

А.Е. ПАЩЕНКО
**ПРИМЕНЕНИЕ БАЙЕСОВСКИХ СЕТЕЙ ДОВЕРИЯ
ДЛЯ РАСЧЕТА ОТНОСИТЕЛЬНЫХ ОЦЕНОК ПОКАЗАТЕЛЕЙ
ПРОЦЕССОВ, АССОЦИИРОВАННЫХ С РИСКОМ,
В УСЛОВИЯХ ИНФОРМАЦИОННОГО ДЕФИЦИТА**

Пащенко А.Е. Применение байесовских сетей доверия для расчета относительных оценок показателей процессов, ассоциированных с риском, в условиях информационного дефицита.

Аннотация. В статье развита модель расчета относительных оценок частот событий процессов, базирующаяся на использовании байесовских сетей доверия. Использование отношения рисков и ассоциированных с ними матриц перехода является удобным и понятным инструментом для отслеживания изменений в процессах, ассоциированных с риском. Описана необходимость и подходы к дискретизации случайных величин. Рассмотрена параметризация сети для кумулятивного риска, относительных частот. Рассмотрены численные примеры, реализованные в среде GeNIe&SMILE, которая позволяет конструировать вероятностные графические модели и осуществляет вероятностный вывод.

Ключевые слова: рискованное поведение, последние эпизоды, ВИЧ-инфекция, отношение рисков, матрицы переходов, байесовские сети.

Pashchenko A.E. Application of Bayesian networks to relative estimates of process indicators associated with the risk under information deficiency.

Abstract. The article developed a model calculation of the relative frequencies of events count processes based on the use of Bayesian networks of trust. Use of the ratio of risk and the associated transition matrix is a convenient and intuitive tool for tracking changes in the processes associated with risk. Describes the need for and approaches to sampling random variables. Parameterization of the network considered for the cumulative risk of relative frequencies. Numerical examples are considered as implemented in GeNIe&SMILE which allows to design and implement graphical models probabilistic inference.

Keywords: risky behavior, last episodes, HIV-infection, risk ratio, the transition matrix, Bayesian networks.

1. Введение. Задачи моделирования социально-значимого поведения респондентов и разработки методов оценивания показателей такого поведения возникают во многих отраслях социологических, психологических, маркетинговых исследований, а также в области информационной безопасности и социального компьютеринга [1, 4, 18]. В области эпидемиологии выделяется класс задач, в которых требуется сравнить риски передачи/приобретения неизлечимых инфекционных заболеваний в различных группах в популяции или же выявить динамику уровня риска в популяции со временем, например, до и после внедрения превентивных мер. Подобные задачи лежат в основе принятия экономических и социальных решений в области общественного

здоровья, например, позволяют ответить на вопрос о том, являются ли определенные превентивные действия, направленные на модификацию поведения людей, эффективными.

Одними из численных показателей динамики риска является отношение рисков (RR, relative risk). Отношение рисков, равное 1, означает, что риск приобретения/передачи инфекции в сравниваемых группах совпадает. Кроме того целесообразно рассматривать отношение частот (frequency ratio — FR).

В настоящей работе мы концентрируем своё внимание на пуассоновской случайной модели процесса. В рассматриваемой модели основным параметром процесса является интенсивность, которая характеризует в среднем число эпизодов поведения, произошедших в течение заданного интервала времени.

Для оценки интенсивности необходимо иметь инструментарий, который позволил бы измерить (оценить) данный параметр. Таким инструментарием может являться подход к оценке интенсивности рискованного поведения, основанный на сведениях о последних эпизодах рискованного поведения [5, 9–12, 16–18, 23, 28].

Раскроем более детально суть указанного подхода [9, 11, 16] на примере. В связи с высокой стоимостью использования прямых методов получения информации о поведении индивида, требуется разработка косвенных методов [9–11, 16, 17]. Доступным источником информации о поведении являются самоотчёты индивидов, находившихся под риском заражения, об их рискованном поведении, например, интервью или опрос. Обозначим начало рассматриваемого периода времени как 0, момент интервью — t_0 ; а сам рассматриваемый период времени будем называть периодом наблюдения. Исследователя интересует число эпизодов, которые произошли в этом интервале времени; если число эпизодов поведения в определённом интервале времени известно, то, согласно модели Бэлла–Тревина [30], возможно вычислить кумулятивный риск приобретения ВИЧ в популяции.

Однако использование самоотчетов как источника информации о поведении индивида сопряжено с рядом трудностей. Пилотные исследования [9–11, 26, 27, 28] показали, что респондент может предоставить данные лишь о нескольких последних эпизодах поведения и об экстремальных значениях интервалов между последовательными эпизодами за некоторый промежуток времени. Таким образом, исследователь может получить систему ответов, состоящую из длин рекордных интервалов времени между последовательными эпизодами поведения за определённый период времени и длин интервалов между несколь-

кими последними последовательными эпизодами (включая особый интервал между последним эпизодом и моментом интервью t_0). Значения, составляющие такую систему, будут неточными и нечисловыми (в силу того, что ответы даются на естественном языке), а также неполными, так что необходимо производить соответствующую последующую обработку ответов [5, 6, 16, 17, 21].

Таким образом, информация поступает в форме ограниченного числа естественно-языковых формулировок, им свойственна нечеткость, неточность и неполнота, что не позволяет напрямую использовать известные методы для оценки интенсивности процессов. Например, хотя по определению неделя — это семь дней, однако ответ «7 дней назад» отличается от ответа «неделю назад», потому что указывает на точное число дней, прошедших с момента события, тогда как второй ответ может употребляться для указания и на 6, и на 9 дней, прошедших с момента события. В редких случаях респондент может обозначить таким образом и больший период — 10 или 11 дней.

Интерпретируя сказанное, можно говорить о том, что любой ответ описывает некоторое распределение возможных значений рассматриваемой величины, то есть является гранулой, а описываемое распределение, в свою очередь, выступает представлением неопределенности гранулярной информации (рис. 1). С теоретико-вероятностной точки зрения, такая гранула является случайной величиной; причем в рассматриваемой предметной области носитель случайной величины конечен. Однако можно рассмотреть гранулу и иначе, приняв во внимание носитель, но по-другому учесть вероятности возможных значений.

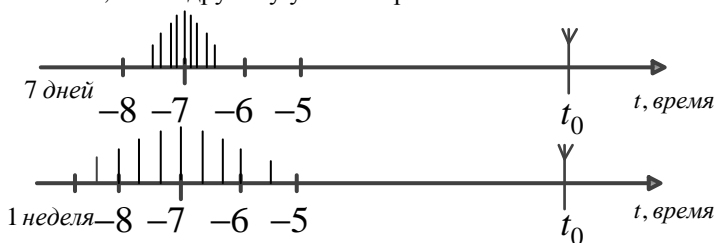


Рис. 1. Распределения, моделирующие гранулярность.

Далее в статье будем рассматривать гранулярные оценки, которые не могут быть напрямую подставлены в формулы, результирующие показатели также носят гранулярный характер, и, следовательно, не могут быть выражены по тем же формулам, что и в точном случае.

Цель настоящей статьи — описать модель расчета оценок кумулятивного риска и относительных оценок частот событий процессов с использованием байесовских сетей доверия.

2. Эффективность ВААРТ и интенсивность поведения. Одной из наиболее существенных проблем в эпидемиологии ВИЧ является приверженность пациентов лечению: насколько правильно, последовательно, строго пациент следует рекомендациям врача по применению схем лечения. Пропуск приема дозы препаратов может обусловить развитие устойчивости вируса к лекарствам. В этом случае терапия будет неэффективной, и врач будет вынужден произвести замену одного или даже всех антиретровирусных препаратов, составляющих данную схему. Кроме того, плохая приверженность режиму лечения способствует повышению риска передачи устойчивых штаммов ВИЧ.

Успех лечения непосредственно зависит от приверженности пациента лечению. Только для приверженности более 95% (то есть при крайне редких пропусках и несвоевременных приемах препаратов) наблюдается существенный прогресс в лечении ВИЧ [3].

При этом приверженность терапии зависит от [2]

- 1) уровня стресса, который испытывает пациент, и эффективности его преодоления;
- 2) веры в эффективность терапии;
- 3) состояния баланса между ощущаемой выгодой от приема терапии и наличием ограничений, которые накладывает ее прием;
- 4) наличия социальной поддержки.

Нетрудно заметить, что врач может реально повлиять на первые три из четырех составляющих, от которых зависит приверженность терапии. Это возможно благодаря использованию консультирования, то есть особым образом организованного собеседования по вопросам приверженности ВААРТ [2]. Однако врачу необходимо иметь инструмент, чтобы оценить результаты проведенных профилактических мер. Далее будет рассмотрен инструментарий, которые позволяет решить поставленные задачи.

3. Байесовские сети доверия и отношение рисков. Следует отметить, что одним из вариантов конечной задачи исследований в данной области является получение не оценки интенсивности поведения респондентов, а показателей изменения интенсивности со временем. По изменению интенсивности можно судить о рискованном поведении индивида, и, самое главное, об успешности превентивных (интервенционных) программ, направленных на модификацию поведения лю-

дей. Без положительной динамики все усилия можно было бы свести к простому мониторингу ситуации в рассматриваемых областях.

Сведение задачи оценки частот событий процессов к построению байесовской сети доверия позволит воспользоваться уже существующим мощным алгоритмическим аппаратом теории байесовских сетей доверия и свободно распространяемым программным инструментарием, например, GeNIe& SMILE [31]. В этом случае появление большого числа наблюдений, скрытых переменных, необходимость отобразить более сложные связи сводятся к частичной модификации построенной ранее модели: к появлению большего числа узлов, установлению новых связей, пересчету некоторых условных вероятностей. Можно представить элементы, входящие в модель, и связи между ними или связи между исходными данными и интересующими нас показателями [13, 19]. Далее мы будем строить байесовские сети доверия для оценки трех показателей — кумулятивного риска, относительного риска и относительных частот, на основе исходных данных о последних эпизодах поведения, а также сведений о рекордных интервалах поведения. Более подробно модели байесовских сетей доверия и область их возможных применений рассмотрены в [7, 8, 24, 25, 29, 32–34].

3.1. Формальная модель. При исследовании процесса нас интересует не общая характеристика поведения респондента «в течение жизни», а недавние события, ограниченные определенным отрезком времени «в прошлом», как следствие, мы концентрируемся на ретроспективе процесса от «настоящего момента» t_0 до момента $t_0 - T$, то есть до T единиц времени назад (рис. 2).

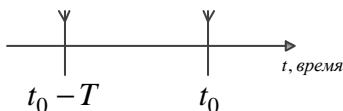


Рис.2. Исследуемый отрезок времени.

Пусть за время T произошло n_0 событий в одной группе и n_1 — событий во второй группе (например, событий до интервенции и событий после интервенции, либо событий в группе, в которой не действует значимый фактор, и в группе, где он действует). Тогда по определению частота событий выражается соответственно:

$$v_0 = \frac{n_0}{T}, \quad v_1 = \frac{n_1}{T},$$

а формула для определения относительной частоты будет выглядеть так:

$$FR = \frac{\nu_1}{\nu_0} = \frac{\frac{n_1}{T}}{\frac{n_0}{T}} = \frac{n_1}{n_0}.$$

Поскольку мы считаем рассматриваемый процесс пуассоновским, то соответствующие плотности распределения выражаются следующим образом:

$$p(n_0; \lambda_0, T) = \frac{(\lambda_0 T)^{n_0}}{n_0!} e^{-\lambda_0 T},$$
$$p(n_1; \lambda_1, T) = \frac{(\lambda_1 T)^{n_1}}{n_1!} e^{-\lambda_1 T},$$

где через λ_i ($i \in \{0, 1\}$) обозначена интенсивность поведения в соответствующих группах.

В последующих выкладках будем также использовать априорно заданную совместную функцию распределения:

$$F_{\lambda_0, \lambda_1}(l_0, l_1) = p(l_0 < \lambda_0, l_1 < \lambda_1),$$

Также введем обозначения для функций распределения:

$$F_{\lambda_0}(l_0) = p(l_0 < \lambda_0),$$

$$F_{\lambda_1}(l_1) = p(l_1 < \lambda_1).$$

Для плотностей вероятности (в тех случаях, когда они определены), зафиксируем обозначение $f_{\lambda_0, \lambda_1}(l_0, l_1)$, $f_{\lambda_0}(l_0)$, $f_{\lambda_1}(l_1)$ соответственно.

Предполагается, что априорные сведения о функциях распределения (или о плотностях распределений) могут быть получены путем обработки экспертной и (или) статистической информации. При этом экспертами могут выступить специалисты по предметной области, специалисты–математики, изучающие случайные процессы, а также группа, включающая в себя представителей обеих указанных категорий.

Если для рассматриваемого распределения существуют какие-либо предположения, дополнительные сведения или ожидаемые свойства, то формула функции распределения может быть выведена аналитически, либо снабжена алгоритмами для расчета своих значений.

Наконец, при наличии статистики по результатам полевых исследований или имитационного моделирования в качестве F_{λ_0, λ_1} , F_{λ_0} либо F_{λ_1} можно использовать эмпирические функции распределения. В случае недоступности информации о F_{λ_0, λ_1} и наличии сведений о $(F_{\lambda_0}, F_{\lambda_1})$ для построения грубых (наивных) оценок можно выразить $F_{\lambda_0, \lambda_1}(l_0, l_1) = F_{\lambda_0}(l_0)F_{\lambda_1}(l_1)$, предполагая независимость соответствующих процессов.

Рассмотрим три примера возможных совместных распределений вероятности.

Пример 1. Функция плотности распределения

$$f_{\lambda_0, \lambda_1}(l_0, l_1) = c_\mu \cdot e^{-\mu\sqrt{l_0^2 + l_1^2}}$$

является одним из способов формализовать предположения о том, что малые значения частот являются более вероятными, чем большие, и при этом не предполагать «асимметричность» двух рассматриваемых частот, где c_μ — нормировочная константа, $l_0 > 0$, $l_1 > 0$.

Пример 2. Функция плотности распределения

$$f_{\lambda_0, \lambda_1}(l_0, l_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(l_0 - l_1 - \mu)^2}{\sigma^2}}$$

формирует предположение о том, что большие различия λ_0 и λ_1 маловероятны при $\mu = 0$.

Пример 3. Функция плотности распределения

$$f_{\lambda_0, \lambda_1}(l_0, l_1) = \mu^2 \cdot e^{-\mu(l_0 + l_1)}$$

также формализует предположение об отсутствии асимметрии частот λ_0 и λ_1 , о малой вероятности их больших значений, а также удобна при разработке численных примеров в иллюстративных целях, представляет удобный аппарат для дальнейших вычислений.

3.2. Синтез байесовской сети доверия. В работах А.В. Суворовой [14, 15, 22] были предложены байесовские сети доверия, в которых были получены оценки λ , интенсивности процесса, зависящие от поступивших свидетельств. Примеры возможных вариантов построения байесовских сетей представлены на рис. 3. Следует отметить, что можно построить несколько вариантов байесовских сетей доверия для расчета одних и тех же показателей, в зависимости от того, какие связи между элементами существенны для нашей модели.

Для общности далее будем использовать рисунки вида 4д, где не конкретизируется доступное сочетание элементов информации.

В зависимости от предположений о связях и доступных сочетаниях элементов информации нас будут сначала явно или неявно интересовать следующие вероятности и условные вероятности (точнее плотности вероятностей):

- 1) $p(\lambda_i), p(t_{01}^{(i)} | \lambda_i), p(t_{12}^{(i)} | \lambda_i), p(t_{23}^{(i)} | \lambda_i),$
 $p(n_i | \lambda_i), p(t_{\min}^{(i)} | n_i, \lambda_i), p(t_{\max}^{(i)} | n_i, \lambda_i),$ где $i \in \{0,1\}$
- 2) $p(r_i | n_i), p(r_i | \lambda_i, n_i), p(\text{FR} | n_0, n_1), p(\text{FR} | n_0, n_1, \lambda_0, \lambda_1),$
 $p(\text{RR} | n_0, n_1), p(\text{RR} | n_0, n_1, \lambda_0, \lambda_1).$

Затем для построения многозначной байесовской сети нам потребуются их дискретные аналоги:

- 3) $p(\hat{\lambda}_i), p(\hat{t}_{01}^{(i)} | \hat{\lambda}_i), p(\hat{t}_{12}^{(i)} | \hat{\lambda}_i), p(\hat{t}_{23}^{(i)} | \hat{\lambda}_i),$
 $p(\hat{n}_i | \hat{\lambda}_i), p(\hat{t}_{\min}^{(i)} | \hat{n}_i, \hat{\lambda}_i), p(\hat{t}_{\max}^{(i)} | \hat{n}_i, \hat{\lambda}_i)$ где $i \in \{0,1\}.$
- 4) $p(\hat{r}_i | \hat{n}_i), p(\hat{r}_i | \hat{\lambda}_i, \hat{n}_i), p(\hat{\text{FR}} | \hat{n}_0, \hat{n}_1), p(\hat{\text{FR}} | \hat{n}_0, \hat{n}_1, \hat{\lambda}_0, \hat{\lambda}_1),$
 $p(\hat{\text{RR}} | \hat{n}_0, \hat{n}_1), p(\hat{\text{RR}} | \hat{n}_0, \hat{n}_1, \hat{\lambda}_0, \hat{\lambda}_1).$

Вероятности и условные вероятности 3) и 4) случайных элементов, получившихся в результате дискретизации соответствующих случайных величин, предназначены для внесения в качестве параметров в байесовской сети доверия (рис. 3).

Вопросы синтеза вероятностей, приведенных выше, рассмотренные в [14, 15, 22], мы будем считать решенными и станем полагаться на результаты, полученные в указанных источниках. Поэтому в дальнейшем нам необходимо будет синтезировать лишь вероятности, приведенные выше в 4).

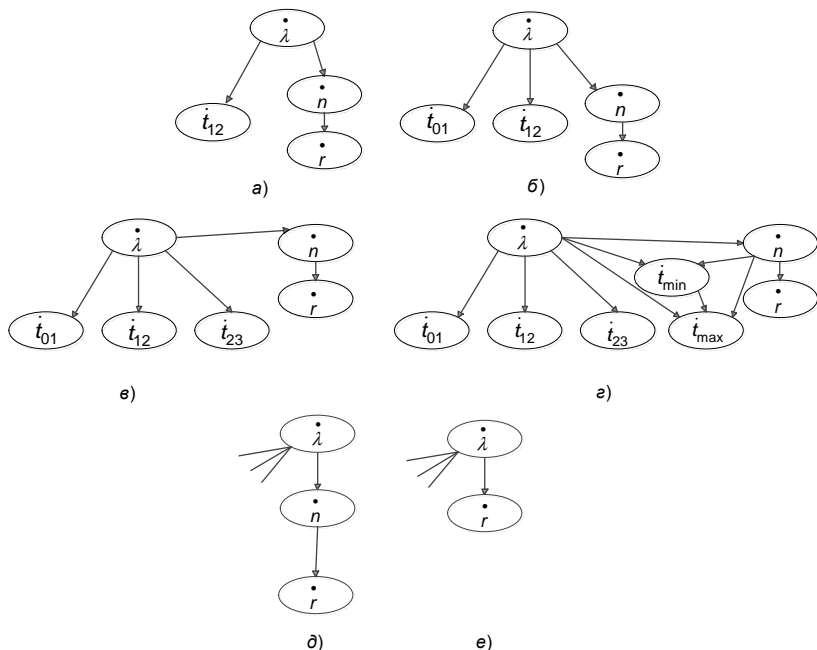


Рис. 3. Примеры байесовских сетей доверия для оценки абсолютных значений кумулятивного риска за время T .

3.3. Дискретизация случайных величин. Первым шагом решения стоящей задачи является дискретизация случайных величин r_i , n_i , FR , RR для получения дискретных случайных элементов \hat{r}_i , \hat{n}_i , \widehat{FR} , \widehat{RR} с небольшим числом возможных значений. Ограниченность числа возможных значений вызвана следующими причинами: во-первых, информация носит гранулярный характер, поэтому оценки не должны иметь высокую чувствительность, что приводит к желательности отбора сравнительно небольшого числа интервалов разбиения. Далее, еще одним желательным свойством является обзорность получаемой таблицы условных вероятностей, которая достигается при достаточно небольшом числе возможных значений. Наконец, наиболее важной причиной является ограничение размера матриц условных вероятностей: для вершины байесовской сети размер этой матрицы равен произведению числа возможных значений как случайного элемента, соответствующего этой вершине, так и всех его родителей.

Для полного определения модели в рамках теории байесовских сетей доверия необходимо вычислить тензоры условной вероятности, характеризующие переходы между узлами сети, а также выполнить дискретизацию исходных данных [20].

Значения случайной величины $\hat{\lambda}$, соответствующей интенсивности поведения, разбиваются на дизъюнктивные промежутки, каждому из которых присваивается вес $p(\hat{\lambda}^{(i)}) = p(\hat{\lambda} \in [\lambda^{(i-1)}; \lambda^{(i)}])$, $i = 1, \dots, m$ — априорная вероятность попадания интенсивности в один из полуоткрытых промежутков $[\lambda^{(i-1)}; \lambda^{(i)})$, причем $\lambda^{(0)} = 0$, $\lambda^{(m)} = +\infty$, $\lambda^{(0)} < \dots < \lambda^{(i)} < \dots < \lambda^{(m)}$. Распределение подобного вида может быть получено по статистическим данным либо по предположениям о типе непрерывного распределения случайной величины, характеризующей интенсивность. Возможные значения $\hat{t}_{j,j+1}$, $j = 0, 1, 2$, \hat{t}_{\min} и \hat{t}_{\max} также разбиваются на дизъюнктивные промежутки $t_{j,j+1}^{(1)}, \dots, t_{j,j+1}^{(k_j)}$ и $t_{\min}^{(1)}, \dots, t_{\min}^{(k_3)}$, $t_{\max}^{(1)}, \dots, t_{\max}^{(k_4)}$. Например, $t_{12}^{(1)} = (0; 0, 1)$, $t_{12}^{(2)} = [0, 1; 1)$, $t_{12}^{(3)} = [1; 7)$, $t_{12}^{(4)} = [7; 30)$, $t_{12}^{(5)} = [30; 180)$, $t_{12}^{(6)} = [180; +\infty)$.

Подробнее возможные подходы к дискретизации, а также численные примеры рассмотрены в [20].

Отметим, что вопрос подбора, поиска удачной дискретизации является отдельной и, в значительной степени, независимой исследовательской задачей, актуальной как в целом, так и в рамках рассматриваемой работы.

Следующим шагом решения является рассмотрение более сложного вопроса, а именно — расчет условных вероятностей получившихся случайных элементов. Далее мы будем последовательно рассматривать синтез недостающих параметров сети для кумулятивного риска, относительного риска и относительных частот.

4. Параметризация сети для кумулятивного риска. Для кумулятивного риска в иллюстративных целях построим пример, с конкретными численными значениями. Обратимся к расчету вероятностей, связанных со случайным элементом \hat{r}_i , соответствующим кумулятивному риску:

$$r_n = 1 - (1 - p)^n,$$

$$\begin{aligned}
 r &= 1 - \sum_{n=0}^{\infty} (1-p)^n \cdot e^{-\lambda T} \frac{(\lambda T)^n}{n!} = 1 - e^{-\lambda T} \cdot \sum_{n=0}^{\infty} \frac{[\lambda T(1-p)]^n}{n!} = \\
 &= 1 - e^{-\lambda T} \cdot e^{\lambda T(1-p)} = 1 - e^{-\lambda T p}.
 \end{aligned}$$

Заметим, что для расчета этого значения нам не требуется число эпизодов. Таким образом, мы можем исключить из рассмотрения байесовскую сеть доверия, содержащую узел с числом эпизодов процесса, поскольку интересующие нас показатели могут быть выражены только через его интенсивность. Это, в свою очередь, приведет к тому, что для расчета соответствующих оценок мы можем использовать БСД с наиболее простой структурой (рис. 3е). Далее мы будем строить условные оценки вероятности для интересующих нас параметров при условии заданной интенсивности.

Необходимо выразить условные вероятности

$$p(r \in [r_i; r_{i+1}) \mid \lambda \in [\lambda_j; \lambda_{j+1})).$$

$$\begin{aligned}
 p(r \in [r_i; r_{i+1}) \mid \lambda \in [\lambda_j; \lambda_{j+1})) &= \frac{p(r \in [r_i; r_{i+1}), \lambda \in [\lambda_j; \lambda_{j+1}))}{p(\lambda \in [\lambda_j; \lambda_{j+1}))} = \\
 &= \frac{\int_{\lambda_j}^{\lambda_{j+1}} I_{r \in [r_i; r_{i+1})}(l) dF_{\lambda_j}(l)}{\int_{\lambda_j}^{\lambda_{j+1}} dF_{\lambda}(l)}, \tag{1}
 \end{aligned}$$

где $I_{r \in [r_i; r_{i+1})}(l)$ — это характеристическая функция.

Чтобы выразить интеграл, распишем индикаторную функцию, выразив в ней r_i через λ . Соответственно, поскольку индикаторная функция принимает только значения 0 и 1, найдем те значения, где она не равна 0:

$$\begin{aligned}
 r_i &\leq 1 - e^{-\lambda T p} \leq r_{i+1}, \\
 -\frac{\ln(1 - r_i)}{T p} &\leq \lambda \leq -\frac{\ln(1 - r_{i+1})}{T p}.
 \end{aligned}$$

Тогда числитель в (1) выражается следующим образом:

$$p(r \in [r_i; r_{i+1}), \lambda \in [\lambda_j; \lambda_{j+1})) = \max \left\{ 0, \frac{\int_{\lambda_j}^{\lambda_{j+1}} dF_\lambda(l)}{\int_{\lambda_j}^{\lambda_{j+1}} dF_\lambda(l)} \right\}$$

В последнем случае берется максимум, поскольку вероятность должна быть неотрицательной. Если нижний предел интеграла оказывается больше верхнего, это означает, что

$$p(r \in [r_i; r_{i+1}) | \lambda \in [\lambda_j; \lambda_{j+1})) = 0).$$

5. Параметризация сети для относительных частот. Сформируем байесовскую сеть для расчета относительных частот. Как и в случае кумулятивного риска, необходимые вероятности не зависят от числа эпизодов, поэтому для построения достаточно предложить метод расчета вероятностей

$p(\text{FR}_i | \lambda_0^{(j)}, \lambda_1^{(k)})$, примем в рамках нашей настоящей работы

$$\text{FR} = \frac{n_1}{n_0} \approx \frac{\lambda_1}{\lambda_0}, \text{ тогда}$$

$\text{FR} \approx \frac{\lambda_1}{\lambda_0}$ — функция описания зависимости.

$$p\left(\text{FR} \in [\text{FR}_i; \text{FR}_{i+1}) | \lambda_0 \in [\lambda_0^j, \lambda_0^{j+1}), \lambda_1 \in [\lambda_1^k, \lambda_1^{k+1})\right) = \frac{\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} I_{(\text{FR} \in [\text{FR}_i; \text{FR}_{i+1}))}(l_0, l_1) dF(l_0, l_1)}{\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} dF(l_0, l_1)}.$$

Рассмотрим отдельно числитель. Он равен

$$\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} I_{(\text{FR}_i \leq \frac{l_1}{l_0} \leq \text{FR}_{i+1})}(l_0, l_1) dF(l_0, l_1) = \xi.$$

Преобразуем неравенство $FR_i \leq \frac{l_1}{l_0} \leq FR_{i+1}$ к виду

$$l_0 \cdot FR_i \leq l_1 \leq l_0 \cdot FR_{i+1},$$

тогда числитель можно переписать следующим образом:

$$\xi = \max \left\{ 0, \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \min\{\lambda_1^{k+1}, l_0 \cdot FR_{i+1}\}}^{l_1 = \max\{\lambda_1^k, l_0 \cdot FR_i\}} dF(l_0, l_1) \right\},$$

как и раньше, максимум появляется, поскольку рассматриваемое значение — вероятность, то есть не может быть отрицательной.

В общем виде

$$p \left(FR \in [FR_i; FR_{i+1}) \mid \lambda_0 \in [\lambda_0^j, \lambda_0^{j+1}), \lambda_1 \in [\lambda_1^k, \lambda_1^{k+1}) \right) = \frac{\max \left\{ 0, \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \min\{\lambda_1^{k+1}, l_0 \cdot FR_{i+1}\}}^{l_1 = \max\{\lambda_1^k, l_0 \cdot FR_i\}} dF(l_0, l_1) \right\}}{\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} dF(l_0, l_1)}.$$

Для иллюстрации выберем совместное распределение, которое позволит ограничиться аналитическими выкладками (если получающееся под знаком интеграла выражение не интегрируется в явном виде, расчеты следует производить с помощью соответствующих пакетов или систем, например MathCad). В разделе 3.1 были предложены три примера возможных совместных функций распределения. Далее рассмотрим пример 3: $f_{\lambda_0, \lambda_1}(l_0, l_1) = \mu^2 e^{-\mu(l_0+l_1)}$. Покажем, что $f_{\lambda_0, \lambda_1}(l_0, l_1)$ является функцией плотности:

$$\begin{aligned} \int_{l_1=0}^{\infty} \int_{l_0=0}^{\infty} f_{\lambda_0, \lambda_1}(l_0, l_1) dl_0 dl_1 &= \int_{l_1=0}^{\infty} \int_{l_0=0}^{\infty} \mu^2 e^{-\mu(l_0+l_1)} dl_0 dl_1 = \\ &= \int_0^{\infty} \mu e^{-\mu l_1} dl_1 \cdot \int_0^{\infty} \mu e^{-\mu l_0} dl_0 = (1 - e^{-\mu l_1}) \Big|_0^{\infty} \cdot (1 - e^{-\mu l_0}) \Big|_0^{\infty} = 1. \end{aligned}$$

Далее рассчитаем значение условной вероятности для каждого означиваний дискретных случайных элементов.

$$\int_{l_1^{(i)}}^{l_1^{(i+1)}} \int_{l_0^{(j)}}^{l_0^{(j+1)}} f(l_0, l_1) dl_0 dl_1 = \int_{l_1^{(i)}}^{l_1^{(i+1)}} \int_{l_0^{(j)}}^{l_0^{(j+1)}} \mu^2 e^{-\mu(l_0+l_1)} dl_0 dl_1 =$$

$$= \int_{l_1^{(i)}}^{l_1^{(i+1)}} \mu e^{-\mu l_1} dl_1 \int_{l_0^{(j)}}^{l_0^{(j+1)}} \mu e^{-\mu l_0} dl_0 =$$

$$= \left(1 - e^{-\mu l_1}\right) \Big|_{l_1^{(i)}}^{l_1^{(i+1)}} \left(1 - e^{-\mu l_0}\right) \Big|_{l_0^{(j)}}^{l_0^{(j+1)}} = \left(e^{-\mu l_1^{(i)}} - e^{-\mu l_1^{(i+1)}}\right) \left(e^{-\mu l_0^{(j)}} - e^{-\mu l_0^{(j+1)}}\right).$$

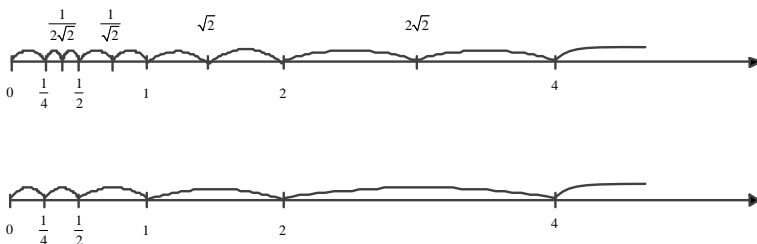


Рис. 4. Примеры возможных разбиений интервалов

Рассчитаем совместную вероятность λ_0 и λ_1 для предложенной выше дискретизации (результаты даны в табл. 1).

Таблица 1. Совместная вероятность λ_0 и λ_1

$\lambda_0 \setminus \lambda_1$	(0, 0.25]	(0.25, 0.5]	(0.5, 1]	(1, 2]	(2, 4]	(4, ∞]
(0, 0.25]	0,05	0,04	0,05	0,05	0,03	0,00
(0.25, 0.5]	0,04	0,03	0,04	0,04	0,02	0,00
(0.5, 1]	0,05	0,04	0,06	0,06	0,03	0,00
(1, 2]	0,05	0,04	0,06	0,05	0,03	0,00
(2, 4]	0,03	0,02	0,03	0,03	0,01	0,00
(4, ∞]	0,00	0,00	0,00	0,00	0,00	0,00

6. Параметризация сети для относительных рисков защищенности. Наконец, рассчитаем условную вероятность для относительных рисков и относительной защищенности. Точно так же, как и в предыдущем случае, нам достаточно будет рассмотреть лишь сеть с простейшей структурой, в которой вероятности значения относительного риска зависят только от интенсивности.

$$RR = \frac{r_1}{r_0} = \frac{1 - e^{-\lambda_1 Tp}}{1 - e^{-\lambda_0 Tp}}.$$

Кроме того, предлагается рассмотреть отношение защищенности (safety ratio, SR). Такой показатель интересен и сам по себе (например, если в задаче нужно рассчитать не уровень риска, а уровень защищенности):

$$\begin{aligned} SR &= \frac{1 - r_1}{1 - r_0} = \frac{e^{-\lambda_1 Tp}}{e^{-\lambda_0 Tp}} = e^{-(\lambda_1 - \lambda_0)Tp}, \\ SR_i &\leq SR \leq SR_{i+1}, \\ SR_i &\leq e^{-(\lambda_1 - \lambda_0)Tp} \leq SR_{i+1}, \\ \ln SR_i &\leq -\ln(\lambda_1 - \lambda_0)Tp \leq \ln SR_{i+1}, \\ -\frac{\ln SR_i}{Tp} &\leq (\lambda_1 - \lambda_0) \leq -\frac{\ln SR_{i+1}}{Tp}, \\ \lambda_0 - \frac{\ln SR_{i+1}}{Tp} &\leq \lambda_1 \leq \lambda_0 - \frac{\ln SR_i}{Tp}. \end{aligned}$$

При наличии такого соотношения

$$\begin{aligned} &P\left(SR \in [SR_i, SR_{i+1}) \mid \lambda_0 \in [\lambda_0^j, \lambda_0^{j+1}), \lambda_1 \in [\lambda_1^k, \lambda_1^{k+1})\right) = \\ &= \frac{\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \min\{\lambda_1^{k+1}, l_0 - \frac{\ln(SR_i)}{Tp}\}}^{\int_{l_1 = \max\{\lambda_1^k, l_0 - \frac{\ln(SR_i)}{Tp}\}}} dF(l_0, l_1)}{\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} dF(l_0, l_1)}. \end{aligned}$$

Как и в случае относительных частот, достаточным для рассмотрения является БСД со сравнительно простой структурой (рис 4.а–д).

Рассмотрим пример и рассчитаем SR

$$p(\text{SR} \mid \lambda_0, \lambda_1) = \frac{\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \min\{\lambda_1^{k+1}, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\}}^{\int_{l_1 = \max\{\lambda_1^k, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\}}} dF(l_0, l_1)}{\iint_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} dF(l_0, l_1)}. \quad (2)$$

Обозначим

$$b = \max\{\lambda_1^k, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\},$$

$$a = \min\{\lambda_1^{k+1}, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\}.$$

Рассмотрим $f_{\lambda_0, \lambda_1}(l_0, l_1) = \mu^2 e^{-\mu(l_0 + l_1)}$, и пусть $\frac{\ln(\text{SR}_{i+1})}{T_p} = x_{i+1}$,

$\frac{\ln(\text{SR}_i)}{T_p} = x_i$, тогда рассмотрим числитель

$$\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \max\{\lambda_1^k, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\}}^{l_1 = \min\{\lambda_1^{k+1}, l_0 - \frac{\ln(\text{SR}_i)}{T_p}\}} dF(l_0, l_1) = \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_b^a \mu e^{-\mu l_0} \mu e^{-\mu l_1} d(l_0) d(l_1) =$$

$$= \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} e^{-\mu l_0} \cdot (-e^{-\mu l_1}) \Big|_b^a dl_0.$$

Рассмотрим 4 варианта вычислений формулы (2), выделяющихся возможными сочетаниями пределов интегрирования.

Вариант 1. $\left[\lambda_1^{k+1} > l_0 - \frac{\ln(\text{SR}_i)}{T_p}; \lambda_1^k < l_0 - \frac{\ln(\text{SR}_{i+1})}{T_p} \right]$

Преобразуем числитель

$$\begin{aligned}
& \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} \cdot (e^{-\mu(l_0 - x_{i+1})} - e^{-\mu(l_0 - x_i)}) dl_0 = \\
& = \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} (\mu e^{-\mu(2l_0 - x_{i+1})} - \mu e^{-\mu(2l_0 - x_i)}) dl_0 = \\
& = -\frac{1}{2} (e^{-\mu(2l_0 - x_{i+1})} - e^{-\mu(2l_0 - x_i)}) \Big|_{\lambda_0^j}^{\lambda_0^{j+1}} = -\frac{1}{2} e^{-2\mu l_0} (e^{\mu x_{i+1}} - e^{\mu x_i}) \Big|_{\lambda_0^j}^{\lambda_0^{j+1}} = \\
& = \frac{1}{2} (e^{-2\mu \lambda_0^j} - e^{-2\mu \lambda_0^{j+1}}) (e^{\mu x_{i+1}} - e^{\mu x_i}).
\end{aligned}$$

Знаменатель соответственно

$$\begin{aligned}
& \frac{1}{2} (e^{-\mu \lambda_1^j} - e^{-\mu \lambda_1^{j+1}}) (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}}) \\
p(\text{SR} \mid \lambda_0, \lambda_1) &= \frac{(e^{-2\mu \lambda_0^j} - e^{-2\mu \lambda_0^{j+1}}) (e^{\mu x_{i+1}} - e^{\mu x_i})}{(e^{-\mu \lambda_1^j} - e^{-\mu \lambda_1^{j+1}}) (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})}.
\end{aligned}$$

Вариант 2.

$$\left[\lambda_1^{k+1} < l_0 - \frac{\ln(SR_i)}{Tp}; \lambda_1^k < l_0 - \frac{\ln(SR_{i+1})}{Tp} \right].$$

Числитель будет иметь вид:

$$\begin{aligned}
& \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} (-e^{-\mu l_1}) \Big|_{l_1 = l_0 - x_{i+1}}^{\lambda_1^{k+1}} = \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} \cdot (e^{-\mu(l_1 - x_{i+1})} - e^{-\mu \lambda_1^{k+1}}) dl_0 = \\
& = \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} (\mu e^{-\mu(2l_0 - x_{i+1})} - \mu e^{-\mu l_0} \cdot e^{-\mu \lambda_1^{k+1}}) dl_0 = \\
& = -\frac{1}{2} (e^{-\mu(2l_0 - x_{i+1})} + \mu e^{-\mu l_0} \cdot e^{-\mu \lambda_1^{k+1}}) \Big|_{\lambda_0^j}^{\lambda_0^{j+1}} = \\
& = \frac{1}{2} (e^{-\mu(2\lambda_0^{j+1} - x_{i+1})} - e^{-\mu(2\lambda_0^j - x_{i+1})}) + (e^{-\mu \lambda_0^{j+1}} - e^{-\mu \lambda_0^j}) \cdot e^{-\mu \lambda_1^{k+1}} \\
p(\text{SR} \mid \lambda_0, \lambda_1) &= \frac{(e^{-\mu(2\lambda_0^{j+1} - x_{i+1})} - e^{-\mu(2\lambda_0^j - x_{i+1})}) + (e^{-\mu \lambda_0^{j+1}} - e^{-\mu \lambda_0^j}) \cdot e^{-\mu \lambda_1^{k+1}}}{(e^{-\mu \lambda_1^j} - e^{-\mu \lambda_1^{j+1}}) (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})}.
\end{aligned}$$

Вариант 3.

$$\left[\lambda_1^{k+1} < l_0 - \frac{\ln(SR_i)}{Tp}; \lambda_1^k > l_0 - \frac{\ln(SR_{i+1})}{Tp} \right].$$

Числитель будет иметь вид:

$$\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} (-e^{-\mu l_1}) \Bigg|_{\lambda_1^k}^{\lambda_1^{k+1}} = (e^{-\mu \lambda_1^k} - e^{-\mu \lambda_1^{k+1}})(e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})$$

$$p(\text{SR} | \lambda_0, \lambda_1) = \frac{2(e^{-\mu \lambda_1^k} - e^{-\mu \lambda_1^{k+1}})(e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})}{(e^{-\mu \lambda_1^j} - e^{-\mu \lambda_1^{j+1}})(e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})}.$$

Вариант 4.

$$\left[\lambda_1^{k+1} > l_0 - \frac{\ln(SR_i)}{Tp}; \lambda_1^k > l_0 - \frac{\ln(SR_{i+1})}{Tp} \right].$$

Числитель будет иметь вид:

$$\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} (-e^{-\mu l_1}) \Bigg|_{\lambda_1^k}^{l_1 = l_0 - x_i} dl_0 =$$

$$= e^{-\mu \lambda_1^k} (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}}) - \int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \mu e^{-\mu l_0} \cdot e^{-\mu(l_0 - x_i)} dl_0 =$$

$$= e^{-\mu \lambda_1^k} (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}}) + \frac{1}{2} \left(e^{-\mu(2\lambda_0^{j+1} - x_i)} - e^{-\mu(2\lambda_0^j - x_i)} \right),$$

$$p(\text{SR} | \lambda_0, \lambda_1) = \frac{e^{-\mu \lambda_1^k} (e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}}) + \frac{1}{2} \left(e^{-\mu(2\lambda_0^{j+1} - x_i)} - e^{-\mu(2\lambda_0^j - x_i)} \right)}{(e^{-\mu \lambda_1^j} - e^{-\mu \lambda_1^{j+1}})(e^{-\mu \lambda_0^j} - e^{-\mu \lambda_0^{j+1}})}.$$

Формула для расчета условных вероятностей для узла БСД, соответствующего $RR = \frac{1 - e^{-\lambda_i Tp}}{1 - e^{-\lambda_0 Tp}}$, представлена следующим образом:

$$RR_i \leq \frac{1 - e^{-\lambda_i Tp}}{1 - e^{-\lambda_0 Tp}} \leq RR_{i+1},$$

$$RR_i(1 - e^{-\lambda_0 Tp}) \leq 1 - e^{-\lambda_i Tp} \leq RR_{i+1}(1 - e^{-\lambda_0 Tp}),$$

$$1 - RR_{i+1}(1 - e^{-\lambda_0 Tp}) \leq e^{-\lambda_i Tp} \leq 1 - RR_i(1 - e^{-\lambda_0 Tp}),$$

$$-\frac{\ln(1 - \text{RR}_{i+1}(1 - e^{-\lambda_0 T_p}))}{T_p} \leq \lambda_1 \leq -\frac{\ln(1 - \text{RR}_i(1 - e^{-\lambda_0 T_p}))}{T_p}.$$

Получим, что

$$\text{RR} = \frac{\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 \in [\lambda_0^j, \lambda_0^{j+1})} \frac{\ln(1 - \text{RR}_i(1 - e^{-\lambda_0 T_p}))}{T_p} dF(l_0, l_1)}{\iint_{\substack{l_0 \in [\lambda_0^j, \lambda_0^{j+1}) \\ l_1 \in [\lambda_1^k, \lambda_1^{k+1})}} dF(l_0, l_1)}.$$

Аналогичным образом выведем формулу для расчета условных вероятностей для узла БСД, соответствующего $\text{FR} \approx \frac{\lambda_1}{\lambda_0}$

$$\begin{aligned} \text{SR}_i &\leq \text{SR} \leq \text{SR}_{i+1} \\ \lambda_0 \text{FR}_i &\leq l_0 \leq \lambda_0 \text{FR}_{i+1} \\ p(\text{FR} | \lambda_0, \lambda_1) &= \frac{p(\text{FR}, \lambda_0, \lambda_1)}{p(\lambda_0, \lambda_1)} = \frac{\int_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} \int_{l_1 = \min\{\lambda_1^{j+1}, l_0 \text{FR}_{i+1}\}}^{l_1 = \max\{\lambda_1^j, l_0 \text{FR}_{i+1}\}} dF(l_0, l_1)}{\iint_{l_0 \in [\lambda_0^j, \lambda_0^{j+1})} dF(l_0, l_1)}. \end{aligned}$$

7. Численные примеры. Рассмотрим численный пример, позволяющий анализировать кумулятивный риск $r = 1 - e^{-\lambda T_p}$. В данном случае считаем $f_\lambda(l) = \mu e^{-\mu l}$.

$$p(r | \lambda) = \frac{\int_{\lambda_j}^{\lambda_{j+1}} \min\left\{\lambda_{j+1}, \frac{\ln(1-r_{i+1})}{T_p}\right\} \int_{\max\left\{\lambda_j, \frac{\ln(1-r_i)}{T_p}\right\}} dF_\lambda(l)}{\int_{\lambda_j}^{\lambda_{j+1}} dF_\lambda(l)} = \frac{e^{-\mu \cdot \max\left\{\lambda_j, \frac{\ln(1-r_i)}{T_p}\right\}} - e^{-\mu \cdot \min\left\{\lambda_{j+1}, \frac{\ln(1-r_{i+1})}{T_p}\right\}}}{e^{-\mu \lambda_j} - e^{-\mu \lambda_{j+1}}}.$$

Таблица 2. Априорное распределение риска

$(r \lambda)$	[0; 0,25]	[0,25; 0,5]	[0,5; 1]	[1; 2]	[2; 4]	[4; ∞]
[0; 0,05]	1	0,768081	0	0	0	0
[0,05; 0,2]	0	0,231919	1	0,089233	0	0
[0,2; 0,5]	0	0	0	0,422145	0	0
[0,5; 1]	0	0	0	0,488622	1	1

В среде GeNIe&SMILE [31] была построена модель байесовский сети доверия. Данная программная среда предоставляет возможности по конструированию графических моделей и позволяет осуществлять вероятностный вывод в данной модели.

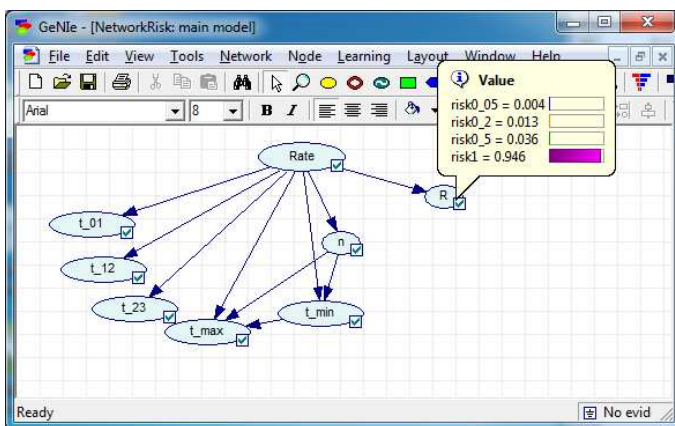


Рис. 4. Первоначальное распределение риска

В рассматриваемом примере априорные оценки кумулятивного риска оказались высоки для высоких значений риска, поскольку было выбрано априорное распределение интенсивности вида $f_{\lambda}(l) = \mu e^{-\mu l}$, (где $\mu = 0.1$ — экспертно определенный параметр распределения) экспоненциально. Рассматриваемый временной интервал $T = 180$ (дней), вероятность передачи инфекции $p = 0.007$.

Предположим, что далее были получены следующие сведения: последний эпизод произошел 7 дней назад, предпоследний эпизод произошел за 1 день до этого, предпредпоследний эпизод произошел за 7 дней до этого. Эти сведения естественным образом образуют свидетельства для построенной сети, пропация которых позволит рассчитать апостериорные оценки вероятности риска передачи.

После пропагации полученных свидетельств значения апостериорного распределения будут иметь вид, представленный на рисунке 5. (распределение 1)

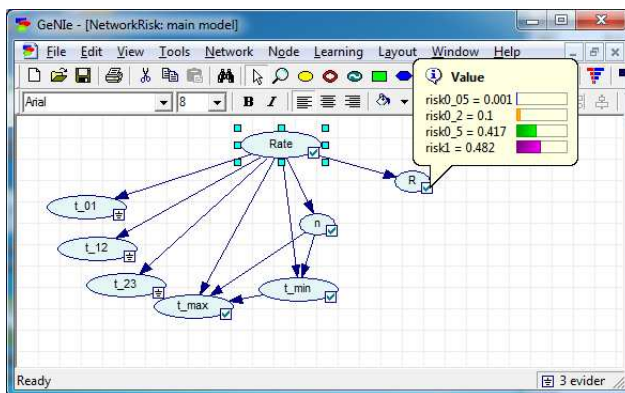


Рис. 5. Распределение риска 1.

Сравним эти результаты со случаем, когда поступили похожие, но, все же, отличные сведения, а именно: последний эпизод произошел 7 дней назад, предпоследний эпизод произошел за 7 дней до этого, предпредпоследний эпизод произошел за 7 дней до этого. В этом случае апостериорные значения соответствующих вероятностей будут иметь вид, представленный на рисунке 6. (распределение 2). Как видно из рисунка, апостериорные вероятности оказались достаточно похожи на те, которые были в предыдущем случае.

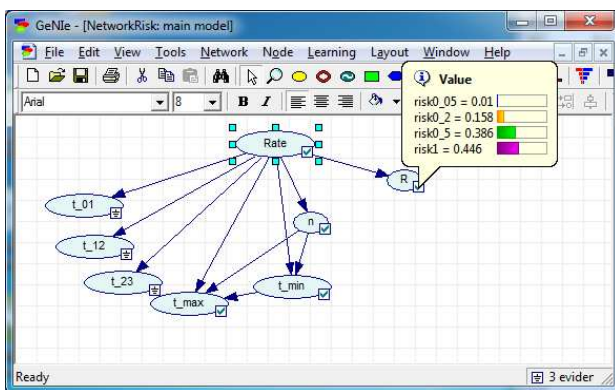


Рис. 6. Распределение риска 2.

Теперь рассмотрим пример, в котором полученные сведения сильно отличаются от двух предыдущих. Так, пусть мы получили следующие сведения для той же модели: последний эпизод произошел 30 дней назад, предпоследний эпизод произошел за 280 дней до этого, предпредпоследний эпизод произошел за 30 дней до этого. В этом случае апостериорные значения будут иметь вид, представленный на рисунке 7. (распределение 3). Как мы видим из полученных результатов, вероятностные оценки рисков существенно изменились.

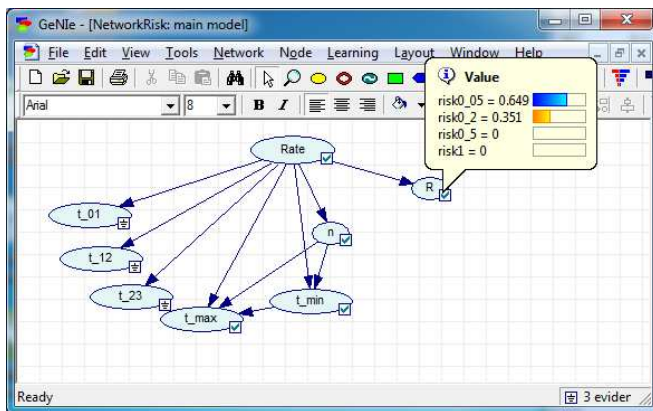


Рис. 7. Распределение риска 3.

8. Заключение. В статье предложен метод, развивающий разработанный ранее подход А.В. Суворовой [14, 15, 20, 22]. Рассмотренный метод позволяет получить относительную оценку интенсивности за счет агрегирования данных и знаний в форме байесовских сетей доверия, а затем — распространения влияния (пропагации) свидетельств, отражающих те сведения, которые были получены апостериорно. Предложенный метод является существенным вкладом в обработку данных с неопределенностью в условиях информационного дефицита. Полученные результаты позволяют разработать комплекс программ, обеспечивающих синтез байесовских сетей доверия на основе полученных оценок интенсивности процесса, зависимых от исходных свидетельств с последующей автоматизацией вычислений для кумулятивного риска, относительного риска и относительных частот, и отношений шансов.

Литература

1. Азаров А.А., Бродовская Е.В., Бубнов А.Ю., Глазков А.А., Гибадулин Р.Я., Дмитриева О.В., Долгоруков А.М., Жданов Р.И., Мусина В.Ф., Нечаев В.Д., Огнев А.С., Сироткин А.В., Суворова А.В., Титов В.В., Тулупьев А.Л., Тулупьева Т.В., Фильченков А.А., Юсупов Р.М. Predictor Mining: применение методов интеллектуального анализа данных в задачах социального компьютинга // Труды СПИИРАН. 2013. Вып. 3(26). С. 136–161.
2. Беляева ВВ., Покровский ВВ., Кравченко АВ. Консультирование при ВИЧ-инфекции. Пособие для врачей различных специальностей. М., 2003. 77 с.
3. В.В. Беляева Приверженность высокоактивной противоретровирусной терапии при вич-инфекции. М.: Медицина для Вас, 2005. 52 с.
4. Ванюшичева О.Ю., Тулупьева Т.В., Пашенко А.Е., Тулупьев А.Л., Азаров А.А. Количественные измерения поведенческих проявлений уязвимостей пользователя, ассоциированных с социоинженерными атаками. // Труды СПИИРАН. 2011. Вып. 19. С. 34–47.
5. Зельтерман Д., Суворова А.В., Пашенко А.Е., Мусина В.Ф., Тулупьев А.Л., Тулупьева Т.В., Гро Л., Хаймер Р. Диагностика регрессионных уравнений в анализе интенсивности рискованного поведения по его последним эпизодам // Труды СПИИРАН. 2011. № 17. С. 33–46.
6. Зельтерман Д., Тулупьев А.Л., Суворова А.В., Пашенко А.Е., Мусина В.Ф., Тулупьева Т.В., Красносельских Т.В., Гро Л., Хаймер Р. Обработка систематической ошибки, связанной с длиной временных интервалов между интервью и последним эпизодом в гамма-пуассоновской модели поведения // Труды СПИИРАН. 2011. № 16. С. 160–185.
7. Мусина В.Ф. Байесовские сети доверия как вероятностная графическая модель для оценки медицинских рисков // Труды СПИИРАН. 2013. № 24. С. 135–151.
8. Мусина В.Ф. Байесовские сети доверия как вероятностная графическая модель для оценки экономических рисков // Труды СПИИРАН. 2013. № 25. С. 235–254.
9. Пашенко А.Е. Идентификация интенсивности пуассоновского процесса, моделирующего поведение респондента, в условиях дефицита информации // Информационно-измерительные и управляющие системы. 2009. № 4. Т. 7. С. 45–48.
10. Пашенко А.Е., Тулупьев А.Л., Николенко С.И. Моделирование заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Известия высших учебных заведений: Приборостроение. 2006. № 8. С. 33–34.
11. Пашенко А.Е., Тулупьев А.Л., Тулупьева Т.В. Аprobация блока вопросов и обработка ответов о последних эпизодах рискованного поведения ВИЧ-инфицированных // X Санкт-Петербургская международная конференция .Региональная информатика–2006 (РИ-2006).: Труды. СПб., 2007. С. 323–326.
12. Степанов Д.В., Мусина В.Ф., Суворова А.В., Тулупьев А.Л., Сироткин А.В., Тулупьева Т.В. Функция правдоподобия с гетерогенными аргументами в идентификации пуассоновской модели рискованного поведения в случае информационного дефицита // Труды СПИИРАН. 2012. № 4(23). С. 157–184.
13. Суворова А.В. Моделирование социально-значимого поведения по сверхмалой

- неполной совокупности наблюдений // Информационно-измерительные и управляющие системы. 2013. №9, т. 11. С. 34–38.
14. Суворова А.В. Анализ данных о рекордных интервалах между эпизодами социально-значимого поведения в рамках модели, построенной на основе байесовских сетей доверия // Научная сессия НИЯУ МИФИ-2013. Аннотации докладов. В 3 т. Т.2 Проблемы фундаментальной науки. Стратегические информационные технологии. М.: НИЯУ МИФИ, 2013. С. 329.
 15. Суворова А.В. Представление пуассоновской модели социально-значимого поведения в виде байесовской сети доверия // Современные проблемы математики. Тезисы Международной (44-я Всероссийской) молодежной школы-конференции. Екатеринбург: Институт математики и механики УрО РАН, 2013. С. 333–335.
 16. Суворова А.В., Лавренов А.В., Тулупьева Т.В., Тулупьев А.Л., Пашенко А.Е. Моделирование социально значимого поведения респондентов: аналитическая и численная оценки интенсивности в окрестности интервью при информационном дефиците // Труды СПИИРАН. 2012. № 1(20). С. 101–115.
 17. Суворова А.В., Тулупьев А.Л., Пашенко А.Е., Тулупьева Т.В., Красносельских Т.В. Анализ гранулярных данных и знаний в задачах исследования социально значимых видов поведения // Компьютерные инструменты в образовании. 2010. № 4. С. 30–38.
 18. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В. Идентификация параметров вероятностных графических моделей как моделей нечетких стохастических процессов, представляющих социально-значимое поведение индивидов // Научная сессия НИЯУ МИФИ-2013. Аннотации докладов. В 3 т. Т.2 Проблемы фундаментальной науки. Стратегические информационные технологии. М.: НИЯУ МИФИ, 2013. С. 328.
 19. Суворова А.В., Тулупьева Т.В., Сироткин А.В. Построение вероятностных графических моделей социально-значимого поведения с учетом особенностей гранулярных данных о рекордных интервалах между его эпизодами // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VII-й Международной научно-технической конференции (Коломна, 20–22 мая 2013 г.). М.: Физматлит. Т.3. С. 1376–1385
 20. Суворова А.В., Мусина В.Ф., Тулупьева Т.В., Тулупьев А.Л., Красносельских Т.В., Фильченков А.А., Азаров А.А., Абдала Н. Автоматизированный инструментарий для опроса респондентов об эпизодах рискованного поведения: первичный анализ результатов применения // Труды СПИИРАН. 2013. Вып. 3(26). С. 175-193
 21. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Пашенко А.Е. Эвристическая оценка интенсивности поведения по рекордным интервалам между эпизодами: обработка неточности ответов респондентов //XV Международная конференция по мягким вычислениям и измерениям. SCM-2012. (25-27 июня 2012 г. Санкт-Петербург). Сборник докладов. 2012. Т. 2. СПб.: Изд-во СПбГЭТУ „ЛЭТИ“. С. 101–104.
 22. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В. Применение байесовских сетей доверия для моделирования угрозообразующего поведения индиви-

- да по неполным и неточным данным // Тринадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2012 (16–20 октября 2012 г., г. Белгород). Труды конференции. Т. 3. Белгород: Изд-во БГТУ, 2012. С. 292–299.
23. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В., Пашенко А.Е. Вероятностные графические модели социально-значимого поведения индивида, учитывающие неполноту информации // Труды СПИИРАН. 2012. Вып. 22. С. 101–112.
 24. Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
 25. Тулупьев А.Л., Сироткин А.В., Николенко С.И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петербургского ун-та, 2009. 400 с.
 26. Тулупьева Т.В., Пашенко А.Е., Мусина В.Ф., Тулупьев А.Л., Азаров А.А., Жук С.Н., Казакова О.С., Красносельских Т.В., Сироткин А.В., Суворова А.В., Фильченков А.А. Отчет о научно-исследовательской работе .Классификация ответов респондентов о последних эпизодах рискованного поведения и косвенная оценка его интенсивности. (заключительный),инвентарный № 02201259425 от 2012.06.26, по теме .Моделирование и измерение количественных характеристик ВИЧ-рискованного поведения на основе обработки ответов респондентов., регистрационный № 01201262070. СПб.: СПИИРАН. 2012. 34 с. (Депонировано в ЦИТиС.)
 27. Тулупьева Т.В., Пашенко А.Е., Мусина В.Ф., Тулупьев А.Л., Жук С.Н., Азаров А.А., Суворова А.В., Сироткин А.В., Фильченков А.А. Отчет о научно-исследовательской работе .Опросный инструментарий для выявления особенностей рискованного поведения в контексте адаптивных стилей и анализ результатов пилотного исследования., инвентарный № 02201259423 от 2012.06.26, по теме "Взаимосвязь адаптивных стилей ВИЧ-инфицированных и степени рискованности их поведения регистрационный № 01201262071. СПб.: СПИИРАН. 2012. 78 с. (Депонировано в ЦИТиС.)
 28. Тулупьева Т.В., Пашенко А.Е., Тулупьев А.Л., Красносельских Т.В., Казакова О.С. Модели ВИЧ-рискованного поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.
 29. Феллер В. Введение в теорию вероятностей и её приложения. В 2-х томах. Т. 2. М.: Мир, 1984. 738 с.
 30. Bell D.C, Trevino R.A. Modeling HIV Risk [Epidemiology] // JAIDS. 1999. № 22(3). P. 280–287.
 31. GeNIe& SMILE // Decisions systems laboratory. School of Information Sciences. University of Pittsburg. URL: <http://genie.sis.pitt.edu/> (дата обращения 05.05.2013).
 32. Koller D., Friedman N. Probabilistic Graphical Models. Principles and Techniques. Cambridge, Massachusetts, MIT Press, 2009. 1231 pp.
 33. Pearl J. Probabilistic Reasoning in Intelligent Systems. NYC: Morgan Kaufmann, 1988. 552 p.
 34. Sun L., Shenoy P. P. Using Bayesian networks for bankruptcy prediction: Some methodological issues //European Journal of Operational Research. 2007. Vol 180 No 2. P. 738–753.

Поддержка исследований. Результаты, представленные в настоящей работе, были получены в рамках исследования, поддержанного субгрантом № M13A11589 (A06995) гранта NIH № 5 R01AA017389-04 «Alcohol and HIV Risk Reduction in St. Petersburg, R.F.», грантом для молодых ученых и кандидатов наук Комитета по науке и высшей школе Правительства Санкт-Петербурга «Разработка программного комплекса для идентификации интенсивности и производных параметров стохастических моделей рискованного поведения на основе неполных и неточных данных», руководитель — А.Е. Пашенко (2012), грант РФФИ на 2009–2011 гг., проект № 09-01-00861-а «Методология построения интеллектуальных систем поддержки принятия решений на основе баз фрагментов знаний с вероятностной неопределенностью», грант РФФИ на 2012–2014 гг., проект № 12-01-00945-а «Развитие теории алгебраических байесовских сетей и родственных им логико-вероятностных графических моделей систем знаний с неопределенностью».

Пашенко Антон Евгеньевич — научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН. Область научных интересов: математическая статистика, статистическое моделирование, применение методов биостатистики и математического моделирования в эпидемиологии. Число научных публикаций — 52. AEP@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

Paschenko Anton Evgen'evich — researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS. Research interests: mathematical statistics, statistical modeling, application of biostatistics and mathematical modeling in epidemiology. The number of publications — 52. AEP@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupyev.

Рекомендовано лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, зав. лаб. Тулупьев А.Л., д.ф.-м.н., доц.

Статья поступила в редакцию 09.10.2013.

РЕФЕРАТ

Пащенко А.Е. Применение байесовских сетей доверия для расчета относительных оценок показателей процессов, ассоциированных с риском, в условиях информационного дефицита.

В статье рассмотрено еще одно применение разработанного ранее метода оценки интенсивности рискованного поведения в условиях дефицита информации. Предложено решение задачи для одной из наиболее существенных проблем в эпидемиологии ВИЧ, которой является приверженность пациентов лечению, насколько правильно пациент следует рекомендациям врача по применению схем лечения.

Предложено для оценки динамики изменения рискованного поведения, связанного с отклонением от ВААРТ, так и для других областей применения метода оценки интенсивности рискованного поведения использовать матрицы, в которых в простой и наглядной форме отображена позитивная либо негативная динамика изменения регистрируемого параметра, а также отклонение конкретного индивида от среднего значения параметра.

Кроме того развита модель, базирующаяся на использовании байесовских сетей доверия. Использование отношения рисков, построенных на ранее подсчитанных абсолютных оценках интенсивности поведения и ассоциированных с ними матрицами перехода, является удобным и понятным инструментом для динамического отслеживания изменений рискованного поведения.

Описаны необходимость и подходы к дискретизации случайных величин. Рассмотрена параметризация сети для кумулятивного риска, для расчета которой не требуются сведения о числе эпизодов. Таким образом, мы можем исключить из рассмотрения байесовскую сеть доверия, содержащую узел с числом последних эпизодов. Сформирована байесовская сеть для расчета относительных частот. Наконец, рассчитана условная вероятность для относительных рисков относительной защищенности. Рассмотрены численные примеры, реализованные в среде GeNIe&SMILE, которая предоставляет возможности по конструированию графических моделей и осуществляет вероятностный вывод в данной модели.

SUMMARY

Pashchenko A.E. **Application of Bayesian networks to calculate the relative confidence indicator estimates processes associated with the risk in the information deficit.**

The article discusses another application of a previously developed method for estimating the intensity of risk-taking behavior in terms of lack of information. Proposed solution of the problem for one of the most significant problems in the epidemiology of HIV, which is treatment compliance, how well the patient follows the doctor's recommendations on the application of treatments.

Proposed changes for the assessment of risk behaviors associated with the deviation from HAART, and for other applications that are estimates of the intensity of risk behavior in which use matrices in a simple and understandable form shown positive or negative trend recorded parameter changes, as well as the deviation from the mean of a particular individual parameter values.

Also developed a model based on the use of Bayesian networks of trust. Using hazard ratio built on previously calculated estimates of the absolute intensity of the behavior and the associated transition matrices is a convenient and intuitive tool for tracking the dynamic changes of risky behavior.

Describes the need for and approaches to sampling random variables. Considered parameterization network for cumulative risk calculation that does not require information on the number of episodes. Thus, we can eliminate from consideration Bayesian network trust containing the node with the number of the latest episodes. Bayesian network is formed to calculate the relative frequencies. Finally, the conditional probability is calculated relative risks for the relative security. Numerical examples are considered realized in the medium GeNIe & SMILE which provides the ability to design and implement graphical models probabilistic inference in this model.