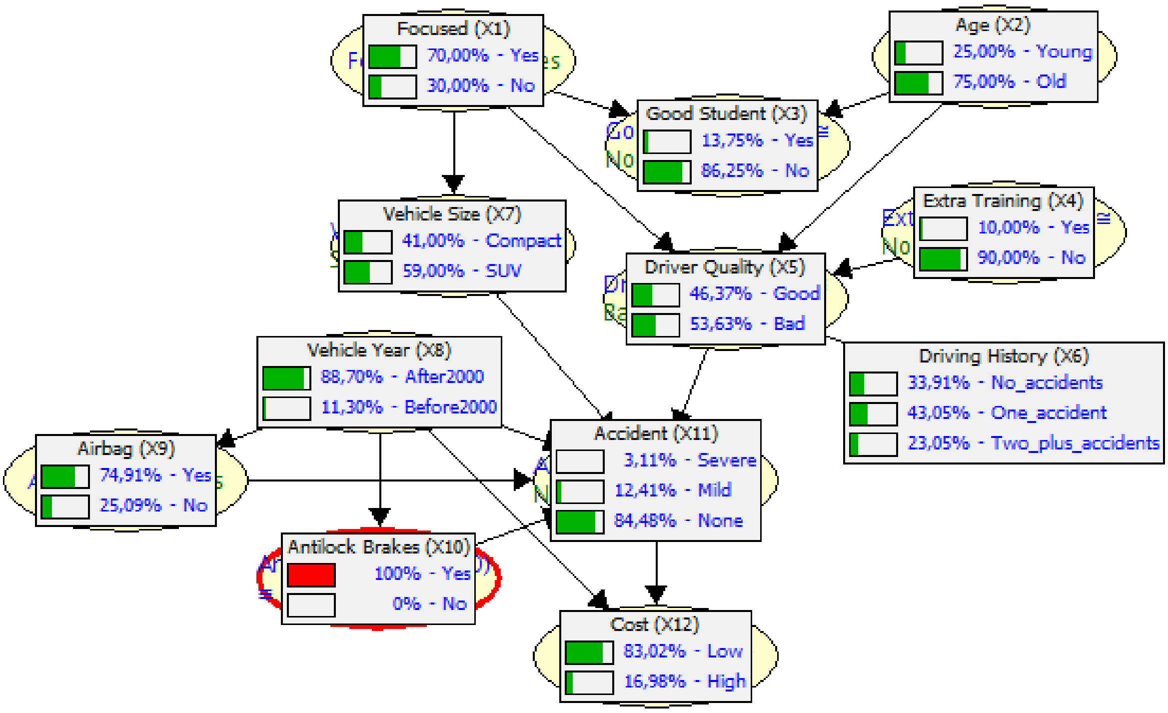


Рис. Б1. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = Yes) = 0{,}1698,\ P(X_{12} = Low | X_{10} = Yes) = 0{,}8302$$

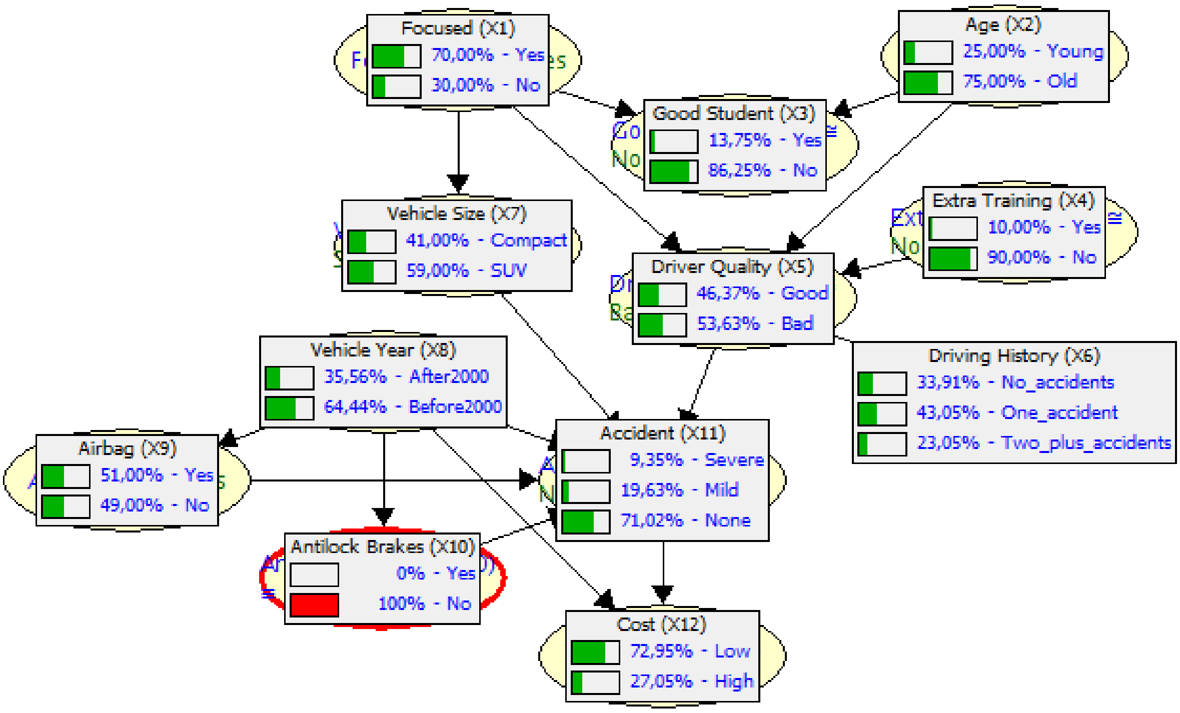


Рис. Б2. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = No) = 0{,}2705,\ P(X_{12} = Low | X_{10} = No) = 0{,}7295$$

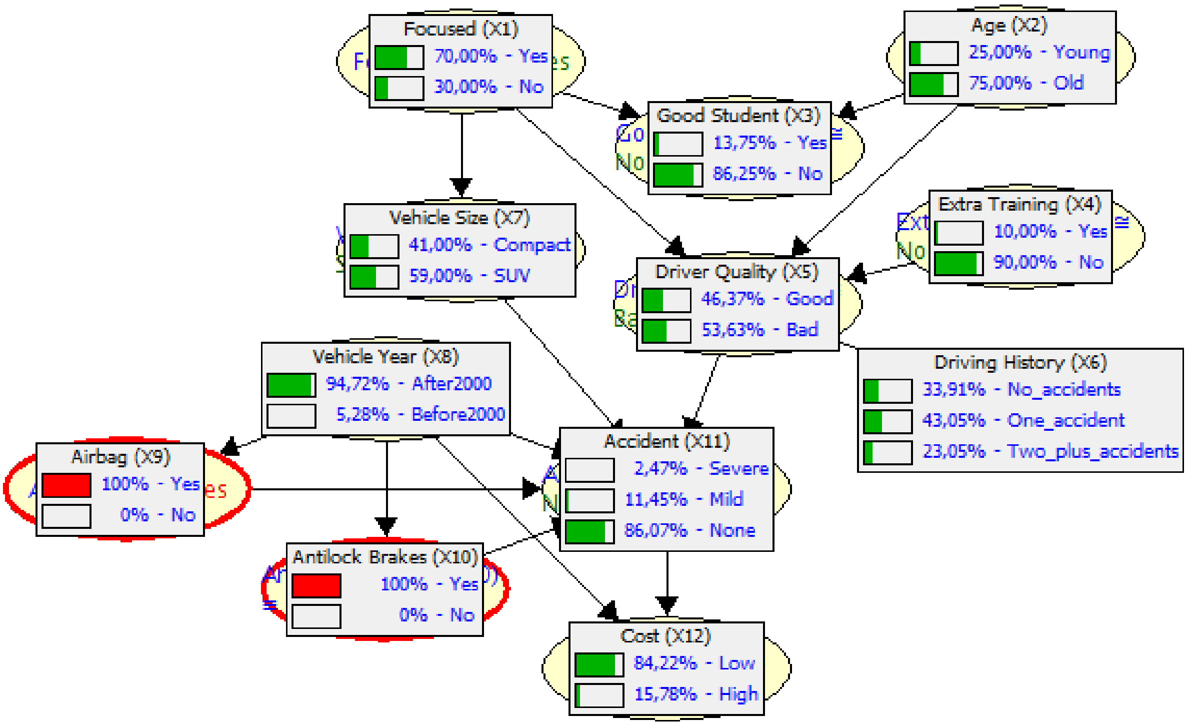


Рис. Б3. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = Yes, X_9 = Yes) = 0{,}1578$$

$$P(X_{12} = Low | X_{10} = Yes, X_9 = Yes) = 0{,}8422$$

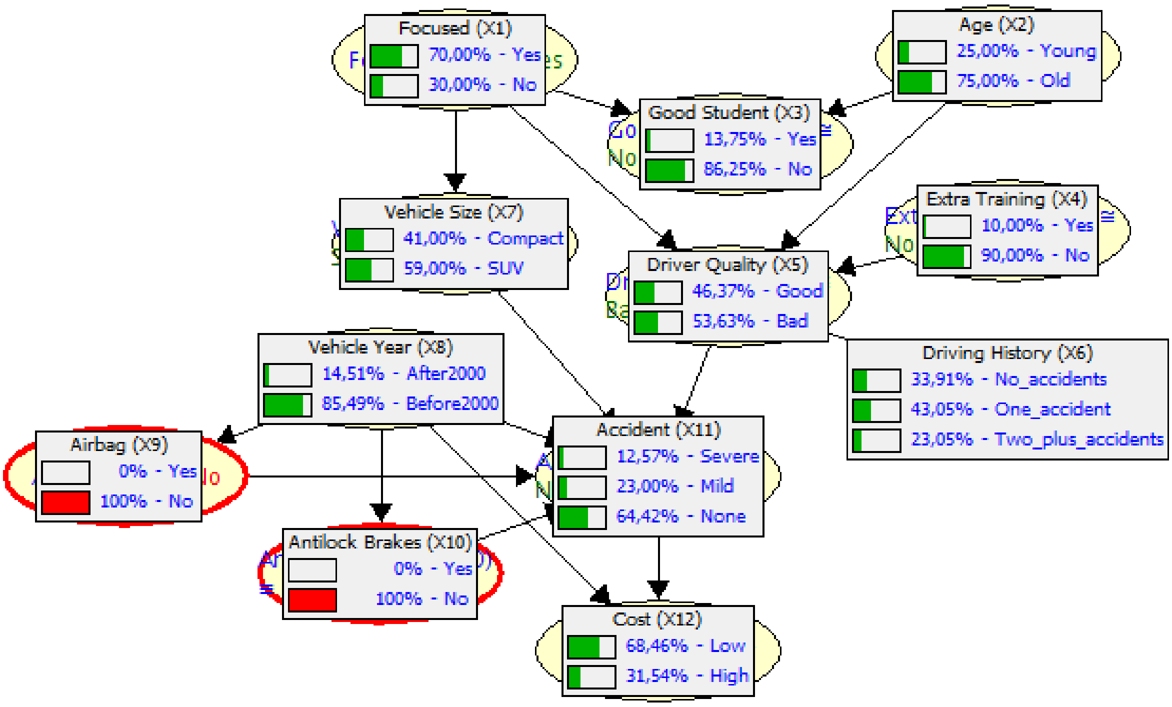


Рис. Б4. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = No, X_9 = No) = 0{,}3154$$

$$P(X_{12} = Low | X_{10} = No, X_9 = No) = 0{,}6846$$

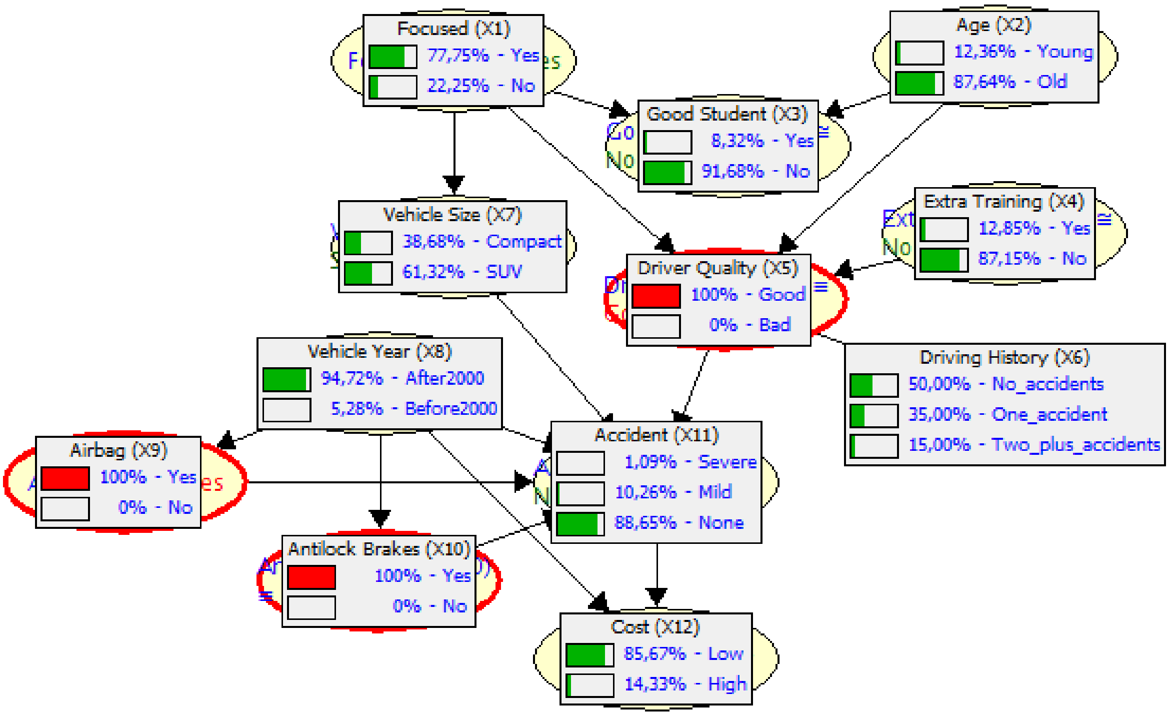


Рис. Б5. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = Yes, X_9 = Yes, X_5 = Good) = 0{,}1433,$$

$$P(X_{12} = Low | X_{10} = Yes, X_9 = Yes, X_5 = Good) = 0{,}8567$$

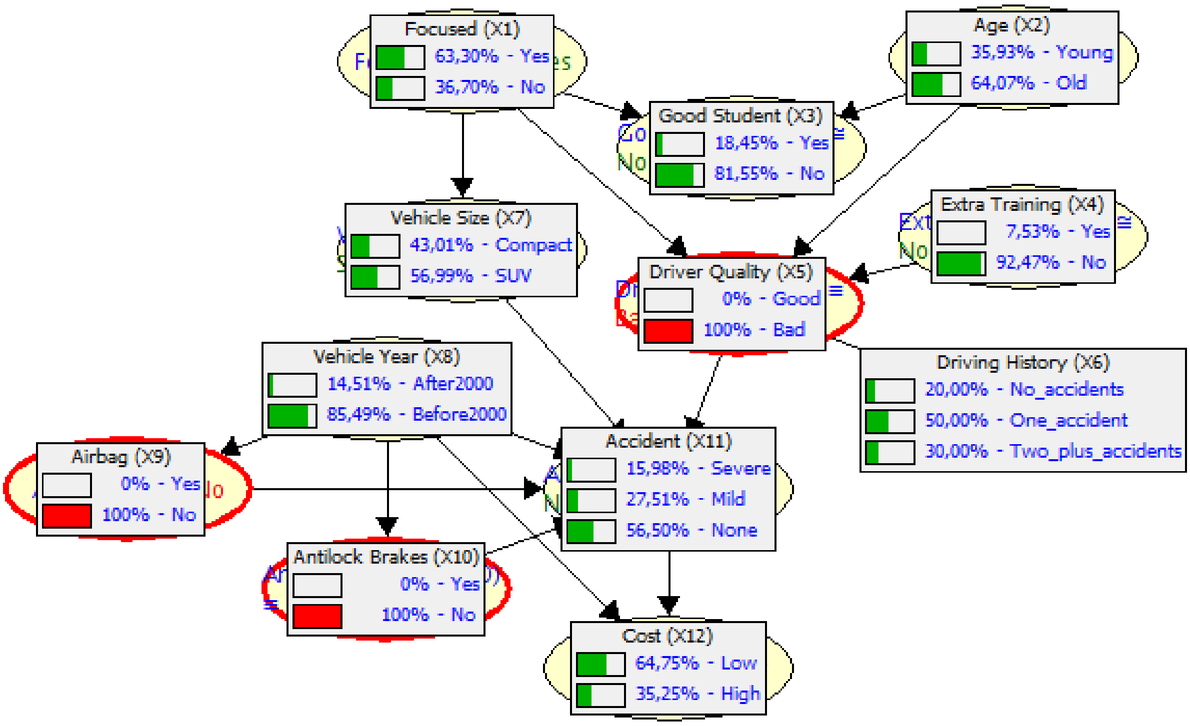


Рис. Б6. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High | X_{10} = No, X_9 = No, X_5 = Bad) = 0{,}3525,$$

$$P(X_{12} = Low | X_{10} = No, X_9 = No, X_5 = Bad) = 0{,}6475$$

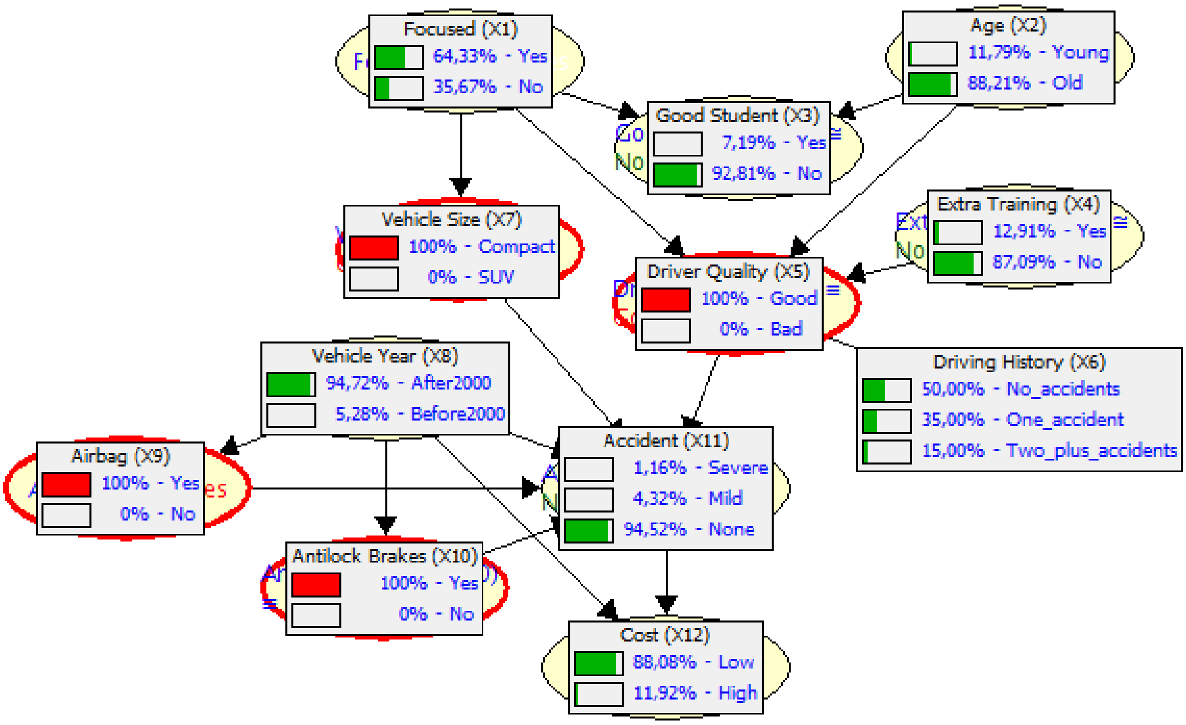


Рис. Б7. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High|X_{10} = Yes, X_9 = Yes, X_5 = Good, X_7 = Compact) = 0{,}1192$$

$$P(X_{12} = Low|X_{10} = Yes, X_9 = Yes, X_5 = Good, X_7 = Compact) = 0{,}8808$$

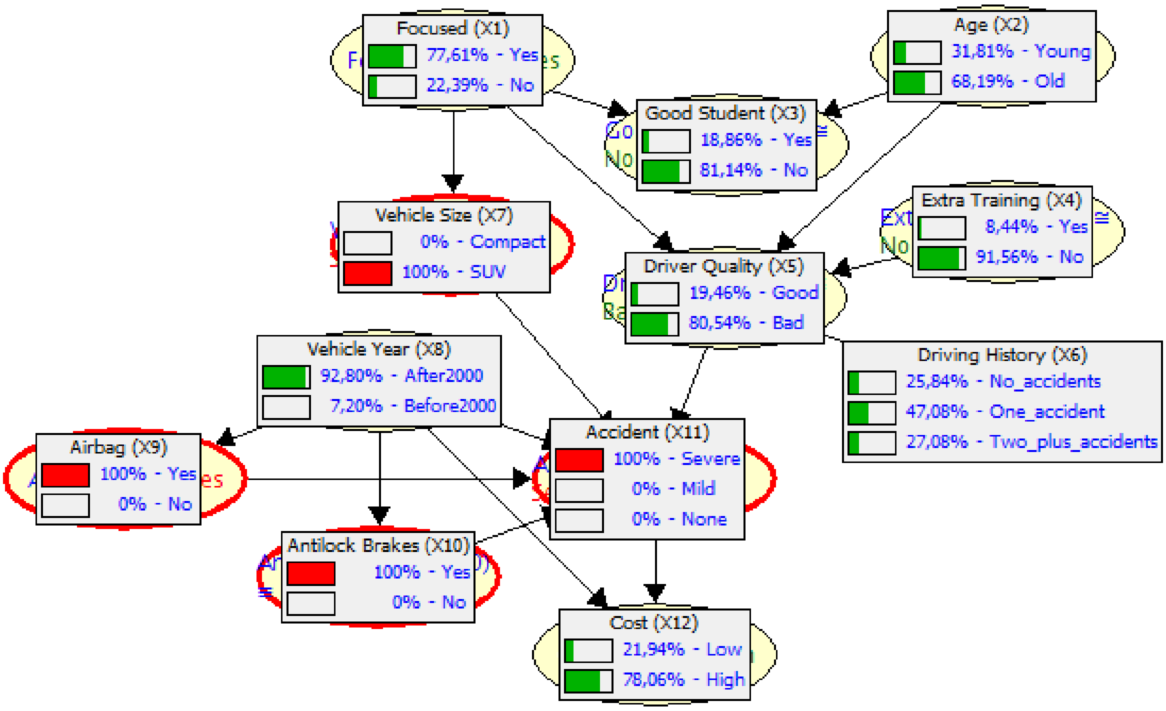


Рис. Б8. Апостеріорний розподіл витрат страхової компанії

$$P(X_{12} = High|X_{10} = Yes, X_9 = Yes, X_7 = SUV, X_{11} = Severe) = 0{,}7806,$$

$$P(X_{12} = Low|X_{10} = Yes, X_9 = Yes, X_7 = SUV, X_{11} = Severe) = 0{,}2194$$

Додаток В

Байєсівські оцінки – спроможні та асимптотично ефективні оцінки параметрів умовних імовірнісних розподілів вершин байєсівської мережі. Спроможність гарантує збіжність оцінки за ймовірністю до справжнього значення параметра зі зростанням обсягу вибірки. Асимптотична ефективність гарантує прямування дисперсії оцінки до нуля зі зростанням обсягу вибірки.

Асимптотична поведінка байєсівських оцінок наведена на рисунку В1. Неперервними лініями позначено байєсівські оцінки параметрів, пунктирними лініями – експертні оцінки параметрів.

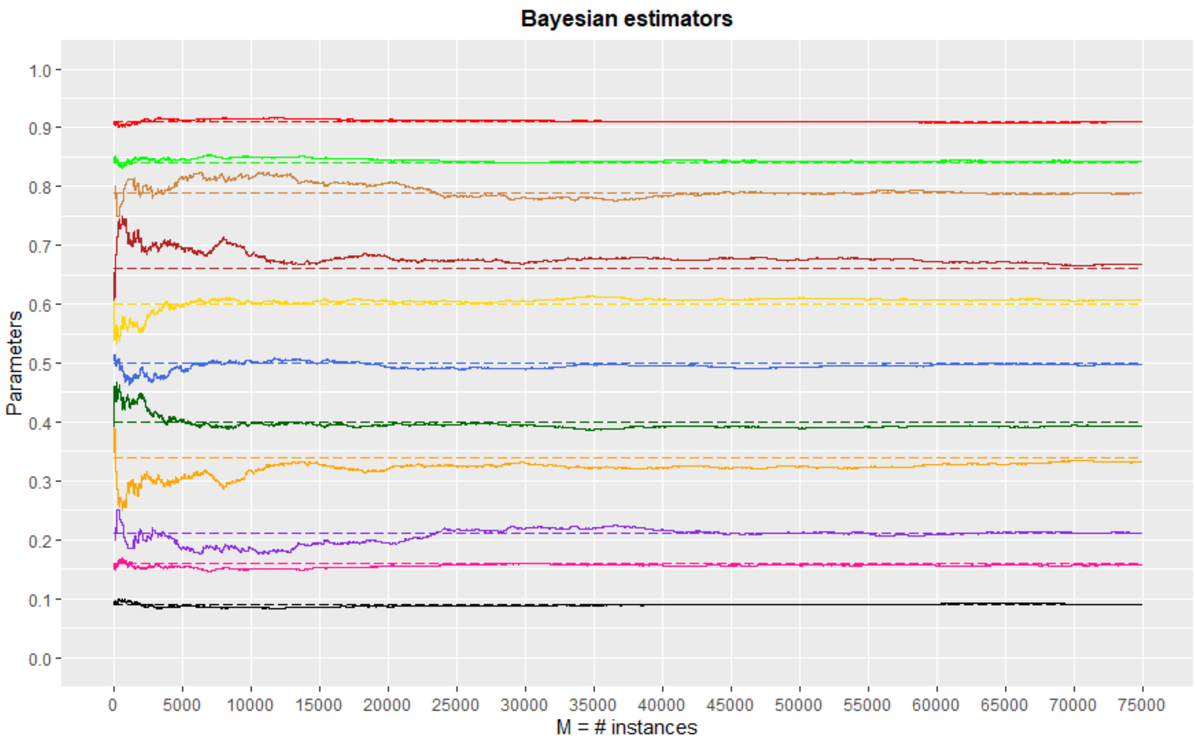


Рис. В1. Асимптотична поведінка байєсівських оцінок параметрів умовних імовірнісних розподілів вершини *Cost*

Зміст