Байесовские сети

Максим Гончаров
maxgon@microsoft.com
maxim.goncharov@spellabs.ru

1. Введение

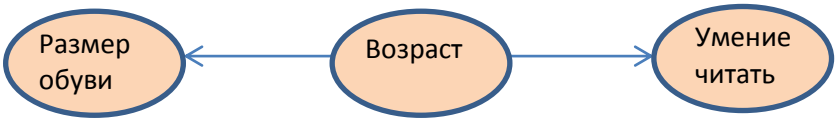
Понятие информационной зависимости между объектами предметной области является естественным для человеческого мышления. Люди имеют тенденцию рассуждать в терминах трех-уровневых связей между факторами: фактор x влияет на y посредством z. Поэтому попытки сконструировать интуитивно-понятную модель предметной области приводят к необходимости использования языка, способного ясно выявлять и формулировать опосредованные зависимости между факторами.

В теории вероятности понятие информационной зависимости моделируется посредством условной зависимости (или строго: отсутствием условной независимости), которая описывает, как наша уверенность в исходе некоего события меняется при получении нового знания о фактах, при условии, что нам был уже известен некоторый набор других фактов.

Удобно и интуитивно понятно представлять зависимости между элементами посредством направленного пути, соединяющего эти элементы в графе. Если зависимость между элементами x и y не является непосредственной и осуществляется посредством третьего элемента z, то логично ожидать, что на пути между x и y будет находиться элемент z. Такие узлы-посредники будут «отсекать» зависимость между x и y, т.е. моделировать ситуацию условной независимости между ними при известном значении непосредственных факторов влияния.

Такими языками моделирования являются байесовские сети, которые служат для описания условных зависимостей между понятиями некой предметной области. Различают два основных сценария применения байесовских сетей:

1. *Описательный анализ*. Предметная область отображается в виде графа, узлы которого представляют понятия, а направленные дуги, отображаемые стрелками, иллюстрируют непосредственные зависимости между этими понятиями. Связь между понятиями x и y означает: знание значения x помогает сделать более обоснованное предположение о значении y. Отсутствие непосредственной связи между понятиями моделирует условную независимость между ними при известных значениях некоторого набора «разделяющих» понятий. Например, размер обуви ребенка, очевидно, связан с умением ребенка читать через возраст. Так, б**о**льший размер обуви дает б**о**льшую уверенность, что ребенок уже читает, но если нам уже известен возраст, то знание размера обуви уже не даст нам дополнительной информации о способности ребенка к чтению.



   В качестве другого, противоположного, примера рассмотрим такие изначально несвязанные факторы как курение и простуда. Но если нам известен симптом, например, что человек страдает по утрам кашлем, то знание того, что человек не курит, повышает нашу уверенность того, что человек простужен.

2. *Классификация и прогнозирование*. Байесовская сеть, допуская условную независимость ряда понятий, позволяет уменьшить число параметров совместного распределения, делая возможным их доверительную оценку на имеющихся объемах данных. Так, при 10

переменных, каждая из которых может принимать 10 значений, число параметров совместного распределения – 10 миллиардов - 1. Если допустить, что между этими переменными друг от друга зависят только 2 переменные, то число параметров становится 8*(10-1) + (10*10-1) = 171. Имея реалистичную по вычислительным ресурсам модель совместного распределения, неизвестное значение какого-либо понятия мы можем прогнозировать как, например, наиболее вероятное значение этого понятия при известных значениях других понятий.

В статье рассмотрены основные свойства байесовских сетей, а также методы оценки их структуры и параметров.

2. Обзор

Сначала мы вводим необходимые понятия из теории графов и получаем нужные в дальнейшем результаты.

Затем мы определяем байесовскую сеть как направленный ациклический граф, узлы которого соответствуют понятиям предметной области, а дуги – непосредственным стохастическим связям между ними. Для вероятностного распределения на множестве значений узлов графа мы определяем гипотезы об условной независимости, факторизации и разделении, соответствующие байесовской сети. Затем мы доказываем эквивалентность этих гипотез, что позволяет далее пользоваться единой гипотезой, накладывающей ограничения на множество параметров вероятностного распределения узлов в сети. Эти ограничения соответствуют условной независимости между узлами, связанными в сети через маршруты, проходящие через разделяющие эти узлы множества.

Затем мы вводим отношение эквивалентности на множестве байесовских сетей: сети эквивалентны, если содержат один и тот же набор высказываний о разделении. Эквивалентным сетям соответствуют одинаковые гипотезы об условной независимости. Мы доказываем два различных необходимых и достаточных условия эквивалентности сетей. Сети эквиваленты тогда и только тогда, когда они имеют одинаковый скелет и v-структуры. Сети эквиваленты тогда и только тогда, когда одну сеть можно получить из другой посредством конечного числа обращений покрытых дуг.

Далее мы получаем оценку максимального правдоподобия для *параметров* сети и показываем, что оценка *структуры* сети методом максимального правдоподобия неприемлема, так как приводит всегда к полному графу (все узлы связаны), что сводит на нет смысл байесовских сетей, как моделей условной независимости.

Затем мы используем байесовский подход (параметры и структура сети – случайные величины с неким априорным распределением) для оценки структуры байесовской сети. Введя ряд ограничений и допущений на априорное распределение параметров и структуры сети, мы получаем формулу для метрики Байеса-Дирихле (BD) в явном виде, которая описывает апостериорную вероятность структуры сети по данным. В качестве частного случая мы получаем метрику K2. Затем мы получаем достаточные условия, при которых BD метрика принимает одинаковые значения на эквивалентных сетях (BDe). Мы получаем формулу для частного случая этой метрики – равномерную эквивалентную метрику Байеса-Дирихле (BDeu). В заключение раздела мы получаем асимптотическую оценку любой байесовской метрики – информационный критерий Байеса (BIC).

Далее мы приводим методы поиска оптимальной с точки зрения выбранной метрики структуры байесовской сети, приводящие к локально-оптимальным результатам, а также метод повторного запуска алгоритма для нахождения глобального максимума метрики.

Статья основана на многочисленных работах Judea Pearl, Thomas Verma, David Heckerman, David Maxwell Chickering, Dan Geiger и других. Часть работ приведена в списке литературы.

3. Определения и обозначения

Под $G=(N,E)$ мы будем обозначать направленный ациклический граф (**НАГ**) с множеством узлов $N=\{x_1,...,x_n\}$ и множеством направленных дуг $E=\{\langle x_i,x_j\rangle | x_i,x_j\in N\}$.

Пусть $Y\subset N$ - некоторое множество узлов, введем следующие обозначения:

$P(Y):=\{x\in N | \exists y\in Y, \exists\langle x,y\rangle\in E\}$ - **множество родителей** Y,

$C(Y):=\{x\in N | \exists y\in Y, \exists\langle y,x\rangle\in E\}$ - **множество детей** Y,

$A_k(Y):=\begin{cases} P(Y), k=1 \\ P(A_{k-1}(Y)), k>1 \end{cases}$ - **множество предков** Y k-ого уровня,

$A(Y):=\bigcup_{i=1}^{n} A_i(Y)$ - **множество предков** Y,

$D_k(Y):=\begin{cases} C(Y), k=1 \\ C(D_{k-1}(Y)), k>1 \end{cases}$ - **множество потомков** Y k-ого уровня,

$D(Y):=\bigcup_{i=1}^{n} D_i(Y)$ - **множество потомков** Y.

Определение: **ребро –** дуга без учета направления.

Определение: **скелет** направленного графа – ненаправленный граф, полученный из направленного заменой дуг на ребра.

Определение: узел *x* и дуга *e* называются **инцидентными**, если *e* входит или исходит из *x.*

Определение: узлы называются **смежными**, если они инциденты одной дуге, т.е. если один из них является родителем другого.

Определение: **маршрут –** это чередующаяся последовательность узлов и дуг, в которой любые два соседних элемента инцидентны.

Определение: два узла называются **связными**, если между ними есть маршрут.

Определение: **путь –** последовательность дуг, такая, что конец одной дуги является началом другой дуги.

Определение: **топологической нумерацией** узлов направленного ациклического графа G называется любая перестановке $\sigma:\{1,...,n\}\to\{1,...,n\}$, ставящая родителям в соответствие номер, меньший, чем детям, т.е. если $x_i\in P(x_j)$, то $\sigma(i)<\sigma(j)$.

Утверждение 1: Для каждого НАГ $G=(N,E)$ существует топологическая нумерация.

Доказательство: Сначала произвольно нумеруем все узлы, у которых нет родителей (такие узлы есть, так как в противном случае существовали бы цепочки бесконечной длины из конечного числа узлов, т.е. в графе были бы циклы). Обозначим множество таким образом пронумерованных узлов $L_1:=\{x\in N | P(\{x\})=\varnothing\}$. Если $N\setminus L_1\neq\varnothing$, то далее произвольно

нумеруем элементы множества $L_2 := \left\{ x \in N \setminus L_1 \middle| P\left(\{x\}\right) \subset L_1 \right\}$ - множество узлов с родителями только из множества L_1. $L_2 \neq \varnothing$, так как в противном случае у всех узлов из $N \setminus L_1$ были бы родители в том же множестве $N \setminus L_1$, а, следовательно, в $N \setminus L_1$ были бы пути бесконечной длина, и, следовательно, циклы. Итеративно, если $N \setminus \bigcup_{i=1}^{k-1} L_i \neq \varnothing$, то на k-ом шаге нумеруем узлы из множества $L_k := \left\{ x \in N \setminus \bigcup_{i=1}^{k-1} L_i \middle| P\left(\{x\}\right) \subset \bigcup_{i=1}^{k-1} L_i \right\}$, т.е. множество всех узлов с родителями только из уже пронумерованных множеств $L_1, \ldots, L_{k-1}$. Множество $L_k \neq \varnothing$, так как в противном случае у всех узлов из $N \setminus \bigcup_{i=1}^{k-1} L_i$ существовали бы родители из того же множества $N \setminus \bigcup_{i=1}^{k-1} L_i$, что означало бы наличие циклов в $N \setminus \bigcup_{i=1}^{k-1} L_i$. Таким образом, через p <= n шагов мы получаем набор попарно непересекающихся множеств $L_1, \ldots, L_p$, такой что $N = \bigcup_{i=1}^{p} L_i$. Приведенный алгоритм гарантировано нумерует сначала родителей, а затем – детей. Действительно, пусть узел $x_i \in L_k$ – родитель узла $x_j \in L_m$. По построению, все родители узлов из L_m, в том числе все родителя узла x_j находятся в $\bigcup_{i=1}^{m-1} L_i$, следовательно, $x_i \in \bigcup_{i=1}^{m-1} L_i$, значит $L_k \subset \bigcup_{i=1}^{m-1} L_i$, т.е. k < m. Это означает, что сначала были пронумерованы все узлы из L_k, а затем – все узлы из L_m, т.е. $\sigma(i) < \sigma(j)$.

□

Утверждение 2: Для каждого НАГ $G = (N, E)$, узла $x \in N$ и множества узлов $Y = \{y_1, \ldots, y_k\} \subset N$, не содержащего x и потомков x, т.е. и $x \notin Y, Y \cap D\left(\{x\}\right) = \varnothing$ существует топологическая нумерация узлов, такая, что номера из множества Y имеют меньшее значения, чем номер узла x, т.е. $\sigma(y_i) < \sigma(x), \forall y_i \in Y$.

Доказательство:

Этап 1.

Построим последовательно множества предков всех уровней для узлов из Y: $A_1(Y), \ldots, A_n(Y)$ (эти множества могут иметь непустое пересечение, так как у каждого узла может быть несколько родителей и поэтому несколько путей различной длины от одного узла к другому). Так как в G нет циклов, то существует $p \leq n$, такое, что множество потомков p+1-ого уровня – пустое, т.е. $A_{p+1}(Y) = \ldots = A_n(Y) = \varnothing$.

Произвольно пронумеруем сначала узлы, входящие в $A_p(Y)$, затем – непронумерованные узлы, входящие в $A_{p-1}(Y)$ и так далее до $A_1(Y)$ и Y. Таким образом, если узел входит в несколько множеств $A_s(Y)$ и $A_l(Y)$, s < l, то он нумеруется в рамках нумерации узлов из множества с б**о**льшим индексом l. Другими словами каждый узел нумеруется в порядке длины максимального пути от узла к множеству Y. При таком способе нумерации номер каждого ребенка будет больше, чем номер родителя. Действительно, пусть пронумерованный узел w – родитель пронумерованного узла v. Тогда длина максимального пути от w до Y будет больше длины максимального пути от узла v до Y. Это означает, что w будет нумероваться раньше, чем v, т.е. $\sigma(w) < \sigma(v)$.

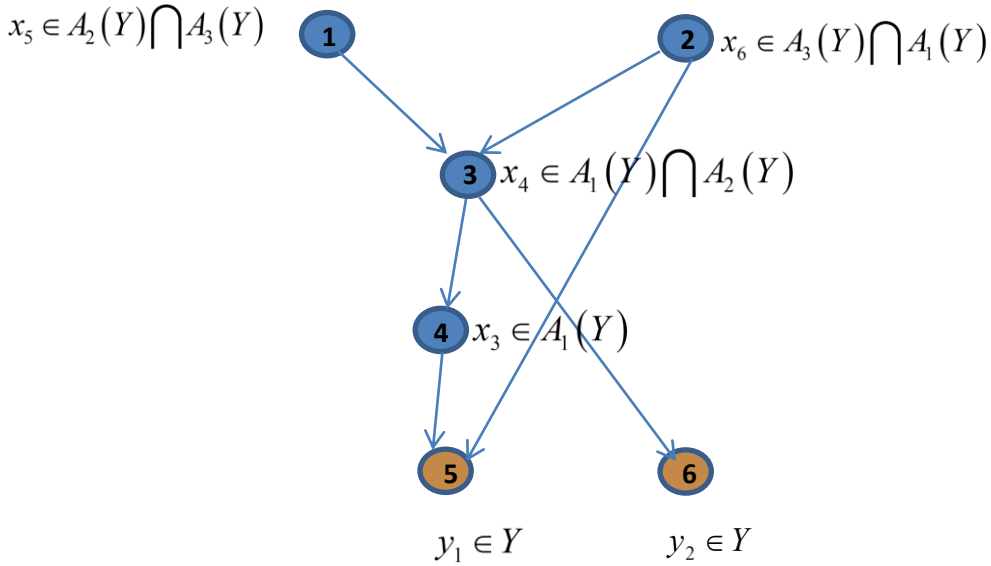


Рис. 1 – *Иллюстрация топологической нумерации узлов - 1*

Таким образом, мы осуществили топологическую нумерацию узлов из $A(Y)$. Среди предков Y не может быть узла x, так как в противном случае хотя бы один из узлов из Y был бы потомком x. Поэтому x остается на данном этапе пока непронумерованным.

Этап 2.

Далее мы нумеруем все непронумерованные узлы по алгоритму, изложенному в утверждении 1: сначала на первом шаге нумеруем все непронумерованные узлы из первого уровня графа – узлы без родителей. Затем, итеративно на i-ом шаге нумеруем непронумерованные узлы, у которых родители находятся только среди уже пронумерованных узлов, принадлежащих меньшим уровням графа (т.е. родители которых принадлежат уровням с 1 по i-1). Таким образом, мы двигаемся по графу «вниз» и не можем «столкнуться» с тем, что мы нумеруем некий узел t, у которого есть уже пронумерованный с меньшим номером ребенок s. Потому что в этом случае s был бы пронумерован на этапе 1 нумерации предков Y, т.е. s был бы предком Y: $s \in A(Y)$, а, следовательно, t также был бы предком Y: $t \in A(Y)$, т.е. t был бы уже пронумерован на этапе 1. Таким образом, мы получаем топологическую нумерацию узлов графа, при которой номер узла x больше, чем все номера из множества Y.

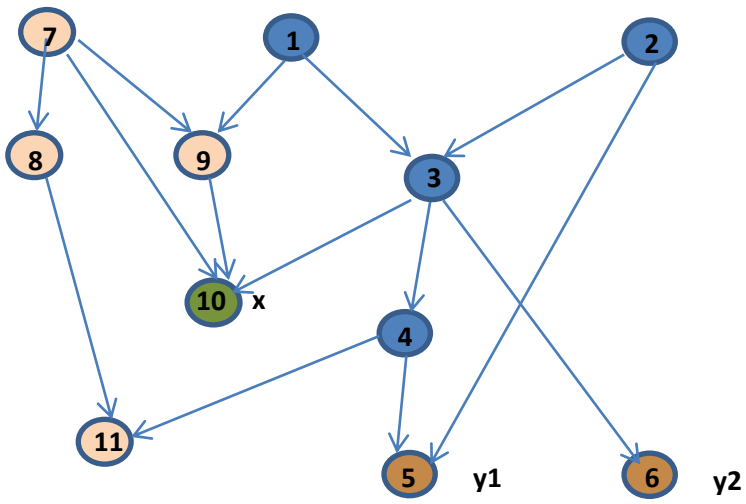


Рис. 2– *Иллюстрация топологической нумерации узлов - 2*

□

Определение:
Пусть Ω - пространство элементарных событий и A $-\sigma$-алгебра на нем. Пусть $\wp\left(\{1,...,r_i\}\right)-\sigma$-алгебра на $\{1,...,r_i\}$, представляющая собой множество всех подмножеств из $\{1,...,r_i\}$.

Далее, пусть $X_i:(\Omega,\mathrm{A})\to\left(\{1,...,r_i\},\wp\left(\{1,...,r_i\}\right)\right);i=1,...,n$ – семейство измеримых функций (случайных величин), определенных на множестве событий Ω и со значениями в $Val(X_i):=\{1,...,r_i\}$.

4. Байесовская сеть

Определение: **байесовской сетью** (БС) мы будем называть любой направленный ациклический граф G(N,E) с узлами, соответствующими множеству случайных величин $X_1,...,X_n$, т.е. $N=\{X_1,...,X_n\}$.

Определим параметрическое множество $r_1...r_n$-мерных векторов:

$$\Theta:=\left\{\theta=\left(\theta_{k_1,...,k_n}\middle|k_i\in\{1,...,r_i\}\right)\in\mathbb{R}^{r_1...r_n}\middle|\theta_{k_1,...,k_n}>0,\sum_{k_1=1}^{r_1}...\sum_{k_n=1}^{r_n}\theta_{k_1,...,k_n}=1\right\}$$

Для каждого $\theta\in\Theta$ определим распределение вероятности на множестве $\{1,...,r_1\}\times...\times\{1,...,r_n\}$ как

$$\mathrm{P}_\theta\left(X_1=k_1,...,X_n=k_n\right):=\theta_{k_1,...,k_n}$$

Без каких-либо дополнительных предположений, размерность θ равняется $\#\theta=\prod_{i=1}^{n}r_i-1$.

4.1 Гипотеза об условной независимости

Определение: **гипотеза об условной независимости**, соответствующая байесовской сети, представляет собой предположение, что каждый узел x при известных значениях родителей x условно не зависит от любого множества Y, не содержащего x и потомков x.

Гипотезу об условной независимости мы будем обозначать G_{ind}, а множество параметров, удовлетворяющее этой гипотезе - $\Theta_{G_{ind}}\subset\Theta$. Таким образом, гипотеза об условной независимости формально имеет вид:

$$\begin{gathered}\mathrm{P}_\theta\left(X_i,Y\middle|P(X_i)\right)=\mathrm{P}_\theta\left(X_i\middle|P(X_i)\right)\mathrm{P}_\theta\left(Y\middle|P(X_i)\right);\\ \forall\theta\in\Theta_{G_{ind}},\forall Y\subset\{X_1,...,X_n\}\setminus\{X_i\},Y\bigcap D(X_i)=\varnothing\end{gathered}\qquad\textbf{(1)}$$

Утверждение 3: Если справедлива гипотеза об условной независимости, то для любой топологической нумерации узлов справедливо:

$$\mathrm{P}_\theta\left(X_i\middle|X_1,...,X_{i-1}\right)=\mathrm{P}_\theta\left(X_i\middle|P(X_i)\right),\forall\theta\in\Theta_{G_{ind}},\forall i\in\{1,...,n\}\qquad\textbf{(2)}$$

Доказательство:
Из определения топологической нумерации следует, что номер ребенка всегда больше номеров его родителей, поэтому i больше любого номера узла из $P(X_i)$, следовательно

$P(X_i) \subset \{X_1,...,X_{i-1}\}$. Далее, множество $Y := \{X_1,...,X_{i-1}\} \setminus P(X_i)$ не может содержать потомков X_i, так как потомки имеют большие номера, чем i, следовательно, $Y \bigcap D(X_i) = \varnothing$. Получаем:

$$\mathrm{P}_\theta\left(X_i \middle| X_1,...,X_{i-1}\right) = \mathrm{P}_\theta\left(X_i \middle| P(X_i), Y\right) = \frac{\mathrm{P}_\theta\left(X_i, Y \middle| P(X_i)\right)}{\mathrm{P}_\theta\left(Y \middle| P(X_i)\right)}$$

С учетом гипотезы об условной независимости (1):

$$\mathrm{P}_\theta\left(X_i \middle| X_1,...,X_{i-1}\right) = \frac{\mathrm{P}_\theta\left(X_i \middle| P(X_i)\right)\mathrm{P}_\theta\left(Y \middle| P(X_i)\right)}{\mathrm{P}_\theta\left(Y \middle| P(X_i)\right)} = \mathrm{P}_\theta\left(X_i \middle| P(X_i)\right)$$

□

4.2 Гипотеза о факторизации

Определение: **гипотеза о факторизации**, соответствующая байесовской сети, представляет собой предположение, что совместная вероятность есть произведение условных вероятностей каждого узла при известных значениях родителей:

$$\mathrm{P}_\theta(X_1,...,X_n) = \prod_{i=1}^{n} \mathrm{P}_\theta\left(X_i \middle| P(X_i)\right), \forall \theta \in \Theta_{G_{fact}} \qquad \textbf{(3)}$$

Гипотезу факторизации будем обозначать G_{fact}, а множество параметров, удовлетворяющее этой гипотезе - $\Theta_{G_{fact}} \subset \Theta$.

Утверждение 4: Из гипотезы об условной независимости следует гипотеза о факторизации, т.е. $G_{ind} \Rightarrow G_{fact}$

Доказательство:
Пусть выполняется гипотеза об условной независимости, т.е. $\theta \in \Theta_{G_{ind}}$

Тогда из утверждения 3 следует, что для любой топологической нумерации узлов справедливо:
$\mathrm{P}_\theta\left(X_i \middle| X_1,...,X_{i-1}\right) = \mathrm{P}_\theta\left(X_i \middle| P(X_i)\right), \forall i \in \{1,...,n\}$

Из определения условной плотности вероятности для любой нумерации узлов справедливо:

$$\mathrm{P}_\theta(X_1,...,X_n) = \prod_{i=1}^{n} \mathrm{P}_\theta\left(X_i \middle| X_1,...,X_{i-1}\right)$$

Комбинируя эти два равенства получаем:

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

Т.е. выполняется гипотеза о факторизации, т.е. $\theta \in \Theta_{G_{fact}}$.

□

Утверждение 5: Из гипотезы о факторизации для любой топологической нумерации узлов следует равенство (2):
$\mathrm{P}_\theta\left(X_i \middle| X_1,...,X_{i-1}\right) = \mathrm{P}_\theta\left(X_i \middle| P(X_i)\right), \forall \theta \in \Theta_{G_{ind}}, \forall i \in \{1,...,n\}$

Доказательство:
Пронумеруем узлы сети согласно некой топологической нумерации. Из определения условной вероятности для любой нумерации узлов следует:

$$\mathrm{P}_\theta\left(X_1,\ldots,X_n\right)=\prod_{i=1}^{n}\mathrm{P}_\theta\left(X_i\middle|X_1,\ldots,X_{i-1}\right) \qquad \textbf{(4)}$$

Так как $\theta\in\Theta_{G_{fact}}$, то

$$\mathrm{P}_\theta\left(X_1,\ldots,X_n\right)=\prod_{i=1}^{n}\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right)\right) \qquad \textbf{(5)}$$

Так как нумерация топологическая, то у первого узла не может быть родителей, т.е. $P\left(X_1\right)=\varnothing$ и, следовательно, $\mathrm{P}_\theta\left(X_1\right)=\mathrm{P}_\theta\left(X_1\middle|P\left(X_1\right)\right)$ **(6)**

Далее, так как узлы i+1,…,n не могут входить в множество родителей i-ого узла, то суммируя (4) и (5) по всем значениям узлов с 3 по n-ый, получаем соответственно:

$$\begin{aligned}
&\mathrm{P}_\theta\left(X_1,X_2\right)=\sum_{X_3}\ldots\sum_{X_n}\mathrm{P}_\theta\left(X_1,\ldots,X_n\right)\underset{(4)}{=}\sum_{X_3}\ldots\sum_{X_n}\prod_{i=1}^{n}\mathrm{P}_\theta\left(X_i\middle|X_1,\ldots,X_{i-1}\right)=\\
&=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|X_2,X_1\right)\ldots\underbrace{\sum_{X_n}\mathrm{P}_\theta\left(X_n\middle|X_1,\ldots,X_{n-1}\right)}_{=1}=\\
&=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|X_2,X_1\right)\ldots\underbrace{\sum_{X_{n-1}}\mathrm{P}_\theta\left(X_{n-1}\middle|X_1,\ldots,X_{n-2}\right)}_{=1}=\\
&=\ldots=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)\underbrace{\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|X_2,X_1\right)}_{=1}=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)
\end{aligned} \qquad \textbf{(7)}$$

$$\begin{aligned}
&\mathrm{P}_\theta\left(X_1,X_2\right)=\sum_{X_3}\ldots\sum_{X_n}\mathrm{P}_\theta\left(X_1,\ldots,X_n\right)\underset{(5)}{=}\sum_{X_3}\ldots\sum_{X_n}\prod_{i=1}^{n}\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right)\right)=\\
&=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|\underbrace{P\left(X_2\right)}_{X_3,\ldots,X_n\notin P\left(X_2\right)}\right)\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|\underbrace{P\left(X_3\right)}_{X_4,\ldots,X_n\notin P\left(X_3\right)}\right)\ldots\underbrace{\sum_{X_n}\mathrm{P}_\theta\left(X_n\middle|P\left(X_n\right)\right)}_{=1}=\\
&=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|P\left(X_3\right)\right)\ldots\underbrace{\sum_{X_{n-1}}\mathrm{P}_\theta\left(X_{n-1}\middle|P\left(X_{n-1}\right)\right)}_{=1}=\\
&=\ldots=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)\underbrace{\sum_{X_3}\mathrm{P}_\theta\left(X_3\middle|P\left(X_3\right)\right)}_{=1}=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)
\end{aligned} \qquad \textbf{(8)}$$

Из (7) и (8) получаем:

$$\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)$$

Следовательно: $\mathrm{P}_\theta\left(X_2\middle|X_1\right)=\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)$ **(9)**

Аналогично, суммируя (4) и (5) по всем значениям узлов с 4-ого по n-ый, получаем:

$$\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|X_1\right)\mathrm{P}_\theta\left(X_3\middle|X_1,X_2\right)=\mathrm{P}_\theta\left(X_1\right)\mathrm{P}_\theta\left(X_2\middle|P\left(X_2\right)\right)\mathrm{P}_\theta\left(X_3\middle|P\left(X_3\right)\right)$$

Следовательно, с учетом (9) получаем

$$\mathrm{P}_\theta\left(X_3\middle|X_1,X_2\right)=\mathrm{P}_\theta\left(X_3\middle|P\left(X_3\right)\right) \qquad \textbf{(10)}$$

Продолжая итеративно эту операцию, получаем при суммировании равенств (4) и (5) по переменным с i+1-ой по n-ую:

$$\mathrm{P}_\theta\left(X_i\middle|X_1,\ldots,X_{i-1}\right)=\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right)\right)$$

□

Утверждение 6: Из гипотезы о факторизации следует гипотеза об условной независимости, т.е. $G_{fact}\Rightarrow G_{ind}$

Доказательство:

Пусть выполняется гипотеза о факторизации, т.е. $\theta\in\Theta_{G_{ind}}$

Пусть Y-некое множество узлов, не содержащее узла X и его потомков. Нам надо показать, что X условно не зависит от Y при известном значении родителей X.

Согласно утверждению 2 существует топологическая нумерация узлов, при которой номера всех элементов из Y будут меньше номера узла $X=X_i$. Следовательно $Y\subset\left\{X_1,\ldots,X_{i-1}\right\}$. Так как нумерация топологическая, то $P\left(X_i\right)\subset\left\{X_1,\ldots,X_{i-1}\right\}$.

Обозначим $Z:=\left\{X_1,\ldots,X_{i-1}\right\}\setminus\left(P\left(X_i\right)\bigcup Y\right)$.

Из утверждения 5 следует, что для любой топологической нумерации узлов справедливо:

$\mathrm{P}_\theta\left(X_i\middle|X_1,\ldots,X_{i-1}\right)=\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right)\right)$, следовательно,

$$\mathrm{P}_\theta\left(X_1,\ldots,X_i\right)\mathrm{P}_\theta\left(P\left(X_i\right)\right)=\mathrm{P}_\theta\left(X_i,P\left(X_i\right)\right)\mathrm{P}_\theta\left(X_1,\ldots,X_{i-1}\right) \qquad \textbf{(11)}$$

Суммируя (11) по всем значениям переменных из Z, получаем:

$$\mathrm{P}_\theta\left(P\left(X_i\right),Y,X_i\right)\mathrm{P}_\theta\left(P\left(X_i\right)\right)=\mathrm{P}_\theta\left(X_i,P\left(X_i\right)\right)\mathrm{P}_\theta\left(P\left(X_i\right),Y\right)\Rightarrow$$

$$\Rightarrow\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right),Y\right)=\mathrm{P}_\theta\left(X_i\middle|P\left(X_i\right)\right)$$

Т.е. выполняется гипотеза об условной независимости, т.е. $\theta\in\Theta_{G_{ind}}$.

□

Из утверждений 4 и 6 непосредственно следует:

Утверждение 7: Гипотеза об условной независимости и гипотеза о факторизации, соответствующие одной байесовской сети, совпадают, т.е. $G_{fact}\Leftrightarrow G_{ind}$ или $\Theta_{G_{ind}}=\Theta_{G_{fact}}$.

□

4.3 Гипотеза о разделении

Определение: для любых попарно непересекающихся множеств $X,Y,Z\subset N$ мы будем называть **утверждением об условной независимости, соответствующим гипотезе G**, выражение $I_G\left(X,Y\middle|Z\right)$ в том случае, если $\mathrm{P}_\theta\left(X,Y,Z\right)\mathrm{P}_\theta\left(Z\right)=\mathrm{P}_\theta\left(X,Z\right)\mathrm{P}_\theta\left(Y,Z\right),\forall\theta\in\Theta_G$ **(12)**

Здесь $\Theta_G \subset \Theta$ - множество параметров, соответствующих ограничениям, налагаемых гипотезой G.

Так как вероятностная мера по нашему условию строго положительная, т.е. $\mathrm{P}_\theta(Z) > 0$ для всех Z, то (12) эквивалентно (13):

$$I_G(X,Y|Z) \Leftrightarrow \mathrm{P}_\theta(X,Y|Z) = \mathrm{P}_\theta(X|Z)\mathrm{P}_\theta(Y|Z), \forall \theta \in \Theta_G \qquad \textbf{(13)}$$

Определение: для любых попарно непересекающихся множеств $X,Y \subset N$ мы будем называть **утверждением о безусловной независимости, соответствующим гипотезе G**, выражение $I_G(X,Y|\varnothing)$ в том случае, если $\mathrm{P}_\theta(X,Y) = \mathrm{P}_\theta(X)\mathrm{P}_\theta(Y), \forall \theta \in \Theta_G$ **(14)**

Утверждение о безусловной независимости является, очевидно, частным случаем утверждения об условной независимости при дополнительном условии $Z = \varnothing$.

Непосредственно из свойств условной независимости получаем:

Утверждение 8:
Для попарно непересекающихся множеств X, Y, Z, W следует:

1. Тривиальная независимость:
$$I_G(X,\varnothing|Z) \qquad \textbf{(15)}$$
2. Симметрия:
$$I_G(X,Y|Z) \Leftrightarrow I_G(Y,X|Z) \qquad \textbf{(16)}$$
3. Разложение:
$$I_G(X,Y\bigcup W|Z) \Rightarrow I_G(X,Y|Z) \wedge I_G(X,W|Z) \qquad \textbf{(17)}$$
4. Слабое объединение:
$$I_G(X,Y\bigcup W|Z) \Rightarrow I_G(X,Y|Z\bigcup W) \wedge I_G(X,W|Z\bigcup Y) \qquad \textbf{(18)}$$
5. Сжатие:
$$I_G(X,Y|Z) \wedge I_G(X,W|Z\bigcup Y) \Rightarrow I_G(X,Y\bigcup W|Z) \qquad \textbf{(19)}$$

Для строго положительных вероятностных мер справедливо также

6. Пересечение:
$$I_G(X,Y|Z\bigcup W) \wedge I_G(X,W|Z\bigcup Y) \Rightarrow I_G(X,Y\bigcup W|Z) \qquad \textbf{(20)}$$

Доказательство:
Докажем для примера равенство (19):

$$I_G(X,Y|Z) \Rightarrow \mathrm{P}_\theta(X,Y,Z)\mathrm{P}_\theta(Z) = \mathrm{P}_\theta(X,Z)\mathrm{P}_\theta(Y,Z) \qquad (*)$$

$$I_G(X,W|Z\bigcup Y) \Rightarrow \mathrm{P}_\theta(X,W,Z\bigcup Y)\mathrm{P}_\theta(Z\bigcup Y) = \mathrm{P}_\theta(X,Z\bigcup Y)\mathrm{P}_\theta(W,Z\bigcup Y) \Rightarrow$$
$$\Rightarrow \mathrm{P}_\theta(X,W,Z,Y)\mathrm{P}_\theta(Z,Y) = \mathrm{P}_\theta(X,Z,Y)\mathrm{P}_\theta(W,Z,Y) \qquad (**)$$

Получаем:

$$\mathrm{P}_\theta\left(X,Y\bigcup W,Z\right)\mathrm{P}_\theta(Z)=\mathrm{P}_\theta(X,Y,W,Z)\mathrm{P}_\theta(Z)\underset{(**)}{=}\frac{\mathrm{P}_\theta(X,Z,Y)\mathrm{P}_\theta(W,Z,Y)}{\mathrm{P}_\theta(Z,Y)}\mathrm{P}_\theta(Z)=$$

$$=\mathrm{P}_\theta(X,Y,Z)\mathrm{P}_\theta(Z)\frac{\mathrm{P}_\theta(W,Z,Y)}{\mathrm{P}_\theta(Z,Y)}\underset{(*)}{=}\mathrm{P}_\theta(X,Z)\mathrm{P}_\theta(Y,Z)\frac{\mathrm{P}_\theta(W,Z,Y)}{\mathrm{P}_\theta(Z,Y)}=$$

$$=\mathrm{P}_\theta(X,Z)\mathrm{P}_\theta(W,Z,Y)=\mathrm{P}_\theta(X,Z)\mathrm{P}_\theta\left(Z,Y\bigcup W\right)$$

Т.е. $\mathrm{P}_\theta\left(X,Y\bigcup W,Z\right)\mathrm{P}_\theta(Z)=\mathrm{P}_\theta(X,Z)\mathrm{P}_\theta\left(Y\bigcup W,Z\right)\Rightarrow I_G\left(X,Y\bigcup W|Z\right)$

□

Мы введем еще одну гипотезу, соответствующую байесовской сети. Для этого нам понадобится ряд определений.

Пусть задан маршрут (последовательность инцидентных друг другу узлов): $R=\langle X,Z_1,\ldots,Z_k,Y\rangle$.

Определение: мы будем говорить, что **узел** Z_i **входит в маршрут** R **с типом 1**, если хотя бы одна из двух инцидентных этому узлу в маршруте дуг выходит из него. Мы будем обозначать вхождение с типом 1: $Z\in_1 R$.

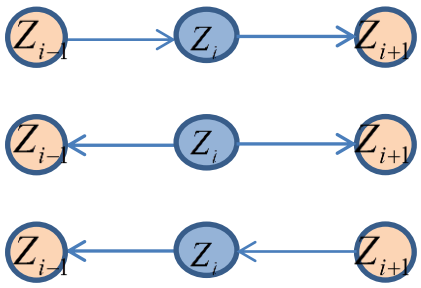


Рис. 3 – *Три варианта соединения узла, входящего в маршрут с типом 1*.

Определение: мы будем говорить, что **узел** Z_i **входит в маршрут** R **с типом 2**, если обе дуги, инцидентные этому узлу в маршруте, входят из него. Мы будем обозначать вхождение с типом 2: $Z\in_2 R$.

Рис. 4 – *Единственный вариант соединения узла, входящего в маршрут с типом 2*.

Принципиальное различие между соединениями с типами 1 и 2 заключается в следующем. Для байесовской сети, состоящей из трех узлов, соединенных как на рис. 3, крайние узлы Z_{i-1} и Z_{i+1} могут быть зависимы друг от друга, но согласно гипотезе об условной независимости, становятся независимыми, если становится известным значение узла Z_i. Действительно, например, для самого верхнего графика на рис. 3, узел Z_{i+1} не зависит от своего не-потомка (узла Z_{i-1}) при известном значении родителя (узла Z_i). В качестве примера можно рассмотреть Z_{i-1} – инвестиции в здравоохранение, Z_i – число обследований, Z_{i+1} – число выявленных заболеваний. Очевидно, Z_{i-1} и Z_{i+1} зависимы через Z_i, т.е. чем больше инвестиций, тем больше обследований, и тем больше выявленных заболеваний. Но если нам уже известно число проведенных обследований, то число выявленных заболеваний от инвестиций, произведенных в эти обследования, уже не зависит. Аналогичные утверждения следуют и для остальных двух графиков на рис. 3. Совершенно другая ситуация соответствует соединению с типом 2, изображенном на рис. 4. В этом случае узлы Z_{i-1} и Z_{i+1} не зависят друг от друга (узел Z_{i-1} не зависит от своего не-потомка – узла Z_{i+1} при известном множестве своих родителей – пустом множестве, что означает безусловную независимость Z_{i-1} от Z_{i+1}). Однако мы не можем утверждать о наличии условной независимости Z_{i-1} от Z_{i+1} при известном

значении узла Z_i. Так, узлы Z_{i-1} и Z_{i+1} могут описывать две независимые причины, а узел Z_i – симптом проявления комбинации этих причин. В этом случае причины могут становиться зависимыми при известном значении симптома. Например, если Z_{i-1} – состояние тормозов автомобиля, а Z_{i+1} – состояние дороги, то эти факторы независимы. Но если Z_i описывает факт аварии и известно, что авария произошла, то знание о состояние дороги позволяет с большей уверенностью делать вывод о состоянии тормозов: если дорога была хорошая, то так как авария произошла, то вероятность того, что с тормоза были неисправны, возрастает.

Определение: Пусть даны два различных узла $X, Y \in N$, маршрут $R = \langle X, Z_1, ..., Z_k, Y \rangle$, соединяющий эти узлы, а также не содержащее X и Y множество $A \subset N$. Мы будем говорить, что **множество A блокирует маршрут R между узлами X и Y**, если выполняется одно из условий:

1. Существует узел Z, входящий в маршрут R с типом 1 и принадлежащий множеству A. $\exists Z \in_1 R : Z \in A$
2. Существует узел Z, входящий в маршрут R с типом 2, причем ни сам узел Z, а также никто из его потомков не входят в A. $\exists Z \in_2 R : (Z \notin A) \wedge \left(D(\{Z\}) \bigcap A \neq \varnothing\right)$.

Пример 1:

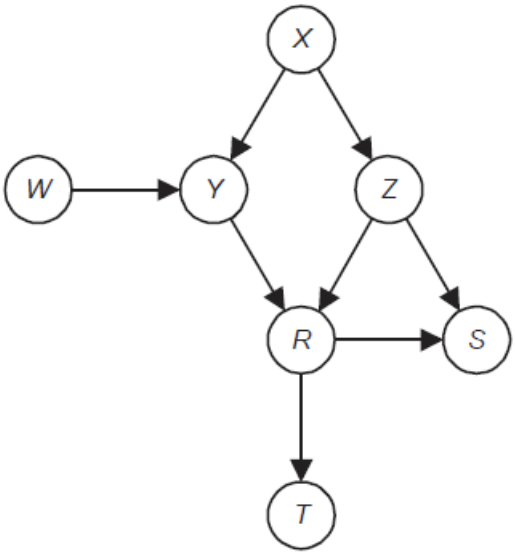


Рис. 5 – *Пример байесовской сети*

1. Маршрут <Y, X, Z, S> блокируется множеством {X}, потому что X входит в маршрут с типом 1.
2. Маршрут <W, Y, R, Z, S> блокируется пустым множеством, потому что R входит в маршрут с типом 2 и ни R, ни его единственный потомок T не принадлежат блокирующему пустому множеству.
3. Маршрут <W, Y, R, S> блокируется множеством {R}, потому что R входит в маршрут с типом 1.
4. Маршрут <W, Y, R, Z, S> не блокируется множеством {R}, потому что R входит в маршрут с типом 2.

Определение: Пусть даны два узла X и Y, и множество A. Мы будем говорить, что **множество A разделяет узлы X и Y**, если оно блокирует все маршруты между X и Y.

Пример 2: Рассмотрим снова байесовскую сеть, изображенную на рис. 5

1. X и R разделяются множеством {Y, Z} потому что маршрут <X, Y, R> блокируется в узле Y, а маршрут <X, Z, R> блокируется в узле Z.
2. X и T разделяются {Y, Z}, потому что маршрут <X, Y, R, T> блокируется в узле Y , маршрут <X, Z, R, T> блокируется в узле Z, а маршрут <X, Z, S, R, T> блокируется в узлах Z и S.

Определение: Пусть A, B, C – попарно непересекающиеся множества узлов направленного ациклического графа G. Мы говорим, что **множество C разделяет множества A и B**, если оно разделяет все пары узлов, входящие в A и B. Мы будем обозначать факт разделения $J_G\left(A,B|C\right)$.

Пример 3: Снова рассмотрим байесовскую сеть, изображенную на рис. 5. Так как каждый маршрут между W и S, W и T, X и S, а также X и T блокируется множеством {R, Z}, то $J_G\left(\{\mathrm{W},\mathrm{X}\},\{\mathrm{S},\ \mathrm{T}\}|\{\mathrm{R},\mathrm{Z}\}\right)$.

Из следующего утверждения получаем тот же набор свойств разделения множеств, как и для условной независимости:
Утверждение 9:
Для попарно непересекающихся множеств X, Y, Z, W следует:

1. Тривиальное разделение:
$$J_G\left(X,\varnothing|Z\right) \qquad \textbf{(20)}$$
2. Симметрия:
$$J_G\left(X,Y|Z\right)\Leftrightarrow J_G\left(Y,X|Z\right) \qquad \textbf{(21)}$$
3. Разложение:
$$J_G\left(X,Y\bigcup W|Z\right)\Rightarrow J_G\left(X,Y|Z\right)\wedge J_G\left(X,W|Z\right) \qquad \textbf{(22)}$$
4. Слабое объединение:
$$J_G\left(X,Y\bigcup W|Z\right)\Rightarrow J_G\left(X,Y\middle|Z\bigcup W\right)\wedge J_G\left(X,W\middle|Z\bigcup Y\right) \qquad \textbf{(23)}$$
5. Сжатие:
$$J_G\left(X,Y|Z\right)\wedge J_G\left(X,W\middle|Z\bigcup Y\right)\Rightarrow J_G\left(X,Y\bigcup W|Z\right) \qquad \textbf{(24)}$$
6. Пересечение:
$$J_G\left(X,Y\middle|Z\bigcup W\right)\wedge J_G\left(X,W\middle|Z\bigcup Y\right)\Rightarrow J_G\left(X,Y\bigcup W|Z\right) \qquad \textbf{(25)}$$

Доказательство:
Докажем для примера опять равенство (24). Рассмотрим произвольный маршрут R между множествами X и $Y\bigcup W$. Этот маршрут соединяет две какие-то точки между этими множествами. Возможны два случая:

1. Маршрут R соединяет $x\in X$ и $y\in Y$. Из условия $J_G\left(X,Y|Z\right)$ следует, что маршрут R блокируется множеством Z. Из определения блокировки возможны два варианта:
   1.1. Существует узел $a\in_1 R$, входящий в маршрут R с типом 1, принадлежащий также множеству Z. Следовательно, **Z блокирует маршрут R**.
   1.2. Существует узел $a\in_2 R$, входящий в маршрут R с типом 2. Т.е. ни он, ни его потомки не входят в Z. Следовательно, **Z блокирует маршрут R**.
2. Маршрут R соединяет $x\in X$ и $w\in W$. Из условия $J_G\left(X,W\middle|Z\bigcup Y\right)$ следует, что маршрут R блокируется множеством $Z\bigcup Y$. Опять из определения блокировки возможны два варианта:
   2.1. Существует узел $a\in_1 R$, входящий в маршрут R с типом 1, принадлежащий также множеству $Z\bigcup Y$. Возможны два варианта:
      2.1.1. Если $a\in Z$, то **Z блокирует этот маршрут**.

2.1.2. Если $a \in Y$, то рассматривая под-маршрут $R' \subset R$, соединяющий $x \in X$ и $a \in Y$, мы получаем уже ранее рассмотренный случай 1, а, следовательно, вывод: Z блокирует маршрут $R' \subset R$, т.е. **Z блокирует маршрут R**.

2.2. Существует узел $a \in_2 R$, входящий в маршрут R с типом 2, при этом ни он, ни его потомки не входят в $Z \bigcup Y$ и, следовательно, ни он, ни его потомки не входят в Z. Следовательно, **Z блокирует маршрут R**.

Таким образом, в любом случае, мы получаем вывод, что Z блокирует все маршруты между всеми парами из X и $Y \bigcup W$, т.е. $J_G\left(X, Y \bigcup W \middle| Z\right)$.

□

Определение: **гипотеза о разделении**, соответствующая байесовской сети G, представляет собой предположение, что если для попарно непересекающихся множеств узлов X, Y и Z справедливо утверждение, что Z разделяет X и Y, то X и Y являются условно независимыми при известных значениях Z.

Гипотезу о разделении мы будем обозначать G_{sep}, а множество параметров, удовлетворяющих этой гипотезе Θ_{sep}. Из определения гипотезы о разделении следует:

$$J_G\left(X, Y \middle| Z\right) \Rightarrow I_\theta\left(X, Y \middle| Z\right), \forall \theta \in \Theta_{G_{sep}} \qquad \textbf{(26)}$$

Утверждение 10: Из гипотезы о разделении следует гипотеза об условной независимости, т.е. $G_{sep} \Rightarrow G_{ind}$.

Доказательство:

Нам надо доказать, что множество родителей узла разделяет сам узел от множества его не-потомков. В этом случае, согласно гипотезе о разделении, этот узел будет независим от не-потомков при известном значении родителей, что влечет справедливость гипотезы об условной независимости.

Пусть $x \in N$ - произвольный узел, $Y := N \setminus \left(D\left(\{x\}\right) \bigcup \{x\} \bigcup P\left(\{x\}\right)\right)$ - множество, не содержащее x, родителей x и потомков x, а $Z := P\left(\{x\}\right)$ - множество родителей x. X, Y и Z – попарно непересекающиеся множества.

Если между X и Y нет маршрута, то X и Y разделяет любое множество, в том числе, Z, следовательно, по гипотезе о разделении X и Y, условно независимы при известных значениях Z, т.е. $I_\theta\left(X, Y \middle| Z\right), \forall \theta \in \Theta_{G_{sep}}$.

Пусть R – маршрут между x и $y \in Y$ и пусть a – первый узел, смежный с x. Возможны следующие варианты соединения дуги между x и a:

1. x <- a, в этом случае a входит в маршрут R с типом 1, а также является родителем x, т.е. $a \in Z$. Следовательно, Z разделяет маршрут R. По гипотезе о разделении $I_\theta\left(X, Y \middle| Z\right), \forall \theta \in \Theta_{G_{sep}}$.
2. x -> a. Здесь возможны два случая:
   2.1. Все последующие соединения в маршруте R имеют вид: x -> a -> … -> y. В этом случае y – потомок x, что противоречит выбору y
   $$y \in Y = N \setminus \left(D\left(\{x\}\right) \bigcup \{x\} \bigcup P\left(\{x\}\right)\right) \Rightarrow Y \bigcap D\left(\{x\}\right) = \varnothing \Rightarrow y \notin D\left(\{x\}\right).$$
   2.2. В маршруте есть узел с соединением типа 2. Пусть w – первый по порядку такой узел, т.е. x -> a ->… ->w<-…y. Если бы w или его потомок принадлежал бы множеству Z, т.е. был бы

родителем x, то в графе существовал бы цикл от x к w, затем (возможно через потомков) к родителю x, а затем – к x. Поэтому, ни сам w, ни его потомки не принадлежат Z, т.е. Z разделяет маршрут R, поэтому $I_\theta\left(X,Y\middle|Z\right),\forall\theta\in\Theta_{G_{sep}}$.

Мы получили, что в любом случае $I_\theta\left(X,Y\middle|Z\right),\forall\theta\in\Theta_{G_{sep}}$, т.е.

$$I_\theta\left(\{x\},N\setminus\left(D\left(\{x\}\right)\bigcup\{x\}\bigcup P\left(\{x\}\right)\right)\middle|P\left(\{x\}\right)\right),\forall\theta\in\Theta_{G_{sep}} \qquad \textbf{(27)}$$

Учитывая $I\left(X,Y\middle|Z\right)\Rightarrow I\left(X,Y\bigcup Z\middle|Z\right)$, получаем из (27):

$$I_\theta\left(\{x\},N\setminus\left(D\left(\{x\}\right)\bigcup\{x\}\right)\middle|P\left(\{x\}\right)\right),\forall\theta\in\Theta_{G_{sep}}. \qquad \textbf{(28)}$$

Итак, (28) означает, что каждый узел байесовской сети условно независим от своих не-потомков при известных значениях родителей, что означает выполнение гипотезы об условной независимости. Получаем $\Theta_{G_{sep}}\subset\Theta_{G_{ind}}$ или $G_{sep}\Rightarrow G_{ind}$.

□

Нам понадобится вспомогательное

Утверждение 11: Свойство слабой транзитивности:

$$J_G\left(X,Y\middle|W\right)\wedge J_G\left(X,Y\middle|W\bigcup\{v\}\right)\Rightarrow J_G\left(X,\{v\}\middle|W\right)\vee J_G\left(Y,\{v\}\middle|W\right), \qquad \textbf{(29)}$$

где X, Y, W, {v} – непересекающиеся множества и {v} состоит из одного элемента.

Доказательство:

Пусть выполнена левая часть (29): $J_G\left(X,Y\middle|W\right)\wedge J_G\left(X,Y\middle|W\bigcup\{v\}\right)$. Предположим, что правая часть (29) не выполняется, т.е. W не разделяет X и {v}, а также Y и {v}: $\neg J_G\left(X,\{\mathrm{v}\}\middle|W\right)\wedge\neg J_G\left(Y,\{\mathrm{v}\}\middle|W\right)$. Следовательно, есть маршрут R_1 между неким узлом x из X и v и маршрут R_2 между неким узлом y из Y и v, оба которых не блокируются множеством W. Маршрут R, полученный соединением R_1 и R_2, соединяет x и y через v. По условию, множества W и $W\bigcup\{v\}$ разделяют X и Y, а, следовательно, блокируют маршрут R. Пусть узел w_1 из R блокируется множеством W, а узел w_2 из R блокируется множеством $W\bigcup\{v\}$.

1. Если w_1 не равен v, то он лежит или на маршруте R_1 от x до v или на маршруте R_2 от y до v. Пусть для определенности $w_1\in R_1$. Так как множество W блокирует маршрут R в узле w_1, то
   1.1. w_1 либо принадлежит W, если w_1 входит в R (и R1) с типом 1, либо
   1.2. w_1 и все его потомки не принадлежат в W, а сам он входит в R (и R_1) с типом 2.
   В обоих этих случаях получается, что множество W блокирует маршрут R_1 через w_1, что **противоречит выбору R_1**. Аналогично:
2. Если w_2 не равен v, то w_2 лежит на маршруте R_1 или R_2. Пусть $w_2\in R_1$. Так как маршрут R блокируется множеством $W\bigcup\{v\}$ в точке w_2, то
   2.1. либо w_2 принадлежит W (так как w_2 не может принадлежать {v}) и входит в R (и в R_1) с типом 1,
   2.2. либо ни сам w_2, ни его потомки не принадлежат $W\bigcup\{v\}$, а, следовательно, W. При этом w_2 входит в R (и в R_1) с типом 2.
   Получаем, что W блокирует маршрут R_1 в точке w_2, что **противоречит выбору R_1**.
3. Если оба w_1 и w_2 равны v, то
   3.1. Если v входит в маршрут R с типом 1, то так как W блокирует R через v, то v должен лежать в W. Но {v} и W по условию не пересекаются, и **мы приходим к противоречию**.
   3.2. Если v входит в маршрут R с типом 2, то так как $W\bigcup\{v\}$ блокирует R через v, то v не должен принадлежать $W\bigcup\{v\}$, что очевидно, **является противоречием**.

Из полученных противоречий следует, что предположение $\neg J_G\left(X,V\middle|W\right)\wedge\neg J_G\left(Y,V\middle|W\right)$ было неверным, и, следовательно, утверждение доказано.

□

Утверждение 12: Из гипотезы об условной независимости следует гипотеза о разделении, т.е. $G_{ind}\Rightarrow G_{sep}$.

Доказательство (Thomas Verma, Judea Pearl)**:**

Пусть дан граф G=(N,E), и вероятностная модель P_θ, удовлетворяющая гипотезе об условной независимости, т.е. $\theta\in\Theta_{G_{ind}}$. Нам нужно показать, что $\theta\in\Theta_{G_{sep}}$, т.е. для параметра θ выполняется также гипотеза о разделении, что означает условную независимость X и Y при Z, если Z разделяет X и Y: $J_G\left(X,Y\middle|Z\right)\Rightarrow I_\theta\left(X,Y\middle|Z\right)$.

Мы будем проводить доказательство по индукции от числа узлов n=#N. При n=1 нечего доказывать. При n=2 два узла X и Y разделены тогда и только тогда, когда между ними нет связи, что влечет на основании гипотезы об условной независимости их безусловную независимость (X не зависит от не-потомка Y при пустом множестве родителей).

Итак, пусть утверждение справедливо для всех n<k. Покажем справедливость утверждения для n=k. Пронумеруем узлы согласно некой топологической нумерации. Тогда из гипотезы об условной независимости следует:

$$I_\theta\left(\{X_i\},\{X_1,..,X_{i-1}\}\middle|P\left(X_i\right)\right),\forall i\in\{1,...,n\} \qquad \textbf{(30)}$$

Определим граф G-n как граф, полученный из G удалением n-ого узла и всех инцидентных ему дуг. По предположению индукции $J_{G-n}\left(X,Y\middle|Z\right)\Rightarrow I_\theta\left(X,Y\middle|Z\right)$ для всех попарно непересекающихся множеств X, Y, Z из N \ {X_n}.

Пусть для попарно непересекающихся множеств X, Y, Z из N справедливо: $J_G\left(X,Y\middle|Z\right)$. В зависимости от вариантов принадлежности n-ого узла X_n мы будем рассматривать 4 различных случая.

1. X_n не входит ни в одно из множеств X, Y и Z, т.е. $X_n\notin X\bigcup Y\bigcup Z$. Следовательно, X, Y и Z принадлежат графу G-n. Z разделяет X и Y в G, следовательно, Z разделяет X и Y в G-n, так как каждый маршрут из G-n является также маршрутом из G и блокируется множеством Z, которое также принадлежит G-n. Т.е. $J_{G-n}\left(X,Y\middle|Z\right)$, и, следовательно, из предположения индукции $I_\theta\left(X,Y\middle|Z\right)$.
2. X_n входит в X, т.е. $X_n\in X$. Тогда X можно представить как $X=\{X_n\}\bigcup X',X'\subset N\setminus\{X_n\}$. Из гипотезы об условной независимости следует: $I_\theta\left(\{X_n\},\{X_1,..,X_{n-1}\}\setminus P\left(X_n\right)\middle|P\left(X_n\right)\right)$.
   Обозначим
   $B:=P\left(X_n\right)$ и $A:=\{X_1,..,X_{n-1}\}\setminus P\left(X_n\right)$, следовательно
   $$I_\theta\left(\{X_n\},A\middle|B\right),A\bigcap B=\varnothing,A\bigcup B=\{X_1,..,X_{n-1}\}=G-n \qquad \textbf{(31)}$$
   Так как множества X', Y и Z попарно непересекающиеся, то можно разложить множества A и B на следующие попарно непересекающиеся множества:

$$B_X := B\bigcap X', B_Y := B\bigcap Y, B_Z := B\bigcap Z, B_0 := B\setminus\left(B_X\bigcup B_Y\bigcup B_Z\right)\Rightarrow B = B_X + B_Y + B_Z + B_0,$$

где под + понимается операция объединения непересекающихся множеств. Аналогично:

$$A_X := A\bigcap X', A_Y := A\bigcap Y, A_Z := A\bigcap Z, A_0 := A\setminus\left(A_X\bigcup A_Y\bigcup A_Z\right)\Rightarrow A = A_X + A_Y + A_Z + A_0$$

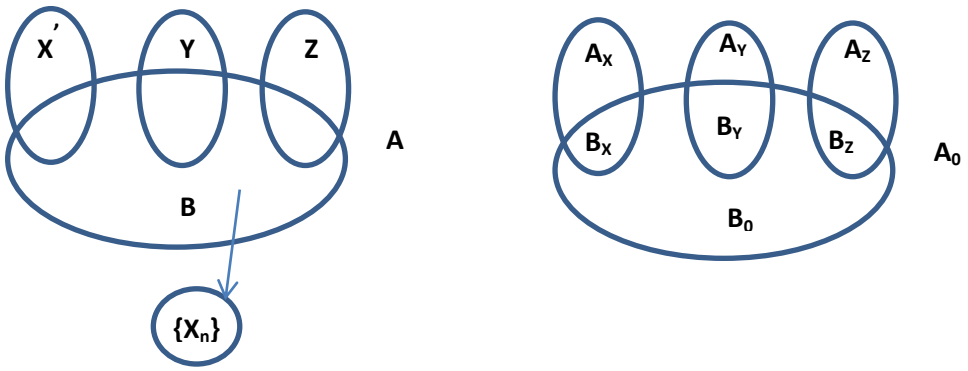


Рис 6. – *Разбиение множеств A и B*

Так как B – множество родителей X_n, то в графе G из каждой точки B есть дуга к X_n, в том числе есть дуга из всех узлов B_Y к X_n. Но так как по условию $J_G\left(\underbrace{X'\bigcup\{X_n\}}_{=X}, Y\middle|Z\right)$, то каждый маршрут из любого узла из Y до X_n должен блокироваться Z, в том числе должны блокироваться все пути из B_Y к X_n, но это невозможно, т.к. дугу нельзя блокировать. Поэтому множество B_Y должно быть пустым. Следовательно, $B = B_X + B_Z + B_0$ и $A_Y = Y \Rightarrow A = A_X + Y + A_Z + A_0$. Таким образом, (31) принимает вид:

$$I_\theta\left(\{X_n\}, A_X + Y + A_Z + A_0\middle|B_X + B_Z + B_0\right) \qquad \textbf{(32)}$$

С учетом (17) и (18) получаем из (32):

$$I_\theta\left(\{X_n\}, A_X + Y + A_Z + A_0\middle|B_X + B_Z + B_0\right)\Rightarrow$$

$$\underset{(18)}{\Rightarrow} I_\theta\left(\{X_n\}, Y + A_0\middle|\underbrace{B_X + A_X}_{X'} + \underbrace{B_Z + A_Z}_{Z} + B_0\right)\Rightarrow$$

$$\underset{(17)}{\Rightarrow} I_\theta\left(\{X_n\}, Y\middle|X' + B_0 + Z\right)$$

Итак $I_\theta\left(\{X_n\}, Y\middle|X' + B_0 + Z\right)$ **(33)**

Предположим, что существует маршрут от Y до B_0, который не блокируется множеством Z в G. Дополним этот маршрут дугой от узла w из B_0 до X_n (такая дуга есть, так как все узлы из B_0 – родители X_n). По условию Z блокирует все маршруты из Y в X_n через G, т.е. Z блокирует узел w, который входит в маршрут с типом 1. Следовательно, w должен принадлежать Z, чего не может быть, так как B_0 и Z не пересекаются вследствие того, что

$$B_0\bigcap B_Z = \varnothing,$$

$B_0 \subset B, A_Z \subset A, B\bigcap A = \varnothing \Rightarrow B_0\bigcap A_Z = \varnothing$ и поэтому $B_0\bigcap\underbrace{\left(B_Z + A_Z\right)}_{Z} = \varnothing$. Из противоречия следует, что Z разделяет Y от B_0 в G, т.е.

$$J_G\left(B_0, Y\middle|Z\right) \qquad \textbf{(34)}$$

Так как Z разделяет Y как от X' (по условию), так и от B_0 (согласно (34)), то Z также разделяет Y от объединения этих множеств, т.е.

$$J_G\left(B_0+X',Y\middle|Z\right) \qquad \textbf{(35)}$$

Узел X_n не входит ни в одно из трех множеств в утверждении (35), поэтому, как уже было показано в случае 1, из (35) следует разделение в графе G-n, а, следовательно, по предположению индукции, справедливость соответствующего утверждения об условной независимости:

$$J_G\left(B_0+X',Y\middle|Z\right) \underset{X_n\notin B_0+X'+Y+Z}{\Rightarrow} J_{G-n}\left(B_0+X',Y\middle|Z\right) \underset{\text{индукция}}{\Rightarrow} I_\theta\left(B_0+X',Y\middle|Z\right) \qquad \textbf{(36)}$$

С учетом свойств (17) и (19) получаем из (36) и (33):

$$I_\theta\left(Y,B_0+X'\middle|Z\right)\wedge I_\theta\left(Y,\{X_n\}\middle|X'+B_0+Z\right)\Rightarrow$$

$$\underset{(19)}{\Rightarrow} I_\theta\left(Y,B_0+\underbrace{X'+\{X_n\}}_{=X}\middle|Z\right)\underset{(17)}{\Rightarrow} I_\theta\left(Y,X\middle|Z\right)$$

Мы получили $I_\theta\left(X,Y\middle|Z\right)$

3. X_n входит в Y. Из-за симметрии разделения и условной независимости этот случай доказывается аналогично случаю 2.

4. X_n входит в Z. Тогда Z можно представить как $Z=\{X_n\}\bigcup Z', Z'\subset N\setminus\{X_n\}$.

   Все маршруты, из X в Y, блокируемые множеством $Z=\{X_n\}\bigcup Z'$, блокируются множеством Z'. Действительно, пусть есть маршрут из X в Y, который блокируется Z и не блокируется Z'. Тогда либо есть узел w, входящий в маршрут с типом 1, который входит в Z, но не входит в Z'. Значит w входит в $Z\setminus Z'=\{X_n\}$, т.е. w=X_n. Но это невозможно, потому что X_n – последний узел согласно топологической нумерации, а, следовательно, у него нет детей, поэтому все инцидентные ему дуги входят в него и он не может входить в маршрут с типом 1. Либо есть узел w, входящий в маршрут с типом 2, причем ни он сам, ни его потомки не входят в Z. Следовательно, ни он сам, ни его потомки не входят в Z', а значит, он также блокируется множеством Z'. Мы получаем:

$$J_G\left(X,Y\middle|Z'\right) \qquad \textbf{(37)}$$

   Из (37) вместе с условием $J_G\left(X,Y\middle|Z\right)$ следует согласно утверждению 11, что

$$J_G\left(\{X_n\},X\middle|Z'\right)\vee J_G\left(\{X_n\},Y\middle|Z'\right) \qquad \textbf{(38)}$$

   Комбинируя (37) и (38), получаем:

$$\left(J_G\left(\{X_n\},X\middle|Z'\right)\vee J_G\left(\{X_n\},Y\middle|Z'\right)\right)\wedge J_G\left(X,Y\middle|Z'\right)\Rightarrow$$

$$\Rightarrow\left(J_G\left(\{X_n\},X\middle|Z'\right)\wedge J_G\left(Y,X\middle|Z'\right)\right)\vee\left(J_G\left(\{X_n\},Y\middle|Z'\right)\wedge J_G\left(X,Y\middle|Z'\right)\right)\Rightarrow$$

$$\Rightarrow J_G\left(\{X_n\}\bigcup Y,X\middle|Z'\right)\vee J_G\left(\{X_n\}\bigcup X,Y\middle|Z'\right)$$

   Т.е. получаем

$$J_G\left(X,Y\bigcup\{X_n\}\middle|Z'\right)\vee J_G\left(X\bigcup\{X_n\},Y\middle|Z'\right) \qquad \textbf{(39)}$$

   Условие (39) соответствует одному из уже рассмотренных случаев 2 или 3, для которых было доказано $I_\theta\left(X,Y\bigcup\{X_n\}\middle|Z'\right)$ или $I_\theta\left(X\bigcup\{X_n\},Y\middle|Z'\right)$ соответственно. Пусть, например, справедливо $I_\theta\left(X\bigcup\{X_n\},Y\middle|Z'\right)$. Тогда с учетом (23) получаем $I_\theta\left(X,Y\middle|Z'\bigcup\{X_n\}\right)$, что означает $I_\theta\left(X,Y\middle|Z\right)$

Таким образом, мы доказали условную независимость X и Y от Z, т.е. выполнение гипотезы о разделении.

□

Непосредственно из утверждений 7, 10, 12 следует:

Утверждение 13: Гипотеза об условной независимости, гипотеза о факторизации и гипотеза о разделении, соответствующие одной байесовской сети, совпадают, т.е. $G_{fact} \Leftrightarrow G_{ind} \Leftrightarrow G_{sep}$ или $\Theta_{G_{ind}} = \Theta_{G_{fact}} = \Theta_{G_{sep}}$.

□

Определение: мы будем называть гипотезу об условной независимости, эквивалентную ей гипотезу о факторизации и эквивалентную гипотезу о разделении, соответствующую байесовской сети G=(N,E), как **гипотезу, соответствующую байесовской сети**, и обозначать ее G. Множество параметров вероятностного распределения на $\{1,...,r_1\} \times ... \times \{1,...,r_n\}$, удовлетворяющее соответствующим гипотезам – Θ_G.

В дальнейшем, когда мы будем рассматривать байесовскую сеть, мы будем по умолчанию предполагать, что для рассматриваемого вероятностного распределения на множестве значений переменных выполняется гипотеза, соответствующая этой сети.

4.3.1 Следствия из гипотезы о разделении

Понятие о разделении множеств позволяет по диаграмме байесовской сети визуально определять условную независимость переменных.

Введем полезное понятие.

Определение: **Марковским покрытием** (Markov Blanket) узла байесовской сети называется множество родителей, детей и других родителей своих детей, т.е.

$$MB(X) := P(X) \bigcup C(X) \bigcup \left\{ y \in N \middle| C(\{y\}) \bigcap C(X) \neq \varnothing \right\}.$$

Утверждение 14: Любой узел байесовской сети условно не зависит при известном значении своего Марковского покрытия от всех остальных узлов сети.

Доказательство: Пусть $x \in N, y \in N \setminus MB(\{x\})$. Достаточно показать, что любой маршрут из y к x блокируется MB(x). Любой маршрут R из y к x проходит либо через множество родителей x, либо через множество детей x.

1. Если R проходит через узел w, находящийся в множестве родителей x, то он входит в R с типом 1, так как все дуги между родителями и x направлены к x. Таким образом, R блокируется множеством родителей.
2. Если R проходит через узел w, находящийся в множестве детей x, то возможны два варианта:
   2.1. Другая инцидентная к w дуга направлена от w, т.е. фрагмент маршрута, содержащий w, выглядит x -> w ->. Тогда w входит в R с типом 1, т.е. R блокируется множеством детей.
   2.2. Другая инцидентная к w дуга направлена к w, т.е. w входит в R с типом 2. Соответствующий фрагмент маршрута выглядит как x -> w <- z. В этом случае z входит в маршрут с типом z, а также входит в множество родителей детей x, которое его блокирует.

□

Утверждение 15: Два узла a и b являются смежными тогда и только тогда, когда они не разделяются объединенным множеством своих родителей. Т.е. для графа G=(N, E) и двух узлов a и b справедливо: $\langle a,b\rangle \in E \vee \langle b,a\rangle \in E \Leftrightarrow \neg J_G\left(a,b \middle| P\left(\{a,b\}\right)\right)$

Доказательство:

1. ->: Если два узла связаны дугой, например a->b, то маршрут, состоящий из этой дуги, нельзя по определению блокировать никаким множеством, так как этот маршрут не содержит внутренних узлов. В том числе его нельзя блокировать множеством $P\left(\{a,b\}\right)$, т.е. $P\left(\{a,b\}\right)$ не разделяет a и b.

2. <-: Пусть справедливо $\neg J_G\left(a,b \middle| P\left(\{a,b\}\right)\right)$, тогда между a и b существует маршрут R, который не блокируется множеством $P\left(\{a,b\}\right)$. Предположим, что узлы a и b не смежные, тогда маршрут R содержит внутренние узлы.

   Если бы все водящие в R внутренние узлы входили маршрут с типом 1, т.е. имели хотя бы одну исходящую дугу, то ближайший к a или ближайший к b узел w в маршруте имел бы дугу, входящую в a или b соответственно. Соответствующий фрагмент маршрута выглядел бы так: a<-w…b или a…w->b. В этом случае w принадлежал бы родителям a или b, т.е. $w \in P\left(\{a,b\}\right)$ и, следовательно, блокировался бы $P\left(\{a,b\}\right)$, что противоречит $\neg J_G\left(a,b \middle| P\left(\{a,b\}\right)\right)$.

   Итак, в маршруте R есть хотя бы один узел, входящий с типом 2. Пусть w_1 такой узел, ближайший к узлу a, а w_2 – ближайший к узлу b (w_1 и w_2 могут совпадать). Маршрут R будет иметь вид: a->…->w_1<-……->w_2<-…<-b. Так как R не блокируется родителями a и b, то сам узел w_1 или его потомок должны находиться в $P\left(\{a,b\}\right)$. Если бы w_1 или его потомок находились в $P\left(\{a\}\right)$, то в G был бы цикл, идущий от a до w_1, а от него к множеству родителей a, и снова к самому a. Итак, $w_1 \in P\left(\{b\}\right)$. Аналогично получаем $w_2 \in P\left(\{a\}\right)$. Мы получаем, что в G есть цикл, идущий от a к w_1, затем к родителям b, от b к w_2, а от w_2 к родителям a, а от них – снова к a.

   Мы пришли к противоречию, и это значит, что узлы a и b смежные.

□

Следствие 1 из утверждения 15: Два узла смежные тогда и только тогда, когда они не разделяются никаким множеством.

Доказательство:
По определению разделения два смежных узла нельзя разделить никаким множеством. Если же два узла нельзя разделить никаким множеством, то их в частности, нельзя разделить объединенным множеством их родителей, следовательно, по утверждению 15, они смежные.

□

Определение: Пусть даны три узла a, b и c. Мы говорим, что a,b и c образуют **v-структуру abc**, если выполняются два условия:

1. Узлы связаны a -> b <- c, т.е. $a,c \in P\left(\{b\}\right)$.
2. Узлы a и c не смежные, т.е. $a \notin P\left(\{c\}\right) \wedge c \notin P\left(\{a\}\right)$.

Утверждение 16: Смежные пары узлов a-b и b-c, образуются v-структуру abc тогда и только тогда, когда выполнено два условия:
1. Существует множество, которое разделяет узлы a и c.
2. Любое разделяющее a и c множество не содержит b.

Доказательство:
1. По следствию 1 из утверждения 15 первый пункт условия утверждения 16 эквивалентен тому, что узлы a и c не смежные, т.е. второму пункту определения v-структуры.

2. Если узлы связаны a->b<c, то любое разделяющее a и c множество не должно содержать b по определению разделения.

   Пусть любое разделяющее a и c множество не содержит b, a Z-некое множество, разделяющее a и c. Так как существует маршрут a-b-c, то если b входит в него с типом 1, то b должен принадлежать Z по определению разделения. Следовательно, b входит в маршрут с типом 2, т.е. a->b<-c.

   Т.е. второй пункт условия утверждения 16 эквивалентен первому пункту определения v-структуры.

□

5. Эквивалентность байесовских сетей

Определение: Две байесовские сети G_1 и G_2 называются **эквивалентными**, если им соответствуют одинаковое множество высказываний о разделении. Т.е. для всех попарно непересекающихся множеств X, Y и Z справедливо $J_{G_1}\left(X,Y\middle|Z\right) \Leftrightarrow J_{G_2}\left(X,Y\middle|Z\right)$. Мы будет обозначать эквивалентность сетей G_1~G_2.

Отношение ~ является отношением эквивалентности на множестве НАГ, т.е. является рефлексивным, симметричным и транзитивным.

Если две сети эквивалентны, то соответствующие этим сетям гипотезы имеют одинаковый набор высказываний об условной независимости, и, следовательно, совпадающие ограничения на множестве параметров распределения. Таким образом, для эквивалентных сетей $G_1 \sim G_2 \Rightarrow \Theta_{G_1} = \Theta_{G_2}$.

Утверждение 17: Две сети эквиваленты тогда и только тогда, когда у них одинаковый скелет и v-структуры.

Доказательство: (Thomas Verma, Judea Pearl).

->: Пусть две сети G_1 и G_2 эквивалентны.
1. Предположим, что существуют два узла a и b, смежные в G_1 и не смежные в G_2. Следовательно, из утверждения 15 a и b разделяются множеством своих объединенных родителей в G_2, но не разделяются никаким множеством в G_1. Значит у G_1 и G_2 разные утверждения о разделении, что противоречит определению эквивалентности G_1 и G_2. Таким образом, два узла смежные в одной сети, если они смежные в другой. Т.е. G_1 и G_2 должны иметь одинаковый набор ребер, следовательно, одинаковый скелет.
2. Предположим, что в G_1 существует v-структура abc, которой нет в G_2. По уже доказанному, G_1 и G_2 имеют одинаковый скелет, следовательно, в G_2 также есть маршрут a-b-c, и узлы a и c также не смежные в G_2. Так как узлы a и c не смежные в обеих сетях, то по утверждению 15 в обеих сетях существует множества, которые разделяют эти узлы.

Разделяющее множество в G_1 не может содержать узел b, иначе бы маршрут a->b<-c с типом 2 не блокировался. Так как abc в G_2 – не v-структура, то узел b входит в маршрут a-b-c в G_2 с типом 1. Следовательно, узел b должен входить в любое разделяющее a и c множество. Мы получили, что непустой набор множеств, разделяющий a и b в обеих сетях, не пересекается (в G_1 все разделяющие множества не содержат b, а все в G_2 – содержат). Это противоречит эквивалентности G_1 и G_2. Таким образом, G_1 и G_2 содержат одинаковые v-структуры.

<-: Пусть у G_1 и G_2 одинаковый скелет и набор v-структур. Мы докажем, что любой, неблокируемый неким множеством маршрут в одной сети не будет блокироваться тем же множеством в другой. Пусть R – маршрут в G_1, не блокируемый множеством S. Пусть R будет также минимальным в следующем смысле: если k – число узлов в R, $a=r_1$ – первый узел, а $b=r_k$ – последний узел, то выполняется:
(a) между r_1 и r_k нет неблокируемого множеством S маршрута с длиной строго меньшей k;
(b) между r_1 и r_k нет другого неблокируемого множеством S маршрута P с числом узлов k и с элементами $p_1,\ldots,p_k$ для которых либо $p_i=r_i$ или p_i – потомок r_i для всех $1<i<k$.

Так как по условию, скелет у G_1 и G_2 одинаков, то R является также маршрутом в G_2. Мы покажем по индукции по числу узлов с входящими дугами (-> x <-) в маршруте R, что соответствующий маршрут в G_2 также не блокируется множеством S. Это доказательство будет разбито на три части. В первой части мы докажем, что если маршрут R не содержит узлы со сходящимися дугами, то он будет также не блокироваться множеством S в G_2 (первый шаг индукции). Во второй части мы покажем, что если R содержит хотя бы один узел x со сходящимися дугами, тогда R не блокируется множеством S в G_2 тогда и только тогда, когда узел x не блокируется множеством S в G_2. В третьей части мы докажем, что такой узел не блокируется множеством S в G_2, а значит, завершив индукционный шаг, покажем, что не блокируется и сам маршрут в G_2.

Часть 1.

Если R не содержит узлов со сходящимися дугами в G_1, то он также не содержит таких узлов в G_2. Действительно, предположим, что $x=r_i$ – узел со сходящимися в G_2 дугами, с родителями $y=r_{i-1}$ и $z=r_{i+1}$.

Возможные конфигурации для G_1 выглядят так:
r_1 … y -> x -> z … r_k
r_1 … y <- x -> z … r_k
r_1 … y <- x <- z … r_k

Для G_2 возможна только одна конфигурация:
r_1 … y -> x <- z … r_k

Если бы узлы y и z не были бы смежными в одной сети, то благодаря тому, что у сетей общий скелет, эти узлы не были бы смежными в другой сети. Значит, они образовывали бы в G_2 v-структуру. А так как и v-структуры у сетей совпадают, то узел x должен был бы входить и в G_1 со сходящимися дугами. Итак, узлы y и z должны быть смежными в обеих сетях. Но последовательность узлов, полученная из R удалением узла x в этом случае также была бы маршрутом. Более того, этот маршрут не блокировался бы в G_1 множеством S, так как этот маршрут содержит только узлы, входящие с типом 1, не блокируемые S. Это противоречит критерию «минимальности» (a).

Итак, мы показали, что маршрут R не содержит узлов со сходящимися дугами и в G_2. Так как ни один из узлов маршрута не входит в S (иначе бы он блокировался в G_1), то этот маршрут не блокируется множеством S и в G_2.

Часть 2.

Пусть теперь маршрут R содержит, по крайней мере, один узел со сходящимися дугами в G_1. Пусть $x=r_i$ – такой узел, с родителями $y=r_{i-1}$ и $z=r_{i+1}$, тогда G_1 выглядит так: a … y -> x <- z … b. Пусть R_1 – часть маршрута между узлами a и y, а R_2 – часть маршрута между узлами z и b. Если i-1 равно 1, то R_1 состоит из одного узла, если i+1 равно k, то R_2 состоит из одного узла.

Как R_1, так и R_2 не блокируются множеством S в G_1, так как они входят в маршрут R, который не блокируется S. Также R_1 и R_2 являются «минимальными» маршрутами в смысле условий (a) и (b), неблокируемыми множеством S. Далее, R_1 и R_2 содержат строго меньше узлов со сходящимися дугами, чем R. Значит, по индукции по числу узлов со сходящимися дугами в «минимальных» неблокируемых S в G_1 маршрутах, маршруты R_1 и R_2 не блокируются множеством S также и в G_2.

Предположим, что y и z смежные в G_1. Так как y и z входят в маршрут R с типом 1, то учитывая то, что S не блокирует R, то y и z не принадлежат S. Узел x входит в маршрут R с типом 2, поэтому он или его потомок w принадлежит S. Получим маршрут R' между a и b из маршрута R удалением узла x и используя дугу между y и z. Если узлы y и z входят в новый маршрут R' с типом 1, то они не блокируются S, так как не входят в S. Если же один из них входит в R' с типом 2, то он также не блокируется S, потому что его потомок w (через x) входит в S. Мы получили, что маршрут R' имеет строго меньше узлов, чем R, и не блокируется S, что противоречит «минимальности» R.

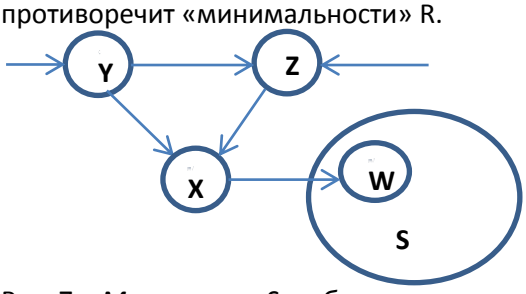


Рис. 7 – *Множество S не блокирует укороченный маршрут.*

Следовательно, y и z не могут быть смежными в G_1, а следовательно, и в G_2, и поэтому, узлы yxz формируют v-структуру в обеих сетях. Как уже было показано по индукции, R_1 и R_2 не блокируются S в G_2. Узлы y и z также не блокируются S в G_2, так как входят с типом 1 и не принадлежат S. Следовательно, маршрут R не блокируется в G_2 тогда и только тогда, когда узел x не блокируется S в G_2.

Часть 3.

Так как x не блокируется S в G_1, то или узел x принадлежит S, или в G_1 существует направленный путь от узла x к узлу w из S. В первом случае узел x очевидно блокируется S также и в G_2. Рассмотрим теперь второй случай.

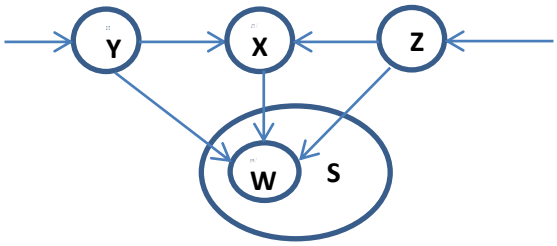


Рис. 8 – *X не блокирует маршрут*

Пусть ф – самый короткий путь из x в w в G_1. Покажем, что соответствующий маршрут в G_2 также является путем из x в w. Тогда x не блокирует S в G_2, и, следовательно, с учетом части 2, маршрут R не блокируется S в G_2. Доказательство будем вести по длине l пути ф.

При **l=0** x=w и доказывать нечего, так как пути нет.

При **l=1** мы имеем одну дугу в G_1.

Если оба узла y и z смежные с w, то инцидентные дуги могут быть направлены только в направлении y -> w и z -> w (рис. 8), так как при другой ориентации в графе G_1 были бы циклы. Итак, y и z родители w. Следовательно, маршрут R', полученный из R заменой узла x на w, является маршрутом из a к b, не блокируется S (w входит с типом 2 и лежит в S), и содержит ровно столько же узлов, как и R. Все узлы R' совпадают с узлами R кроме одного, являющегося потомком соответствующего узла из R. Это противоречит «минимальности» R (b), следовательно, хотя бы один из узлов y или z не связан с w.

Без ограничения общности, будем считать, что y не связан с w в G_1, и, следовательно, в G_2. Значит, в G_2 не может быть связи y -> x <- w, иначе yxw была бы v-структура в G_2, а, следовательно, v-структура в G_1, что не так. Значит, ориентация дуги x -> w в G_2 та же, что и в G_1. Следовательно, в G_2 есть путь от x к w.

При **l>1** путь ф содержит хотя бы две дуги. Рассмотрим две последние дуги к w, изображенные на рис. 9.

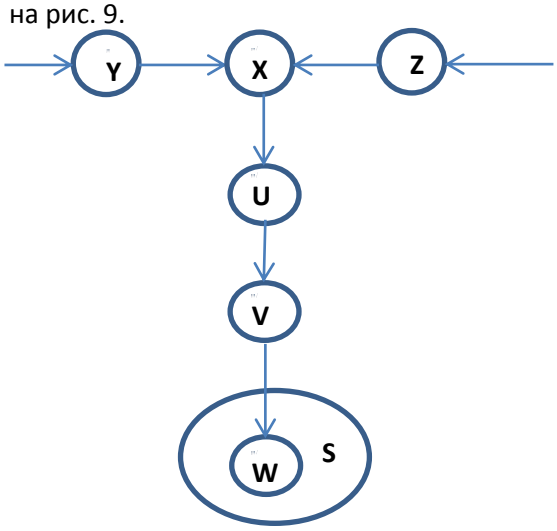


Рис. 9 – *Путь от x к w*

Фрагмент пути $ф_1$ от x до v имеет длину l-1 и, следовательно, по индукции также является путем в G_2. Если бы узел u был связан с w, то направление этой связи было бы u->w (иначе бы был цикл), а значит, существовал бы более короткий путь от x к w в G_1, минуя v, что противоречит выбору ф как кратчайшего пути. Итак, узлы u и w не связаны. В G_2 не может быть связи w->v, так как в противном случае в G_2, а значит, и в G_1 была бы v-структура uvw, чего не может быть, так как u->v->w – путь в G_1. Итак, связь между v и w в G_2 выглядит как v->w, т.е. ф является путем в G_2.

Мы показали, что между x и S существует путь в G_2, значит, узел x не блокируется S в G_2, следовательно (согласно части 2) маршрут R не блокируется S в G_2. Мы завершили доказательство индукционного перехода по числу узлов со сходящимися дугами.

Таким образом, из трех доказанных частей мы имеем, что любой «минимальный» маршрут, не блокирующийся S в G_1, также не блокируется S в G_2. В результате, что если какое-то множество S

не блокирует узлы a и b в G_1, то по определению блокировки, существует хотя бы один не блокируемый в G_1 множеством S маршрут между a и b. Выберем из множества неблокируемых маршрутов «минимальный», т.е. маршрут R, удовлетворяющий условиям (a) и (b). Из доказанного следует, что соответствующий маршрут в G_2 будет также не блокироваться множеством S. Значит, множество S не блокирует узлы a и b в G_2.

Следовательно, все высказывания о не разделении из G_1 справедливы для G_2 (и наоборот), а значит G_1 и G_2 – эквивалентны.

□

Мы получим еще одно необходимое и достаточное условие эквивалентности байесовских сетей. Для этого нам понадобится

Определение: дуга x -> y называется **покрытой** (covered edge), если у узлов x и y одинаковое множество родителей, за исключением того, что x является родителем y, т.е. $P(\{x\}) = P(\{y\}) \setminus \{x\}$.

Утверждение 18: Пусть дана байесовская сеть G_1 и некий граф G_2, отличающийся от G_1 только ориентацией одной дуги. Тогда G_2 – эквивалентная G_1 байесовская сеть тогда и только тогда, когда эта дуга является покрытой.

Доказательство: (David Maxwell Chickering).

Пусть две сети G_1 и G_2 отличаются только ориентацией дуги x -> y в G_1 и y -> x в G_2.

<-: Пусть дуга x -> y является покрытой в байесовской сети G_1, т.е. $P_{G_1}(\{y\}) = P_{G_1}(\{x\}) \bigcup \{x\}$.

Сначала мы покажем, что G_2 не содержит циклов. Если бы G_2 содержал цикл, то этот цикл содержал бы дугу y -> x, потому что G_1 циклов по условию не содержит. Значит, в G_1 есть направленный путь от x к y, который не содержит дугу x -> y. Пусть z будет последним узлом в этом пути, т.е. z – родитель y, отличный от x. Следовательно, z – родитель x в G_1, т.е. в G_1 есть путь x -> … -> z -> x, следовательно, G_1 содержит цикл, что противоречит тому, что G_1 - НАГ. Таким образом, мы показали, что G_2 не содержит циклов, т.е. является байесовской сетью.

Теперь покажем, что G_2 эквивалентна G_1. Так как G_1 и G_2 имеют один и тот же скелет, то по утверждению 17, если бы G_1 и G_2 были не эквивалентны, то у них отличались бы v-структуры. Предположим, что в G_2 есть v-структура, которой нет в G_1. Эта v-структура должна содержать дугу y -> x, потому что эта единственная дуга, которая отличается в обеих графах. Но это значит, что y x есть родитель, не смежный с y в обоих графах. Это противоречит тому, что любой родитель x есть также родитель y в обоих графах. Аналогичный аргумент применим, если мы предположим, что в G_1 есть v-структура, которой нет в G_2. Мы показали, что v-структуры графов совпадают, а, значит, они эквивалентны.

->: Пусть G_2 – байесовская сеть и $G_1 \sim G_2$ Мы покажем, что если дуга x -> y не покрытая в G_1, то G_2 или содержит цикл, или байесовская сеть G_2 не эквивалентна G_1. Если дуга x -> y не покрытая в G_1, то одно из двух условий должно выполняться в G_1: (1) существует узел z, являющийся родителем y, но не родителем x; (2) существует узел w, родитель x, но не родитель y. Пусть $z \neq x$ - родитель y в G_1, который не является родителем x в G_1. Если бы z и x не были бы смежные, то xyz была бы v-структура в G_1, которой нет в G_2. Это противоречит эквивалентности графов. Следовательно, узлы z и x смежные. Если x родитель z в G_1, то x – родитель z также и в G_2, так как эти графы отличаются только одной дугой, а, следовательно, в G_2 есть цикл x -> z -> y -> x. Это противоречит тому, что G_2 не содержит циклов.

Пусть узел w – родитель x в G_1 и не родитель y в G_1. Если бы узлы w и y не были бы смежными, то в G_2 существовала бы v-структура wxy, которой нет в G_1. Значит, графы не эквивалентны. Узел y не может быть родителем w в G_1, так как в G_1 тогда бы был цикл w -> x -> y -> w.

Из полученных противоречий, мы приходим к выводу, что дуга x -> y является покрытой в G_1.

□

Утверждение 19: Пусть G_1 и G_2 – две эквивалентные байесовские сети, отличающиеся направлением хотя бы одной дуги, т.е. множество дуг, отличающееся направлением в двух сетях не пустое: $\Delta(G_1,G_2) := \{(x->y \in E_{G_1} \wedge y->x \in E_{G_2}) \vee (x->y \in E_{G_2} \wedge y->x \in E_{G_1})\} \neq \varnothing$.

Тогда в этом множестве $\Delta(G_1,G_2)$ есть хотя бы одна покрытая дуга.

Доказательство: (David Maxwell Chickering)

Введем обозначение: $P_v := \{u \in G_1 | u->v \in \Delta(G_1,G_2)\}$ - множество родителей узла v, с которыми v соединяет дуга из $\Delta(G_1,G_2)$.

Рассмотрим следующий алгоритм:

1. Выполним топологическую сортировку узлов G_1.
2. Выбираем узел y с минимальным номером, для которого $P_y \neq \varnothing$.
3. Выбираем узел x с максимальным номером, $x \in P_y$

Мы покажем, что дуга x -> y покрытая.

Предположим, что x -> y – не покрытая дуга в G_1. Пусть узел $z \neq x$ - родитель узла y, не являющийся родителем узла x. Если узел z не смежный с узлом x, тогда дуга x -> y участвует в v-структуре x -> y <- z в G_1, которой нет в G_2. Следовательно, согласно утверждению 17, сети G_1 и G_2 не эквивалентны, что противоречит условию. Итак, узлы z и x смежные, т.е. в G_1 есть дуга x -> z. Если в G_2 тоже была бы дуга x -> z, а также дуга z -> y, то в G_2 был бы цикл x -> z -> y -> x. Значит, или x -> z или z -> y имеют в G_2 противоположное направление, т.е.

$(\mathrm{x}->\mathrm{z} \in \Delta(G_1,G_2)) \vee (\mathrm{z}->\mathrm{y} \in \Delta(G_1,G_2))$.

Если $\mathrm{x}->\mathrm{z} \in \Delta(G_1,G_2)$, то P_z – непустое множество, а номер узла z меньше номера узла y (z – родитель y), следовательно, узел z был бы выбран вместо узла y на шаге 2 алгоритма (рис. 10).

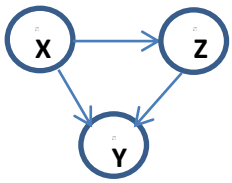


Рис. 10

Если $z->y \in \Delta(G_1,G_2)$, то оба узла x и z принадлежат P_y. Но так как номер узла z больше номера узла x (x родитель z),то узел z был бы выбран вместо узла x на шаге 3 алгоритма.

Итак, из полученного противоречия мы получаем, что дуга x -> y – покрытая дуга.

□

Утверждение 20: Две байесовские сети G_1 и G_2 эквивалентны тогда и только тогда, когда одну сеть можно получить из другой посредством $|\Delta(G_1,G_2)|$ изменений направлений покрытых дуг.

Доказательство:

->: Пусть G_1~G_2. Если G_1 и G_2 идентичны, то утверждение справедливо тривиальным образом. Пусть $\Delta(G_1,G_2) \neq \varnothing$. Тогда по утверждению 19 существует хотя бы одна покрытая дуга из

$\Delta(G_1, G_2)$. Получим граф H_1 из G_1 изменением направления этой дуги. Этот граф H_1 по утверждению 18 будет эквивалентен G_1, а число дуг в G_2, отличающихся направлением с H_1, будет меньше на единицу по сравнению с G_1: $|\Delta(G_2, H_1)| = |\Delta(G_2, G_1)| - 1$. Продолжая эту процедуру итеративно, получим последовательность эквивалентных байесовских сетей G_1=:H_0, H_1, …, H_k:=G_2, при этом соседние сети в последовательности будут отличаться только направлением одной покрытой дуги.

<-: Пусть дан набор байесовских сетей G_1=:H_0, H_1, …, H_k:=G_2, где H_i получается из H_{i-1} изменением направления одной покрытой дуги. Согласно утверждению 18 изменение направления одной покрытой дуги приводит к эквивалентной байесовской сети, т.е. H_i~H_{i-1} для всех i. Так как отношение эквивалентности транзитивно (эквивалентные сети содержат одинаковый набор высказываний о разделении), то H_0~H_k, что означает эквивалентность G_1 и G_2.

□

Из утверждения 20 следует, что при доказательстве, что какое-то свойство является общим для любых двух эквивалентных сетей G_1 и G_2, мы можем ограничиться рассмотрением случая, когда эти сети отличаются друг от друга направлением только одной покрытой дуги. Действительно, тогда это свойство будет справедливо для каждой пары H_i~H_{i-1} в цепочке G_1=:H_0, H_1, …, H_k:=G_2, и, следовательно, для G_1 и G_2.

6. Оценка параметров и структуры БС методом максимального правдоподобия

Выполнение гипотезы, соответствующей байесовской сети G, очевидно, накладывает ограничения на множество значений параметров распределения Θ_G.

Из (1) следует:

$$\mathrm{P}_\theta\left(X_i, Y, P(X_i)\right)\mathrm{P}_\theta\left(P(X_i)\right) = \mathrm{P}_\theta\left(X_i, P(X_i)\right)\mathrm{P}_\theta\left(Y, P(X_i)\right) \Rightarrow$$

$$\Rightarrow \sum_{\{X_1,\dots,X_n\}\setminus\left(\{X_i\}\bigcup Y\bigcup P(X_i)\right)} \mathrm{P}_\theta(X_1,\dots,X_1) \sum_{\{X_1,\dots,X_n\}\setminus P(X_i)} \mathrm{P}_\theta(X_1,\dots,X_1) =$$

$$= \sum_{\{X_1,\dots,X_n\}\setminus\left(\{X_i\}\bigcup P(X_i)\right)} \mathrm{P}_\theta(X_1,\dots,X_1) \sum_{\{X_1,\dots,X_n\}\setminus\left(Y\bigcup P(X_i)\right)} \mathrm{P}_\theta(X_1,\dots,X_1)$$

Здесь $\sum_Z \mathrm{P}_\theta(X_1,\dots,X_1)$ означает сумму по всем значениям переменных из $Z \subset \{X_1,\dots,X_n\}$.

Последнее равенство эквивалентно:

$$\sum_{\{X_1,\dots,X_n\}\setminus\left(\{X_i\}\bigcup Y\bigcup P(X_i)\right)} \theta_{x_1,\dots,x_n} \sum_{\{X_1,\dots,X_n\}\setminus P(X_i)} \theta_{x_1,\dots,x_n} =$$

$$= \sum_{\{X_1,\dots,X_n\}\setminus\left(\{X_i\}\bigcup P(X_i)\right)} \theta_{x_1,\dots,x_n} \sum_{\{X_1,\dots,X_n\}\setminus\left(Y\bigcup P(X_i)\right)} \theta_{x_1,\dots,x_n};$$

$$\forall X_i \in \{X_1,\dots,X_n\}, \forall Y \subset \{X_1,\dots,X_n\}: X_i \notin Y, Y \bigcap D(X_i) = \varnothing \qquad \textbf{(40)}$$

Таким образом, $\Theta_G = \left\{\theta \in \Theta \middle| \textit{справедливо } (40)\right\}$ **(41)**

С другой стороны, мы можем определить другое, специальное параметрическое множество, соответствующее гипотезе G следующим образом: для каждой i-той переменной пронумеруем

множество значений ее родителей: $\{1,...,q_i\}$, где $q_i := \prod_{s \in P(X_i)} r_s$ - число возможных значений множества родителей i-ой переменной, и определим множество $\sum_{i=1}^{n} r_i q_i$ -мерных векторов

$$\Theta^G := \left\{ \theta^G = \left(\theta^G_{i,j,k} \middle| i \in \{1,...,n\},\ j \in \{1,...,q_i\}, k \in \{1,..,r_i\} \right) \in \mathbb{R}^{\sum_{i=1}^{n} r_i q_i} \middle| \theta^G_{i,j,k} > 0, \sum_{k=1}^{r_i} \theta^G_{i,j,k} = 1 \right\} \quad \textbf{(42)}$$

В (42) каждый индекс i соответствует номеру переменной, индекс k – номеру значения i-ой переменной, а j – номеру значения родителей i-ой переменной.

Для $\theta^G \in \Theta^G$ определим:

$$\mathrm{P}_{\theta^G}\left(X_i = k \middle| P(X_i) = j \right) := \theta^G_{i,j,k} \quad \textbf{(43)}$$

Равенство (43) с учетом факторизации (3) $\mathrm{P}_{\theta^G}(X_1,...,X_n) := \prod_{i=1}^{n} \mathrm{P}_{\theta^G}\left(X_i \middle| P(X_i) \right)$ определяет распределение вероятности на $\{1,...,r_1\} \times ... \times \{1,...,r_n\}$, соответствующее гипотезе о факторизации, задаваемой байесовской сетью G. Между элементами множества Θ_G и Θ^G существует взаимно-однозначное соответствие:

$$\forall \theta \in \Theta_G : \theta_{x_1,...,x_n} = \mathrm{P}_\theta(X_1 = x_1,...,X_n = x_n) \underset{(3)}{=} \prod_{i=1}^{n} \mathrm{P}_\theta\left(X_i = x_i \middle| P(X_i) = j \right) = \prod_{i=1}^{n} \theta^G_{i,j,k} \quad \textbf{(44)}$$

Из (44) следует, что параметры распределение вероятности Θ_G, соответствующие гипотезе G, определяются параметрами из множества Θ^G. С другой стороны

$$\forall \theta^G \in \Theta^G : \theta^G_{i,j,k} = \mathrm{P}_{\theta^G}\left(X_i = x_i \middle| P(X_i) = j \right) = \frac{\mathrm{P}_{\theta^G}\left(X_i = x_i, P(X_i) = j \right)}{\mathrm{P}_{\theta^G}\left(P(X_i) = j \right)} =$$

$$= \frac{\sum_{\{X_1,...,X_n\} \setminus \left(\{X_i\} \bigcup P(X_i) \right)} \mathrm{P}_{\theta^G}(X_1 = x_1,...,X_n = x_n)}{\sum_{\{X_1,...,X_n\} \setminus \left(P(X_i) \right)} \mathrm{P}_{\theta^G}(X_1 = x_1,...,X_n = x_n)} = \frac{\sum_{\{X_1,...,X_n\} \setminus \left(\{X_i\} \bigcup P(X_i) \right)} \theta_{x_1,...,x_n}}{\sum_{\{X_1,...,X_n\} \setminus \left(P(X_i) \right)} \theta_{x_1,...,x_n}} \quad \textbf{(45)}$$

Из (45) следует, что параметры из множества Θ^G определяются при помощи параметров распределения вероятности $\theta_{x_1,...,x_n} \in \Theta_G$, для которых дополнительно должны быть выполнены ограничения (40), соответствующие гипотезе G.

Факторизацию (3) совместного распределения вероятности переменных $X_1,...,X_n$ при соблюдении гипотезы G можно записать, таким образом, как

$$\mathrm{P}_\theta(X_1,...,X_n) = \prod_{i=1}^{n} \mathrm{P}_\theta\left(X_i \middle| P(X_i) \right) = \prod_{i=1}^{n} \theta^G_{i,j,k}, \forall \theta \in \Theta_G \quad \textbf{(46)}$$

$$\theta^G_{i,j,k} = \mathrm{P}_\theta\left(X_i = k \middle| P(X_i) = j \right) \quad \textbf{(47)}$$

Пусть дана база данных $D = \left\{ X^l = \left(X^l_1,...,X^l_n \right) \middle| l = 1,...,N \right\}$, представляющая собой набор из N независимых и одинаково распределенных (для всех параметров из множества Θ) случайных величин. Пусть распределение всех X^l удовлетворяет гипотезе G, соответствующей некой

байесовской сети, т.е. справедливо (46) и (47). Оценим параметры распределения θ^G из (47) методом максимального правдоподобия. Из (46) следует, что вероятность любой одной строки из D имеет вид:

$$\mathrm{P}_{\theta^G}\left(X_1^l,...,X_n^l\right)=\prod_{i=1}^{n}\theta_{i,j,k}^G=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\left(\theta_{i,j,k}^G\right)^{1_{l_{i,j,k}}}, \quad \textbf{(48)}$$

где $1_{l_{i,j,k}}:=\begin{cases}1, X_i=k, P\left(X_i\right)=j\\ 0, \text{ в противном случае}\end{cases}$ **(49)**

С учетом независимости и одинакового распределения всех строк, получаем из (48):

$$\mathrm{P}_{\theta^G}\left(D\right)=\prod_{l=1}^{N}\mathrm{P}_{\theta^G}\left(X_1^l,...,X_n^l\right)=\prod_{l=1}^{N}\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\left(\theta_{i,j,k}^G\right)^{1_{l_{i,j,k}}}=$$
$$=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\prod_{l=1}^{N}\left(\theta_{i,j,k}^G\right)^{1_{l_{i,j,k}}}=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\left(\theta_{i,j,k}^G\right)^{\sum_{l=1}^{N}1_{l_{i,j,k}}} \quad \textbf{(50)}$$

Выражение $\sum_{l=1}^{N}1_{l_{i,j,k}}$ равно числу строк в D, для которых i-ая переменная принимает значение k, а ее родители – значение j. Введем обозначение:

$$N_{i,j,k}:=\sum_{l=1}^{N}1_{l_{i,j,k}}=\#\left\{l\in\{1,...,N\}\middle|X_i=k,P\left(X_i\right)=j\right\} \quad \textbf{(51)}$$

С учетом (51) выражение (50) принимает вид:

$$\mathrm{P}_{\theta^G}\left(D\right)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\left(\theta_{i,j,k}^G\right)^{N_{i,j,k}} \quad \textbf{(52)}$$

Соответственно, логарифм совместного распределения (функции правдоподобия) равен:

$$L\left(G,\theta^G,D\right)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}N_{i,j,k}\ln\theta_{i,j,k}^G \quad \textbf{(53)}$$

Здесь в аргументах указана структура сети G, набор ее параметров θ^G и база данных D.

Оценка максимального правдоподобия параметров θ^G должна максимизировать (53) с учетом ограничений: $\theta_{i,j,k}^G>0,\sum_{k=1}^{r_i}\theta_{i,j,k}^G=1$. Функция Лагранжа, составленная из (53) и ограничений имеет вид:

$$\mathrm{Lagr}\left(G,\theta^G,D\right)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}N_{i,j,k}\ln\theta_{i,j,k}^G+\sum_{i=1}^{n}\sum_{j=1}^{q_i}u_{i,j}\sum_{k=1}^{r_i}\theta_{i,j,k}^G, \quad \textbf{(54)}$$

где $u_{i,j}$ – коэффициенты Лагранжа, соответствующие ограничениям на параметры.

Так как (53) вогнутая по параметрам функция, а ограничения на параметры заданы линейными равенствами, то необходимым и достаточным условием для решения задачи максимизации (53) является такое допустимое значение параметра, для которого существует коэффициенты Лагранжа, при которых дифференциал от (54) по параметрам равен нулю.

$$\frac{\partial}{\partial\theta_{i,j,k}^G}\mathrm{Lagr}\left(G,\theta^G,D\right)=\frac{N_{i,j,k}}{\theta_{i,j,k}^G}+u_{i,j}=0\Rightarrow\frac{N_{i,j,k}}{\theta_{i,j,k}^G}=-u_{i,j},\forall k\Rightarrow$$
$$\Rightarrow\sum_{k=1}^{r_i}N_{i,j,k}=-u_{i,j}\underbrace{\sum_{k=1}^{r_i}\theta_{i,j,k}^G}_{=1}\Rightarrow N_{i,j}:=\sum_{k=1}^{r_i}N_{i,j,k}=-u_{i,j} \quad \textbf{(55)}$$

Здесь мы определили $N_{i,j} := \sum_{k=1}^{r_i} N_{i,j,k} = \#\left\{ l \in \{1,...,N\} \middle| P(X_i) = j \right\}$ **(56)**

Из (55) и (56) следует:

$$\frac{N_{i,j,k}}{\theta_{i,j,k}^G} = N_{i,j}, \forall k \Rightarrow \theta_{i,j,k}^G = \frac{N_{i,j,k}}{N_{i,j}}$$

Т.е. оценка максимального правдоподобия параметров распределения, соответствующих гипотезе, заданной байесовской сетью, имеет вид:

$$\hat{\theta}_{i,j,k}^G = \frac{N_{i,j,k}}{N_{i,j}} \quad \textbf{(57)}$$

Логарифм вероятности данных (53) с параметрами распределения, соответствующими оценке максимального правдоподобия (57), принимает вид:

$$L(G,D) := L\left(G, \hat{\theta}^G, D\right) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}} \quad \textbf{(58)}$$

В выражение (58) уже не входят параметры распределения (они заменены на ОМП-оценку). Значение (58) зависит только от структуры сети (через множества узлов-родителей для каждого узла) и данных. Таким образом, это выражение является логарифмом правдоподобия структуры G на данных D.

Из (58) получаем, что оценка максимального правдоподобия структуры сети имеет вид:

$$\hat{G} = \arg\max_{G \in \{БС\}} L(G,D) = \arg\max_{G \in \{БС\}} L\left(G, \hat{\theta}^G, D\right) = \arg\max_{G \in \{БС\}} \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}} \quad \textbf{(59)}$$

В (59) максимум ищется среди всех возможных байесовских сетей.

К сожалению, оценка структуры сети (59) является непрактичной, так как всегда приводит к полностью связанному графу, т.е. вероятностной модели, не содержащей утверждений об условной независимости. Другими словами, руководствуясь оценкой (59) при отыскании структуры сети, мы всегда придем к гипотезе об отсутствии отношений условной независимости между переменными. Докажем это:

Утверждение 21: Добавление дуги в байесовскую сеть увеличивает правдоподобие структуры этой сети на данных.

Доказательство:

Пусть байесовские сети G_1 и G_2 отличаются только тем, что в G_2 есть дуга от s-ого узла к p-ому узлу, а в G_1 такой дуги нет. Покажем, что $L(G_1,D) \le L(G_2,D)$, что будет означать, что добавление дуги в байесовскую сеть может только увеличить правдоподобие этой сети.

Множество родителей узла p в G_2 равно множеству родителей p в G_1 с добавлением s-ого узла, т.е. $P_{G_2}\left(\{X_p\}\right) = P_{G_1}\left(\{X_p\}\right) \bigcup \{X_s\}$. По определению, $N_{p,j,k}$ – число строк в D, для которых p-ый узел принимает значение k, а множество его родителей в G_1, т.е. $P_{G_1}\left(\{X_p\}\right)$, принимает j-ое значение. Логарифм правдоподобия сети G_1 можно записать как

$$L(G_1,D) = \underbrace{\sum_{i=1, i \ne p}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}}}_{=:C_1} + \sum_{j=1}^{q_p} \sum_{k=1}^{r_p} N_{p,j,k} \ln \frac{N_{p,j,k}}{\sum_{l=1}^{r_p} N_{p,j,l}} \quad \textbf{(60)}$$

Так как множества родителей и частоты $N_{i,j,k}$ для всех узлов кроме p-ого в обеих сетях совпадают, то совпадают и все слагаемые в выражении (58) для обеих сетей кроме i=p. Обозначим $N_{p,j,k}^m$ -

число строк в D, для которых p-ый узел принимает значение k, множество его родителей в G_1, т.е. $P_{G_1}\left(\left\{X_p\right\}\right)$, принимает j-ое значение, а s-ый узел принимает значение m. $N^m_{p,j,k}$ описывает частоты, соответствующие конфигурации родителей p-ого узла в сети G_2. Логарифм правдоподобия сети G_2 тогда можно записать как

$$L\left(G_2,D\right)=\underbrace{\sum_{i=1,i\neq p}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}N_{i,j,k}\ln\frac{N_{i,j,k}}{N_{i,j}}}_{=C_1}+\sum_{j=1}^{q_p}\sum_{m=1}^{r_s}\sum_{k=1}^{r_p}N^m_{p,j,k}\ln\frac{N^m_{p,j,k}}{\sum_{l=1}^{r_p}N^m_{p,j,l}} \qquad \textbf{(61)}$$

Вычитая (60) из (61), получаем:

$$L\left(G_2,D\right)-L\left(G_1,D\right)=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\left(\sum_{m=1}^{r_s}N^m_{p,j,k}\ln\frac{N^m_{p,j,k}}{\sum_{l=1}^{r_p}N^m_{p,j,l}}-N_{p,j,k}\ln\frac{N_{p,j,k}}{\sum_{l=1}^{r_p}N_{p,j,l}}\right) \qquad \textbf{(62)}$$

Из определения $N^m_{p,j,k}$ следует: $N_{p,j,k}=\sum_{m=1}^{r_s}N^m_{p,j,k}$, подставляя это равенство в (62), получаем:

$$L\left(G_2,D\right)-L\left(G_1,D\right)=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\left(\sum_{m=1}^{r_s}N^m_{p,j,k}\ln\frac{N^m_{p,j,k}}{\sum_{l=1}^{r_p}N^m_{p,j,l}}-\sum_{m=1}^{r_s}N^m_{p,j,k}\ln\frac{\sum_{t=1}^{r_s}N^t_{p,j,k}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}\right)=$$

$$=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\sum_{m=1}^{r_s}N^m_{p,j,k}\left(\ln\frac{N^m_{p,j,k}}{\sum_{l=1}^{r_p}N^m_{p,j,l}}-\ln\frac{\sum_{t=1}^{r_s}N^t_{p,j,k}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}\right)=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\sum_{m=1}^{r_s}N^m_{p,j,k}\ln\frac{N^m_{p,j,k}\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}{\sum_{l=1}^{r_p}N^m_{p,j,l}\sum_{t=1}^{r_s}N^t_{p,j,k}}=$$

$$=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\sum_{m=1}^{r_s}N^m_{p,j,k}\left(\ln\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s}N^t_{p,j,k}}-\ln\frac{\sum_{l=1}^{r_p}N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}\right)$$

Получаем:

$$L\left(G_2,D\right)-L\left(G_1,D\right)=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\sum_{m=1}^{r_s}\left(N^m_{p,j,k}\ln\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s}N^t_{p,j,k}}-N^m_{p,j,k}\ln\frac{\sum_{l=1}^{r_p}N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}\right)=$$

$$=\sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\left(\sum_{t=1}^{r_s}N^t_{p,j,k}\right)\sum_{m=1}^{r_s}\left(\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s}N^t_{p,j,k}}\ln\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s}N^t_{p,j,k}}-\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s}N^t_{p,j,k}}\ln\frac{\sum_{l=1}^{r_p}N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s}N^t_{p,j,l}}\right) \qquad \textbf{(63)}$$

Для фиксированных j и k обозначим

$$\alpha_m := \frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}}, \beta_m := \frac{\sum_{l=1}^{r_p} N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s} N^t_{p,j,l}}$$

Очевидно $\alpha_m, \beta_m \geq 0; \sum_{m=1}^{r_s} \alpha_m = \sum_{m=1}^{r_s} \beta_m = 1$.

Так как логарифм – вогнутая функция, то для выпуклых комбинаций, т.е. для любых $x_m; \alpha_m \geq 0; \sum_{m=1}^{r_s} \alpha_m = 1$ справедливо:

$\ln\left(\sum_{m=1}^{r_s} \alpha_m x_m\right) \geq \sum_{m=1}^{r_s} \alpha_m \ln(x_m)$, следовательно

$$0 = \ln\left(\underbrace{\sum_{m=1}^{r_s} \beta_m}_{=1}\right) = \ln\left(\sum_{m=1}^{r_s} \alpha_m \frac{\beta_m}{\alpha_m}\right) \geq \sum_{m=1}^{r_s} \alpha_m \ln\left(\frac{\beta_m}{\alpha_m}\right) = \sum_{m=1}^{r_s} \alpha_m \ln \beta_m - \sum_{m=1}^{r_s} \alpha_m \ln \alpha_m$$

Мы получили $\sum_{m=1}^{r_s} \alpha_m \ln \alpha_m \geq \sum_{m=1}^{r_s} \alpha_m \ln \beta_m$ или

$$\sum_{m=1}^{r_s}\left(\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} \ln \frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} - \frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} \ln \frac{\sum_{l=1}^{r_p} N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s} N^t_{p,j,l}}\right) \geq 0$$, следовательно, для (63) справедливо:

$$L(G_2, D) - L(G_1, D) = \sum_{j=1}^{q_p}\sum_{k=1}^{r_p}\left(\sum_{t=1}^{r_s} N^t_{p,j,k}\right) \underbrace{\sum_{m=1}^{r_s}\left(\frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} \ln \frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} - \frac{N^m_{p,j,k}}{\sum_{t=1}^{r_s} N^t_{p,j,k}} \ln \frac{\sum_{l=1}^{r_p} N^m_{p,j,l}}{\sum_{l=1}^{r_p}\sum_{t=1}^{r_s} N^t_{p,j,l}}\right)}_{\geq 0} \geq 0.$$

Итак, $L(G_2, D) \geq L(G_1, D)$.

□

Утверждение 21 доказывает неприменимость метода максимального правдоподобия для определения структуры байесовской сети.

7. Байесовский подход к оценке структуры БС

Рассмотрим байесовский подход к получению оценки структуры байесовской сети. Байесовский подход заключается в том, что мы считаем, что структура сети G и параметры сети θ^G являются случайными величинами с некими априорными распределениями. Случайная величина, соответствующая структуре G, принимает значения на дискретном конечном множестве всех возможных НАГ, а вектор параметров $\theta^G = \left(\theta^G_{i,j,k} \middle| i \in \{1,...,n\}, j \in \{1,...,q_i\}, k \in \{1,..,r_i - 1\}\right)$ - на непрерывном множестве

$$\Theta^G = \left\{\theta^G \middle| \theta^G_{i,j,k} > 0, \sum_{k=1}^{r_i - 1} \theta^G_{i,j,k} < 1\right\} \subset \mathbb{R}^{\sum_{i=1}^{n}(r_i - 1)q_i}.$$

Введем обозначения:

$\theta_{i,j}^G := \left(\theta_{i,j,1}^G, ..., \theta_{i,j,r_i-1}^G\right)$ - вектор параметров, соответствующий одной переменной и одному значению ее родителей.

$\theta_i^G := \bigcup_{j=1}^{q_i} \theta_{i,j}^G$ - вектор параметров, соответствующий одной переменной при всех возможных значениях ее родителей.

$\theta^G := \bigcup_{i=1}^{r_i} \theta_i^G$ - полный вектор параметров, соответствующий всем переменным байесовской сети при всех возможных значениях родителей.

Условная вероятность распределения i-ой переменной байесовской сети X_i, при известном значении структуры сети G, вектора ее параметров θ^G, а также значении родителей X_i имеет вид:

$$\mathrm{P}\left(X_i = k \middle| P\left(\{X_i\}\right) = j, G, \theta^G\right) = \theta_{i,j,k}^G \quad \textbf{(64)}$$

Гипотеза о факторизации, соответствующая байесовской сети принимает вид:

$$\mathrm{P}\left(X_1, ..., X_n \middle| G, \theta^G\right) = \prod_{i=1}^{n} \mathrm{P}\left(X_i \middle| P\left(\{X_i\}\right), G, \theta^G\right) = \prod_{i=1}^{n} \theta_{i,j,k}^G \quad \textbf{(65)}$$

Здесь предполагается, что i-ая переменная имеет k-ое значение, а ее родители принимают j-ое значение. В общем случае (65) можно переписать:

$$\mathrm{P}\left(X_1, ..., X_n \middle| G, \theta^G\right) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^G\right)^{1_{\{X_i = k, P(X_i) = j\}}} \quad \textbf{(66)}$$

Условная вероятность всей базы данных D (правдоподобие данных) при известной структуре и параметрах с учетом (66) равно:

$$\mathrm{P}\left(D \middle| G, \theta^G\right) = \prod_{l=1}^{N} \mathrm{P}\left(X_1^l, ..., X_n^l \middle| G, \theta^G\right) = \prod_{l=1}^{N} \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^G\right)^{1_{l,j,k}} =$$
$$= \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^G\right)^{\sum_{l=1}^{N} 1_{l,j,k}} = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^G\right)^{N_{i,j,k}} \quad \textbf{(67)}$$

7.1 Метрика Байеса-Дирихле – BD

Чтобы получить выражение для апостериорного распределения структуры сети в явном виде, введем дополнительные допущения и ограничения на априорные распределения структуры и параметров.

1. Мы предполагаем, что вектор параметров для каждой переменной при фиксированном значении родителей имеет **распределение Дирихле**:

$$\mathrm{p}\left(\theta_{i,j}^G \middle| G\right) = \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\prod_{s=1}^{r_i} \Gamma\left(\alpha_{i,j,s}\right)} 1_{\left\{\theta_{i,j,1}^G, ..., \theta_{i,j,r_i-1}^G > 0; \sum_{k=1}^{r_i-1} \theta_{i,j,k}^G < 1\right\}} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^G\right)^{\alpha_{i,j,k}-1}, \quad \textbf{(68)}$$

где $\alpha_{i,j,k} > 0$ - параметры распределения Дирихле, а $\theta_{i,j,r_i}^G = 1 - \sum_{k=1}^{r_i-1} \theta_{i,j,k}^G > 0$.

Выражение (68) описывает r_i -1 –мерную плотность распределения параметров относительно меры Лебега, т.е.

$$\mathrm{P}\left(\theta_{i,j}^{G} \in A \middle| G\right) = \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\prod_{s=1}^{r_i} \Gamma\left(\alpha_{i,j,s}\right)} \int_A 1_{\left\{\theta_{i,j,1}^{G},\ldots,\theta_{i,j,r_i-1}^{G}>0;\sum_{k=1}^{r_i-1}\theta_{i,j,k}^{G}<1\right\}} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^{G}\right)^{\alpha_{i,j,k}-1} d\theta_{i,j,1}^{G}\ldots d\theta_{i,j,r_i-1}^{G} \quad \textbf{(69)}$$

В дальнейшем мы будем предполагать, что $\theta_{i,j,k}^{G} > 0, \sum_{k=1}^{r_i-1} \theta_{i,j,k}^{G} < 1$, так что множитель $1_{\left\{\theta_{i,j,1}^{G},\ldots,\theta_{i,j,r_i-1}^{G}>0;\sum_{k=1}^{r_i-1}\theta_{i,j,k}^{G}<1\right\}}$ всегда равен 1, и мы его будем опускать.

Из свойств распределения Дирихле следует:

$$E\left(\theta_{i,j,k}^{G}\right) = \frac{\alpha_{i,j,k}}{\sum_{s=1}^{r_i} \alpha_{i,j,s}} \quad \textbf{(70)}$$

2. **Глобальная независимость** параметров – векторы параметров байесовской сети, соответствующие каждой переменной, независимы друг от друга, т.е.

$$\mathrm{p}\left(\theta^{G} \middle| G\right) = \prod_{i=1}^{n} \mathrm{p}\left(\theta_{i}^{G} \middle| G\right) \quad \textbf{(71)}$$

3. **Локальная независимость** параметров – векторы параметров байесовской сети, соответствующие каждому значению родителей одной переменной, независимы друг от друга, т.е.

$$\mathrm{p}\left(\theta_{i}^{G} \middle| G\right) = \prod_{j=1}^{q_i} \mathrm{p}\left(\theta_{i,j}^{G} \middle| G\right) \quad \textbf{(72)}$$

Из (68), (71) и (72) следует:

$$\mathrm{p}\left(\theta^{G} \middle| G\right) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma\left(\sum_{k=1}^{r_i} \alpha_{i,j,k}\right)}{\prod_{k=1}^{r_i} \Gamma\left(\alpha_{i,j,k}\right)} \prod_{k=1}^{r_i} \left(\theta_{i,j,k}^{G}\right)^{\alpha_{i,j,k}-1} \quad \textbf{(73)}$$

4. Мы также зафиксируем наше априорное незнание относительно структуры байесовской сети тем, что будем предполагать равномерное априорное распределение структуры байесовской сети, т.е.

$\mathrm{P}(G) = \frac{1}{\#\{G - \textit{байесовская сеть}\}}$, что означает

$P(G_1) = P(G_2)$ для любых двух байесовских сетей **(74)**

Из (67) и (73) следует

$$\mathrm{P}\left(D\middle|G\right)=\int \mathrm{P}\left(D,\theta^G\middle|G\right)d\theta^G=\int \mathrm{P}\left(\theta^G\middle|G\right)\mathrm{P}\left(D\middle|G,\theta^G\right)d\theta^G=$$

$$=\int 1_{\left\{\theta^G_{i,j,1},\ldots,\theta^G_{i,j,r_i-1}>0;\sum_{k=1}^{r_i-1}\theta^G_{i,j,k}<1\right\}}\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma\left(\sum_{s=1}^{r_i}\alpha_{i,j,s}\right)}{\prod_{s=1}^{r_i}\Gamma\left(\alpha_{i,j,s}\right)}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{\alpha_{i,j,k}-1}\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{N_{i,j,k}}d\theta^G= \qquad \textbf{(75)}$$

$$=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma\left(\sum_{s=1}^{r_i}\alpha_{i,j,s}\right)}{\prod_{s=1}^{r_i}\Gamma\left(\alpha_{i,j,s}\right)}\int 1_{\left\{\theta^G_{i,j,1},\ldots,\theta^G_{i,j,r_i-1}>0;\sum_{k=1}^{r_i-1}\theta^G_{i,j,k}<1\right\}}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{N_{i,j,k}+\alpha_{i,j,k}-1}d\theta^G_{i,j,1}\ldots d\theta^G_{i,j,r_i-1}$$

Из определения плотности распределения Дирихле (68) следует, что

$$\frac{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}+\alpha_{i,j,s}\right)\right)}{\prod_{s=1}^{r_i}\Gamma\left(N_{i,j,s}+\alpha_{i,j,s}\right)}1_{\left\{\theta^G_{i,j,1},\ldots,\theta^G_{i,j,r_i-1}>0;\sum_{k=1}^{r_i-1}\theta^G_{i,j,k}<1\right\}}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{N_{i,j,k}+\alpha_{i,j,k}-1}$$ - тоже плотность распределения

Дирихле с параметрами $N_{i,j,s}+\alpha_{i,j,s}$, следовательно,

$$\frac{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}+\alpha_{i,j,s}\right)\right)}{\prod_{s=1}^{r_i}\Gamma\left(N_{i,j,s}+\alpha_{i,j,s}\right)}\int 1_{\left\{\theta^G_{i,j,1},\ldots,\theta^G_{i,j,r_i-1}>0;\sum_{k=1}^{r_i-1}\theta^G_{i,j,k}<1\right\}}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{N_{i,j,k}+\alpha_{i,j,k}-1}d\theta^G_{i,j,1}\ldots d\theta^G_{i,j,r_i-1}=1,$$ следовательно,

$$\int 1_{\left\{\theta^G_{i,j,1},\ldots,\theta^G_{i,j,r_i-1}>0;\sum_{k=1}^{r_i-1}\theta^G_{i,j,k}<1\right\}}\prod_{k=1}^{r_i}\left(\theta^G_{i,j,k}\right)^{N_{i,j,k}+\alpha_{i,j,k}-1}d\theta^G_{i,j,1}\ldots d\theta^G_{i,j,r_i-1}=\frac{\prod_{s=1}^{r_i}\Gamma\left(N_{i,j,s}+\alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}+\alpha_{i,j,s}\right)\right)} \qquad \textbf{(76)}$$

Подставляя (76) в (75), получаем:

$$\mathrm{P}\left(D\middle|G\right)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma\left(\sum_{s=1}^{r_i}\alpha_{i,j,s}\right)}{\prod_{s=1}^{r_i}\Gamma\left(\alpha_{i,j,s}\right)}\frac{\prod_{s=1}^{r_i}\Gamma\left(N_{i,j,s}+\alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}+\alpha_{i,j,s}\right)\right)}=$$

$$=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma\left(\sum_{s=1}^{r_i}\alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}+\alpha_{i,j,s}\right)\right)}\prod_{s=1}^{r_i}\frac{\Gamma\left(N_{i,j,s}+\alpha_{i,j,s}\right)}{\Gamma\left(\alpha_{i,j,s}\right)} \qquad \textbf{(77)}$$

Далее получаем: $\mathrm{P}\left(G\middle|D\right)=\frac{\mathrm{P}\left(D\middle|G\right)\mathrm{P}\left(G\right)}{\mathrm{P}\left(D\right)}$, при этом от структуры сети зависит только числитель, следовательно $\mathrm{P}\left(G\middle|D\right)\propto\mathrm{P}\left(D\middle|G\right)\mathrm{P}\left(G\right)$. Так как согласно (74) априорная вероятность структуры P(G) одинакова для всех структур байесовских сетей, то $\mathrm{P}\left(G\middle|D\right)\propto\mathrm{P}\left(D\middle|G\right)$ или

$\mathrm{P}(G|D) = C \cdot \mathrm{P}(D|G)$, где C – константа, не зависящая от структуры и параметров сети. Таким образом, с учетом (77) получаем

$$\mathrm{P}(G|D) = C \cdot \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i} \left(N_{i,j,s} + \alpha_{i,j,s}\right)\right)} \prod_{s=1}^{r_i} \frac{\Gamma\left(N_{i,j,s} + \alpha_{i,j,s}\right)}{\Gamma\left(\alpha_{i,j,s}\right)} \quad \textbf{(78)}$$

Выражение (78) описывает апостериорную вероятность структуры сети.

Логарифм апостериорной вероятности сети равен:

$$\ln \mathrm{P}(G|D) = C + \sum_{i=1}^{n} \sum_{j=1}^{q_i} \left(\ln \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i} \left(N_{i,j,s} + \alpha_{i,j,s}\right)\right)} + \sum_{s=1}^{r_i} \ln \frac{\Gamma\left(N_{i,j,s} + \alpha_{i,j,s}\right)}{\Gamma\left(\alpha_{i,j,s}\right)} \right) \quad \textbf{(79)}$$

Следовательно, оценка максимальной апостериорной гипотезы (MAP) структуры байесовской сети, которая минимизирует вероятность ошибки, равна

$$\hat{G}_{BD} = \arg\max_{G} \mathrm{P}(G|D) = \arg\max_{G} \sum_{i=1}^{n} \sum_{j=1}^{q_i} \left(\ln \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i} \left(N_{i,j,s} + \alpha_{i,j,s}\right)\right)} + \sum_{s=1}^{r_i} \ln \frac{\Gamma\left(N_{i,j,s} + \alpha_{i,j,s}\right)}{\Gamma\left(\alpha_{i,j,s}\right)} \right) \quad \textbf{(80)}$$

В выражении (80) мы берем максимум по всем байесовским сетям, которые описываются множеством родителей для каждого узла и соответствующими значениями $N_{i,j,s}$ и $\alpha_{i,j,s}$.

Определение: Выражение

$$\ln \mathrm{P}(D|G) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \left(\ln \frac{\Gamma\left(\sum_{s=1}^{r_i} \alpha_{i,j,s}\right)}{\Gamma\left(\sum_{s=1}^{r_i} \left(N_{i,j,s} + \alpha_{i,j,s}\right)\right)} + \sum_{s=1}^{r_i} \ln \frac{\Gamma\left(N_{i,j,s} + \alpha_{i,j,s}\right)}{\Gamma\left(\alpha_{i,j,s}\right)} \right), \quad \textbf{(81)}$$

которое мы максимизируем при поиске оптимальной структуры сети в (80), называется **метрикой Байеса-Дирихле** (BD-метрика).

BD-метрики отличается друг от друга в зависимости от значений априорных параметров распределения $\alpha_{i,j,s}$.

Определение: Метрика Байеса-Дирихле называется **эквивалентной**, если она принимает одинаковые значения для эквивалентных сетей.

7.2 Метрика K2

Определение: Если в выражении (81) мы присвоим всем параметрам $\alpha_{i,j,s}$ значение 1, то мы получим частный случай DB-метрики, называемой **K2-метрикой**.

$$\ln \mathrm{P}\left(D|G\right)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(\ln\frac{\left(r_i-1\right)!}{\left(\sum_{s=1}^{r_i}N_{i,j,s}+r_i-1\right)!}+\sum_{s=1}^{r_i}\ln\left(N_{i,j,s}!\right)\right) \quad \textbf{(82)}$$

Следовательно, K2-оценка структуры сети принимает вид:

$$\hat{G}_{K_2}=\arg\max_{G}\mathrm{P}\left(G|D\right)=\arg\max_{G}\sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(\ln\frac{\left(r_i-1\right)!}{\left(\sum_{s=1}^{r_i}N_{i,j,s}+r_i-1\right)!}+\sum_{s=1}^{r_i}\ln\left(N_{i,j,s}!\right)\right) \quad \textbf{(83)}$$

Как показал David Heckerman и соавторы, метрика K2 не является эквивалентной: существуют эквивалентные сети, для которых (82) принимает различные значения.

7.3 Метрики, равные для эквивалентных сетей

Требование эквивалентности к метрике является логичным: если две байесовские сети эквивалентны, то множество высказываний об условной независимости у них совпадают, значит, совпадают ограничения на множества параметров распределения, а значит и сами множества допустимых параметров распределения. В этом случае, логично ожидать, что апостериорная вероятность обеих сетей будет также совпадать.

Мы получим достаточное условие для эквивалентности метрики Байеса-Дирихле (81).
Утверждение 22:
Пусть на множестве $\{1,...,r_1\}\times...\times\{1,...,r_n\}$ задана некая положительная конечная мера α . Если каждый параметр $\alpha_{i,j,k}>0$ в (81) равен мере α того события, что i-ая компонента принимает k-ое значение, а совокупность компонент, являющихся родителями i-ого узла, принимает j-ое значение, т.е. если $\alpha_{i,j,k}=\alpha\left(x_i=k,P\left(x_i\right)=j\right)$, то метрика (81) является эквивалентной.
Доказательство: (David Heckerman, Dan Geiger, David M. Chickering)
Пусть G_1 и G_2 – две эквивалентные сети, α а – положительная вероятностная мера, $\alpha_{i,j,k}=\alpha\left(x_i=k,P\left(x_i\right)=j\right)$. Покажем, что $\ln \mathrm{P}\left(D|G_1\right)=\ln \mathrm{P}\left(D|G_2\right)$.

Согласно утверждению 20, мы можем преобразовать G_1 в G_2 при помощи конечного числа изменений направлений покрытых дуг. Следовательно, достаточно доказать утверждения только для случая, когда G_1 и G_2 отличаются только направлением одной покрытой дуги. Итак, пусть в G_1 есть покрытая дуга x_i -> x_s, а в G_2 – дуга x_s -> x_i, а множества родителей всех остальных узлов в G_1 и G_2 совпадают.

Множество родителей у i-ого и s-ого узла в G_1 и G_2 также совпадают, за исключением того, что i-ый узел – родитель s-ого узла в G_1, а s-ый узел – родитель i-ого узла в G_2. Т.е. $P_{G_2}\left(i\right)=P_{G_1}\left(i\right)\bigcup\{s\}$ и $P_{G_1}\left(s\right)=P_{G_2}\left(s\right)\bigcup\{i\}$, где $P_{G_1}\left(i\right)=P_{G_2}\left(s\right)$, см. рис. 11.

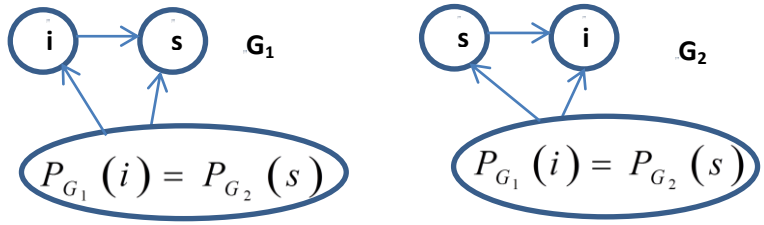


Рис. 11

Обозначим

$R := P_{G_1}(i) = P_{G_2}(s)$,

$q := \prod_{p \in R} r_p$ - число возможных значений переменных из множества R.

$N_{i,j,k}^m := \#\{l \in \{1,...,N\} | X_i = k, R = j, X_s = m\}$

$\alpha_{i,j,k}^m := \alpha(X_i = k, R = j, X_s = m)$

Очевидно,

$$N_{i,j,k} = \#\{l \in \{1,...,N\} | X_i = k, R = j\} = \sum_{m=1}^{r_s} N_{i,j,k}^m,$$

$$\alpha_{i,j,k} = \alpha(X_i = k, R = j) = \sum_{m=1}^{r_s} \alpha_{i,j,k}^m$$

Аналогично,

$N_{s,j,k}^m := \#\{l \in \{1,...,N\} | X_s = k, R = j, X_i = m\}$

$\alpha_{s,j,k}^m := \alpha(X_s = k, R = j, X_i = m)$

Также имеем $N_{s,j,k} = \sum_{m=1}^{r_i} N_{s,j,k}^m, \alpha_{i,j,k} = \sum_{m=1}^{r_i} \alpha_{s,j,k}^m$

Из сделанных определений следует:

$$N_{i,j,k}^m = \#\{l \in \{1,...,N\} | X_i = k, R = j, X_s = m\} = N_{s,j,m}^k$$

$$\alpha_{i,j,k}^m = \alpha(X_i = k, R = j, X_s = m) = \alpha_{s,j,m}^k \qquad \textbf{(84)}$$

Получим отношения метрик для двух сетей из (77), производя следующие операции:

- сокращая в числителе и знаменателе одинаковые множители, соответствующие всем узлам кроме i-ого и s-ого,
- сокращая в числителе и знаменателе одинаковые множители, соответствующие общим родителям i-узла в обеих сетях, учитывая то, что у i-ого узла в G_2 множество родителей состоит из $R \bigcup \{s\}$, а в G_1 – из R,
- сокращая в числителе и знаменателе одинаковые множители, соответствующие общим родителям s-узла в обеих сетях учитывая то, что у s-ого узла в G_1 множество родителей состоит из $R \bigcup \{i\}$, а в G_2 – из R.

$$\frac{\mathrm{P}\left(D|G_1\right)}{\mathrm{P}\left(D|G_2\right)}=$$

$$=\frac{\left(\prod_{j=1}^{q}\frac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}+\alpha_{i,j,k}\right)}{\Gamma\left(\alpha_{i,j,k}\right)}\right)\left(\prod_{j=1}^{q}\prod_{m=1}^{r_i}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)\right)}\prod_{k=1}^{r_s}\frac{\Gamma\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\alpha_{s,j,k}^{m}\right)}\right)}{\left(\prod_{j=1}^{q}\prod_{m=1}^{r_s}\frac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\alpha_{i,j,k}^{m}\right)}\right)\left(\prod_{j=1}^{q}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)\right)}\prod_{k=1}^{r_s}\frac{\Gamma\left(N_{s,j,k}+\alpha_{s,j,k}\right)}{\Gamma\left(\alpha_{s,j,k}\right)}\right)} \quad \textbf{(85)}$$

Далее из (84) получаем:

$$\prod_{j=1}^{q}\prod_{m=1}^{r_s}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\alpha_{i,j,k}^{m}\right)}=\prod_{j=1}^{q}\prod_{m=1}^{r_s}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{s,j,m}^{k}+\alpha_{s,j,m}^{k}\right)}{\Gamma\left(\alpha_{s,j,m}^{k}\right)}=$$

$$=\prod_{j=1}^{q}\prod_{m=1}^{r_i}\prod_{k=1}^{r_s}\frac{\Gamma\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\alpha_{s,j,k}^{m}\right)} \quad \textbf{(86)}$$

Подставляя (86) в (85) и сокращая множители $\prod_{j=1}^{q}\prod_{m=1}^{r_s}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\alpha_{i,j,k}^{m}\right)}$ и $\prod_{j=1}^{q}\prod_{m=1}^{r_i}\prod_{k=1}^{r_s}\frac{\Gamma\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\alpha_{s,j,k}^{m}\right)}$, получаем:

$$\frac{\mathrm{P}\left(D|G_1\right)}{\mathrm{P}\left(D|G_2\right)}=$$

$$=\frac{\left(\prod_{j=1}^{q}\frac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}+\alpha_{i,j,k}\right)}{\Gamma\left(\alpha_{i,j,k}\right)}\right)\left(\prod_{j=1}^{q}\prod_{m=1}^{r_i}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)\right)}\right)}{\left(\prod_{j=1}^{q}\prod_{m=1}^{r_s}\frac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)\right)}\right)\left(\prod_{j=1}^{q}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)\right)}\prod_{k=1}^{r_s}\frac{\Gamma\left(N_{s,j,k}+\alpha_{s,j,k}\right)}{\Gamma\left(\alpha_{s,j,k}\right)}\right)} \quad \textbf{(87)}$$

Далее, имеем:

$$\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)=\sum_{k=1}^{r_i}\left(\#\left\{l\in\{1,\ldots,N\}\middle|X_i=k,R=j\right\}+\alpha\left(X_i=k,R=j\right)\right)=$$

$$=\underbrace{\#\left\{l\in\{1,\ldots,N\}\middle|R=j\right\}}_{=:N_j}+\underbrace{\alpha\left(R=j\right)}_{=:\alpha_j}=N_j+\alpha_j$$

Аналогично,

$$\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)=\sum_{k=1}^{r_s}\left(\#\left\{l\in\{1,\dots,N\}\middle|X_s=k,R=j\right\}+\alpha\left(X_s=k,R=j\right)\right)=$$
$$=\#\left\{l\in\{1,\dots,N\}\middle|R=j\right\}+\alpha\left(R=j\right)=N_j+\alpha_j$$

Из двух последних равенств получаем $\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)=\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)$. Совершенно аналогично: $\sum_{k=1}^{r_i}\left(\alpha_{i,j,k}\right)=\sum_{k=1}^{r_s}\left(\alpha_{s,j,k}\right)$, и, следовательно

$\dfrac{\Gamma\left(\sum_{k=1}^{r_i}\left(\alpha_{i,j,k}\right)\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)\right)}=\dfrac{\Gamma\left(\sum_{k=1}^{r_s}\left(\alpha_{s,j,k}\right)\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)\right)}$. Подставляя это равенство в (87), сокращая в числителе и знаменателе $\prod_{j=1}^{q}\dfrac{\Gamma\left(\sum_{k=1}^{r_i}\left(\alpha_{i,j,k}\right)\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}+\alpha_{i,j,k}\right)\right)}$ и $\prod_{j=1}^{q}\dfrac{\Gamma\left(\sum_{k=1}^{r_s}\left(\alpha_{s,j,k}\right)\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}+\alpha_{s,j,k}\right)\right)}$, имеем:

$$\frac{\mathrm{P}\left(D|G_1\right)}{\mathrm{P}\left(D|G_2\right)}=\frac{\left(\prod_{j=1}^{q}\prod_{k=1}^{r_i}\dfrac{\Gamma\left(N_{i,j,k}+\alpha_{i,j,k}\right)}{\Gamma\left(\alpha_{i,j,k}\right)}\right)\left(\prod_{j=1}^{q}\prod_{m=1}^{r_i}\dfrac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)\right)}\right)}{\left(\prod_{j=1}^{q}\prod_{m=1}^{r_s}\dfrac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)\right)}\right)\left(\prod_{j=1}^{q}\prod_{k=1}^{r_s}\dfrac{\Gamma\left(N_{s,j,k}+\alpha_{s,j,k}\right)}{\Gamma\left(\alpha_{s,j,k}\right)}\right)} \qquad \textbf{(88)}$$

Продолжаем упрощать (88).

$$\sum_{k=1}^{r_i}\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)=\sum_{k=1}^{r_i}\left(\#\left\{l\in\{1,\dots,N\}\middle|X_i=k,R=j,X_s=m\right\}+\alpha\left(X_i=k,R=j,X_s=m\right)\right)=$$
$$=\#\left\{l\in\{1,\dots,N\}\middle|R=j,X_s=m\right\}+\alpha\left(R=j,X_s=m\right)=N_{s,j,m}+\alpha_{s,j,m}$$

Аналогично: $\sum_{k=1}^{r_i}\left(\alpha_{i,j,k}^{m}\right)=\sum_{k=1}^{r_i}\left(\alpha\left(X_i=k,R=j,X_s=m\right)\right)=\alpha\left(R=j,X_s=m\right)=\alpha_{s,j,m}$. Из двух последних равенств имеем:

$$\prod_{j=1}^{q}\prod_{m=1}^{r_s}\frac{\Gamma\left(\sum_{k=1}^{r_i}\alpha_{i,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_i}\left(N_{i,j,k}^{m}+\alpha_{i,j,k}^{m}\right)\right)}=\prod_{j=1}^{q}\prod_{m=1}^{r_s}\frac{\Gamma\left(\alpha_{s,j,m}\right)}{\Gamma\left(N_{s,j,m}+\alpha_{s,j,m}\right)}=\prod_{j=1}^{q}\prod_{k=1}^{r_s}\frac{\Gamma\left(\alpha_{s,j,k}\right)}{\Gamma\left(N_{s,j,k}+\alpha_{s,j,k}\right)} \qquad \textbf{(89)}$$

Подставляя равенство (89) в знаменатель (88), получаем единицу, следовательно:

$$\frac{\mathrm{P}\left(D|G_1\right)}{\mathrm{P}\left(D|G_2\right)}=\left(\prod_{j=1}^{q}\prod_{k=1}^{r_i}\frac{\Gamma\left(N_{i,j,k}+\alpha_{i,j,k}\right)}{\Gamma\left(\alpha_{i,j,k}\right)}\right)\left(\prod_{j=1}^{q}\prod_{m=1}^{r_i}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}^{m}\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}^{m}+\alpha_{s,j,k}^{m}\right)\right)}\right) \qquad \textbf{(90)}$$

Далее,

$$\sum_{k=1}^{r_s}\left(N_{s,j,k}^m+\alpha_{s,j,k}^m\right)=\sum_{k=1}^{r_s}\left(\#\left\{l\in\{1,...,N\}\middle|X_s=k,R=j,X_i=m\right\}+\alpha\left(X_s=k,R=j,X_i=m\right)\right)=$$
$$=\#\left\{l\in\{1,...,N\}\middle|R=j,X_i=m\right\}+\alpha\left(R=j,X_i=m\right)=N_{i,j,m}+\alpha_{i,j,m}$$

$\sum_{k=1}^{r_s}\left(\alpha_{s,j,k}^m\right)=\sum_{k=1}^{r_s}\alpha\left(X_s=k,R=j,X_i=m\right)=\alpha\left(R=j,X_i=m\right)=\alpha_{i,j,m}$, значит

$$\prod_{j=1}^{q}\prod_{m=1}^{r_i}\frac{\Gamma\left(\sum_{k=1}^{r_s}\alpha_{s,j,k}^m\right)}{\Gamma\left(\sum_{k=1}^{r_s}\left(N_{s,j,k}^m+\alpha_{s,j,k}^m\right)\right)}=\prod_{j=1}^{q}\prod_{m=1}^{r_i}\frac{\Gamma\left(\alpha_{i,j,m}\right)}{\Gamma\left(N_{i,j,m}+\alpha_{i,j,m}\right)}=\prod_{j=1}^{q}\prod_{k=1}^{r_i}\frac{\Gamma\left(\alpha_{i,j,k}\right)}{\Gamma\left(N_{i,j,k}+\alpha_{i,j,k}\right)}$$, значит (90)

обращается в единицу, т.е.

$\mathrm{P}\left(D\middle|G_1\right)=\mathrm{P}\left(D\middle|G_2\right)$, т.е. метрика (81) эквивалентная.

□

7.4 Равномерная эквивалентная метрика – BDeu

Воспользуемся утверждением 22 для того, чтобы получить одну эквивалентную метрику Байеса-Дирихле – равномерную эквивалентную метрику Байеса-Дирихле (BDeu).

Определим положительную меру на $\{1,...,r_1\}\times...\times\{1,...,r_n\}$ следующим образом:

$$\alpha\left(x_1,...,x_n\right):=\frac{1}{\prod_{i=1}^{n}r_i}\ , \qquad \textbf{(91)}$$

т.е. α - равномерная вероятностная мера, присваивающая каждому вектору одинаковое значение.

Из (91) следует: $\alpha_{i,j,k}=\alpha\left(x_i=k,P\left(x_i\right)=j\right):=\frac{1}{r_iq_i}$ **(92)**

Подставляем (92) в (81) и получаем:

$$\ln\mathrm{P}\left(D\middle|G\right)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(\ln\frac{\Gamma\left(\frac{1}{q_i}\right)}{\Gamma\left(\sum_{s=1}^{r_i}\left(N_{i,j,s}\right)+\frac{1}{q_i}\right)}+\sum_{s=1}^{r_i}\ln\frac{\Gamma\left(N_{i,j,s}+\frac{1}{q_ir_i}\right)}{\Gamma\left(\frac{1}{q_ir_i}\right)}\right) \qquad \textbf{(93)}$$

Определение: Выражение (93) описывает **равномерную эквивалентную метрику Байеса-Дирихле** (BDeu).

7.5 Асимптотическая байесовская метрика – BIC

Недостатком байесовского подхода является произвол в выборе априорного распределения параметров θ^G. Рассмотрим асимптотический подход к вычислению метрики апостериорной вероятности структуры по данным:

$$\mathrm{P}\left(D\middle|G\right)=\int\mathrm{P}\left(D\middle|G,\theta^G\right)\mathrm{P}\left(\theta^G\middle|G\right)d\theta^G \qquad \textbf{(94)}$$

Обозначим $\tilde{\theta}^G := \arg\max_{\theta^G}\left(\mathrm{P}\left(D\middle|G,\theta^G\right)\mathrm{P}\left(\theta^G|G\right)\right) = \arg\max_{\theta^G}\ln\left(\mathrm{P}\left(D\middle|G,\theta^G\right)\mathrm{P}\left(\theta^G|G\right)\right)$

Используя разложение Тейлора второго порядка для функции $q\left(\theta^G\right) := \ln\left(\mathrm{P}\left(D\middle|G,\theta^G\right)\mathrm{P}\left(\theta^G|G\right)\right)$ в точке $\tilde{\theta}^G$, с учетом $q'\left(\tilde{\theta}^G\right) = 0$ получаем:

$$q\left(\theta^G\right) \approx q\left(\tilde{\theta}^G\right) - \frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t, \qquad \textbf{(95)}$$

где $A := -H\left(\tilde{\theta}^G\right) = -q''\left(\tilde{\theta}^G\right)$ - негативное значение матрицы вторых производных (дифференциал второго порядка) в точке $\tilde{\theta}^G$.

Возводя экспоненту в степень обеих частей равенства (95), получим:

$$\mathrm{P}\left(D\middle|G,\theta^G\right)\mathrm{P}\left(\theta^G|G\right) \approx \mathrm{P}\left(D\middle|G,\tilde{\theta}^G\right)\mathrm{P}\left(\tilde{\theta}^G|G\right)\exp\left(-\frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t\right) \qquad \textbf{(96)}$$

Т.е. выражение (96) представляет собой нормальную (гауссовскую) аппроксимацию подынтегрального выражения в (94). Интегрируя обе части выражения (96) по θ^G и беря логарифм, мы получаем аппроксимацию:

$$\ln\mathrm{P}\left(D|G\right) \approx \ln\mathrm{P}\left(D\middle|G,\tilde{\theta}^G\right) + \ln\mathrm{P}\left(\tilde{\theta}^G|G\right) + \ln\int\exp\left(-\frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t\right)d\theta^G \qquad \textbf{(97)}$$

Так как $\frac{(\det A)^{1/2}}{(2\pi)^{d/2}}\exp\left(-\frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t\right)$ - плотность многомерного нормального распределения, где d – размерность матрицы A, т.е. число параметров θ^G, $d = \sum_{i=1}^{n}\left(r_i - 1\right)q_i$, то

$\frac{(\det A)^{1/2}}{(2\pi)^{d/2}}\int\exp\left(-\frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t\right)d\theta^G = 1$. Следовательно,

$\int\exp\left(-\frac{1}{2}\left(\theta^G - \tilde{\theta}^G\right)A\left(\theta^G - \tilde{\theta}^G\right)^t\right)d\theta^G = \frac{(2\pi)^{d/2}}{(\det A)^{1/2}}$. Из (97) с учетом этого получаем:

$$\ln\mathrm{P}\left(D|G\right) \approx \ln\mathrm{P}\left(D\middle|G,\tilde{\theta}^G\right) + \ln\mathrm{P}\left(\tilde{\theta}^G|G\right) + \frac{d}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(\det A\right) \qquad \textbf{(98)}$$

Приведенная техника аппроксимации интеграла (94) называется аппроксимацией методом Лапласа, а выражение (98) - аппроксимацией Лапласа. В работе Касса (Kass et al., 1988) было показано, что при определенных условиях, относительная ошибка аппроксимации имеет порядок O(1/N), где N - размер выборки. Таким образом, аппроксимация Лапласа может быть достаточно точной.

Можно получить эффективную с точки зрения вычисления (но менее точную) аппроксимацию интеграла (94), оставив в выражении (98) только те члены, которые увеличиваются с ростом N: $\ln\mathrm{P}\left(D\middle|G,\tilde{\theta}^G\right)$, который растет линейно от N и $\ln\left(\det A\right)$, который растет как $d\ln N$. Также, было показано, что с ростом N максимальная апостериорная оценка $\tilde{\theta}^G$ приближается к оценке максимального правдоподобия $\hat{\theta}^G := \arg\max_{\theta^G}\mathrm{P}\left(D\middle|G,\theta^G\right) = \arg\max_{\theta^G}\ln\mathrm{P}\left(D\middle|G,\theta^G\right)$. Таким образом, получается оценка:

$$\ln\mathrm{P}\left(D|G\right) \approx \ln\mathrm{P}\left(D\middle|G,\hat{\theta}^G\right) - \frac{d}{2}\ln N \qquad \textbf{(99)}$$

Из (58) мы знаем, что: $\ln \mathrm{P}\left(D \middle| G, \hat{\theta}^{G}\right) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}}$, а для второго слагаемого в (99) справедливо: $\frac{d}{2}\ln N = \frac{\ln N}{2}\sum_{i=1}^{n}(r_i - 1)q_i$, поэтому (99) принимает вид:

$$\ln \mathrm{P}\left(D \middle| G\right) \approx \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}} - \frac{\ln N}{2}\sum_{i=1}^{n}(r_i - 1)q_i \qquad \textbf{(100)}$$

Так как априорная вероятность структуры байесовской сети равномерна, то находя структуру с максимальным значением (99), мы максимизируем апостериорную вероятность структуры сети $\mathrm{P}\left(G \middle| D\right)$.

Оценка модели на основании выражения (99) привлекательна по нескольким причинам:

1. Она не зависит от выбора априорных распределений параметров моделей, что уменьшает субъективизм при их выборе.
2. Она интуитивно понятна: (99) состоит из слагаемого, описывающего степень «подгонки» модели к данным, а также функции «штрафа» за сложность модели. Последнее обусловлено тем, что при усложнении модели растет число d - число ее параметров.
3. Оценку (99) можно получить как оценку минимальной длины описания (Minimum Description Length - MDL), т.е. на основании совершенно других рассуждений.

Определение: метрика (99) или (100) называется **информационным критерием Байеса** (BIC).

Утверждение 23: Метрика BIC (99) является эквивалентной.
Доказательство:
Действительно, для эквивалентных сетей G_1 и G_2 множества допустимых параметров совпадают: $\Theta_{G_1} = \Theta_{G_2}$. Следовательно, совпадают максимумы вероятности на множестве параметров Θ^{G_1} и Θ^{G_2}. Значит, совпадают первые слагаемые в (99):

$$\ln \mathrm{P}\left(D \middle| G_1, \hat{\theta}^{G_1}\right) = \max_{\theta^{G_1} \in \Theta^{G_1}} \ln \mathrm{P}\left(D \middle| G_1, \theta^{G_1}\right) = \max_{\theta \in \Theta_{G_1}} \ln \mathrm{P}\left(D \middle| G_1, \theta\right) =$$

$$= \max_{\theta \in \Theta_{G_2}} \ln \mathrm{P}\left(D \middle| G_2, \theta\right) = \max_{\theta^{G_2} \in \Theta^{G_2}} \ln \mathrm{P}\left(D \middle| G_2, \theta^{G_2}\right) = \ln \mathrm{P}\left(D \middle| G_2, \hat{\theta}^{G_2}\right)$$

Вторые слагаемые в (99) также совпадают. С учетом утверждения (20) достаточно рассмотреть случай, когда G_1 и G_2 отличаются только направлением одной покрытой дуги. Пусть x_i -> x_s в G_1 и x_s -> x_i – в G_2, $R := P_{G_1}(i) = P_{G_2}(s)$, см рис. 11. Так как множества родителей всех узлов кроме i-ого и s-ого в обеих сетях совпадают, а также совпадают множества родителей i-ого и s-ого узлов кроме того, что i – родитель s в G_1 и s – родитель i в G_2, то

$$d_1 - d_2 = \left((r_i - 1)q + (r_s - 1)qr_i\right) - \left((r_i - 1)qr_s + (r_s - 1)q\right) =$$

$$= r_i q - q + r_s q r_i - q r_i - r_i q r_s + q r_s - r_s q + q = 0$$

Следовательно, d_1=d_2, значит, $\frac{d_1}{2}\ln N = \frac{d_2}{2}\ln N$, т.е.

$\ln \mathrm{P}\left(D \middle| G, \hat{\theta}^{G_1}\right) - \frac{d_1}{2}\ln N = \ln \mathrm{P}\left(D \middle| G, \hat{\theta}^{G_2}\right) - \frac{d_2}{2}\ln N$, что означает равенство метрики (99) для эквивалентных сетей.

□

8. Стратегии поиска локально-оптимальной структуры БС с использованием метрики

M. D. Chickering показал, что задача поиска оптимальной, с точки зрения некой метрики, байесовской сети – NP-полная. Поэтому на практике используются итеративные алгоритмы, использующие методы поиска лучшей по выбранной метрике сети каждый раз в некой локальной окрестности текущей структуры. Эти алгоритмы находят локальный максимум (по выбранной метрике) для структуры сети. При достижении локального максимума сеть подвергается изменениям (случайным или на основании тестов на независимость) для того, чтобы заново начать процесс поиска следующего локального максимума.

Приведенные примеры метрик (58) - ОМП, (82) – K2, (93) - BDeu, (100) – BIC являются **разложимыми** в том смысле, что представляют собой сумму по всем узлам некой функции, зависящей только от множества родителей этого узла. Удобство использования разложимых метрик заключается в том, что мы можем рассчитывать метрику для каждого узла независимо от остальных узлов. Так, если мы добавляем или удаляем одну дугу, то нам необходимо пересчитывать только одно слагаемое в метрике, соответствующее узлу, к которому эта дуга направлена, потому что только его множество родителей затрагивает эта операция. Если мы меняем направление дуги, то нам необходимо пересчитывать только два слагаемых в метрике: для узла, у которого множество родителей уменьшается и для узла, множество родителей которого увеличивается. Под **локальными операциями** с байесовской сетью мы понимаем операции:

- Добавления дуги
- Удаления дуги
- Изменения направления дуги

При этом каждая локальная операция не должна приводить к появлению цикла.

Иногда операция изменения направления дуги усложняется следующим образом: пусть между узлами X и Y есть дуга X -> Y. Мы рассматриваем операцию, состоящую из

- Изменения направления дуги X -> Y
- Изменения множества родителей X на множество $A \subset P(X) \bigcup P(Y)$
- Изменение множества родителей Y на множество $B \subset P(X) \bigcup P(Y)$

Эта усложненная операция также требует перерасчета только двух слагаемых в метрике, соответствующих двум узлам, множество родителей которых мы меняем.

Алгоритм поиска локального максимума метрики состоит из следующих шагов:

1. Рассчитываем метрику для несвязанного графа (байесовской сети без дуг).
2. Рассчитываем метрики для каждой конфигурации узла, полученной из последней конфигурации при помощи локальных операций.
3. Выбираем локальную операцию, выполнение которой приводит к наибольшему изменению суммарной метрики для участвующих в этой операции узлов (одного или двух). Если ни одна локальная операция не приводит к улучшению метрики или все такие операции приводят к появлению цикла, то останавливаем алгоритм.
4. Рассчитываем значения метрики после всех возможных локальных операций для одного или двух узлов, участвовавших в последней локальной операции. Переходим к пункту 3.

Выполнение описанного алгоритма приводит к структуре байесовской сети с локальным максимумом метрики. Т.е. все дальнейшие локальные изменения структуры узлов приводят либо к циклам, либо не улучшают метрику. Далее мы производим следующие операции для повторного старта алгоритма:

1. Для всех несмежных узлов x и y мы тестируем гипотезу об условной независимости $I(x, y|Z)$, где Z – минимальное разделяющее x и y множество. Тестировать можно, например, при помощи теста хи-квадрат с поправкой Бонферрони на условную независимость. Если эта гипотеза отвергается, то мы добавляем связь между узлами x и y.
2. Для всех смежных узлов x и y мы тестируем гипотезу об условной независимости $I(x, y|Z)$, где Z – минимальное разделяющее x и y множество, если убрать связь между x и y. Если какая-либо гипотеза не отвергается, то мы удаляем связь между узлами x и y.

После проведения описанных операций мы снова начинаем выполнение алгоритма по поиску структуры байесовской сети с максимальным значением метрики при помощи операций локального поиска.

9. Литература

1. Thomas Verma, Judea Pearl, “**Causal Networks: Semantics and Expressiveness**”, 1990
2. Thomas Verma, Judea Pearl, “Equivalence and Synthesis of Causal Models”, 1991.
3. Thomas Verma, Judea Pearl, “**An Algorithm for Deciding if a Set of Observed Independencies Has a Causal Explanation**”, 1992.
4. Judea Pearl, Thomas Verma, “The Logic of Representing Dependencies by Directed Graphs”, 1987.
5. Dan Geiger, Judea Pearl, “Logical and Algorithmic Properties of Conditional Independence and Graphical Models”, 1993.
6. R.G. Cowell, P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, “Probabilistic Networks and Expert Systems”, 1999.
7. David Maxwell Chickering, “**A Transformational Characterization of Equivalent Bayesian Network Structures**”, 1995.
8. David Maxwell Chickering, “Learning Equivalence Classes of Bayesian-Network Structures”, 2002.
9. David Maxwell Chickering, “Learning Bayesian Networks is NP-Complete”, 1996.
10. David Heckerman, Dan Geiger, David M. Chickering, “**Learning Bayesian Networks: The Combination of Knowledge and Statistical Data**”, 1995.
11. David M. Chickering , Dan Geiger, David Heckerman, “Learning Bayesian Networks: Search Methods and Experimental Results”, 1995.
12. David Heckerman, Christopher Meek, Gregory Cooper, “A Bayesian Approach to Causal Discovery”, 1997.
13. David Heckerman, “A Bayesian Approach to Learning Causal Networks”, 1995.
14. David Heckerman, Ross Shachter, “Decision-Theoretic Foundations for Causal Reasoning”, 1995.
15. David Heckerman, Dan Geiger, “Likelihoods and Parameter Priors for Bayesian Networks”, 1995.
16. David Heckerman, “A Tutorial on Learning Bayesian Networks”, 1995.
17. David Maxwell Chickering, David Heckerman, “Efficient Approximations for the Marginal Likelihood of a Bayesian Network”, 1997.
18. Luis M. de Campos, Juan M. Fernandez-Luna, J. Miguel Puerta, “**An Iterated Local Search Algorithm for Learning Bayesian Networks with Restarts Based on Conditional Independence Tests**”, 2003
19. Richard E. Neapolitan, “Learning Bayesian Networks”, 2003.
20. Ildik´o Flesch, Peter Lucas, “Markov Equivalence in Bayesian Networks”, 2007.
21. Alexandra M. Carvalho, “Scoring functions for learning Bayesian networks”, 2009.