DOI: http://dx.doi.org/10.21686/2413-2829-2018-2-154-160

ПРОГНОЗИРОВАНИЕ И АНАЛИЗ ОПЕРАЦИОННЫХ ИТ-РИСКОВ С ИСПОЛЬЗОВАНИЕМ БАЙЕСОВСКИХ СЕТЕЙ

Г. С. Петросян
Российский экономический университет имени Г. В. Плеханова,
Москва, Россия

В статье автором предложена модель анализа операционных ИТ-рисков, основанная на математическом аппарате байесовских сетей. Данная модель позволяет прогнозировать величину ущерба от ИТ-рисков в зависимости от качества программного обеспечения, квалификации ИТ-специалистов и использования различных методик тестирования. Модель сопровождается практическим примером, в рамках которого решается задача прямого байесовского вывода и проводится анализ чувствительности, что позволяет получить визуальное представление о влиянии отдельных переменных на величину ущерба от ИТ-инцидентов. Решена задача обратного байесовского вывода для анализа и определения причин рисковых событий. Модель реализована с использованием инструментальных средств RStudio и AgenaRisk. Результаты работы могут быть использованы в практической деятельности банков и их технологических подразделений при прогнозировании потерь от ИТ-инцидентов.
Ключевые слова: управление операционными рисками, ИТ-инцидент, тестирование программного обеспечения, стоимостная мера операционного риска, байесовская сеть, анализ чувствительности, байесовский вывод.

OPERATIONAL IT RISK FORECASTING AND ANALYSIS BASED ON BAYESIAN BELIEF NETWORKS

Grant S. Petrosyan
Plekhanov Russian University of Economics, Moscow, Russia

This article provides the model for IT operational risk analysis, which is based on Bayesian networks. The model allows to predict IT risk losses depending on software quality, IT staff experience and utilized testing practices. The model is provided with hands-on example. In this example, predictive Bayesian inference and sensitivity analysis are performed to get a visual representation of the impact of different input variables on the IT operational losses. The abductive Bayesian inference is performed to analyze risk events and to localize root sources of these events. The model is implemented by means of RStudio and AgenaRisk tools. Results of the work can be used in practical work of banks and its technical departments to predict IT operational losses.
Keywords: operational risk management, IT incident, software testing, operational value at risk, Bayesian network, sensitivity analysis, Bayesian inference.

Байесовская сеть представляет собой направленный граф, вершинами которого являются случайные величины, а дуги соответствуют вероятностным зависимостям между данными случайными величинами [5]. Случайные величины в байесовской сети могут быть как дискретными, так и непрерывными.

Использование байесовских сетей является надежным способом решения широкого спектра задач в области управления операционными рисками. Это объясняется в первую очередь тем, что байесовские сети доверия позволяют строить интуитивно понятные модели с визуальным представлением зависимостей между переменны-

ми, оказывающими влияние на операционный риск. В то же время моделирование байесовских сетей является научно обоснованным подходом, в основе которого лежит аппарат теории вероятностей.

Опишем байесовскую сеть, моделирующую операционный ИТ-риск. Ущерб от ИТ-риска в рамках разработанной модели рассматривается как конечная вершина (output node) байесовской сети, на которую оказывает влияние ряд других случайных величин: количество дефектов программного обеспечения (ПО), квалификация ИТ-специалистов, применяемые методики тестирования. Представленная модель базируется на аппарате теории вероятностей и математической статистики и реализована с использованием языка программирования RStudio, а также инструментального средства AgenaRisk.

Построение байесовской сети операционного ИТ-риска

Моделирование байесовской сети можно разделить на два основных этапа:

1) моделирование непосредственно графа связей между случайными величинами;

2) составление таблиц безусловных и условных вероятностей (node probability table) для каждой случайной величины. Таблицы вероятностей могут быть составлены на основе как статистических данных, так и экспертных оценок [6].

Операционный ИТ-риск – это риск ущерба текущей деятельности банка в виде убытка или недополученного дохода, вызванный используемыми информационными технологиями и ИТ-процессами [4]. Управление ИТ-рисками банка не может рассматриваться отдельно от процессов разработки и тестирования ПО. По этой причине для построения байесовской сети операционного ИТ-риска будем опираться на зависимости, известные из теории тестирования ПО и программной инженерии [3].

В конечном счете потери от ИТ-рисков (y) зависят от количества пропущенных дефектов (x_1) при тестировании очередного релиза автоматизированной системы банка либо при тестировании нового решения.

Количество пропущенных дефектов в свою очередь зависит:

1) от количества дефектов, привнесенных в программный продукт в процессе разработки (x_2);

2) квалификации специалистов по тестированию (x_3);

3) применяемых методик тестирования (x_4, …, x_n).

На рис. 1 представлен граф связей между случайными величинами y, x_1, x_2, …, x_n, построенный при помощи инструментального средства AgenaRisk.

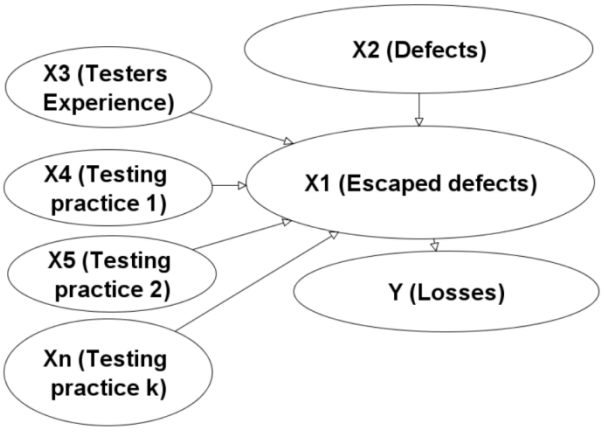

Рис. 1. Граф связей между случайными величинами, влияющими на величину операционного риска

В качестве примера рассмотрим тестирование очередного релиза автоматизированной банковской системы.

Сделаем предположение, что количество дефектов, привнесенных в процессе разработки данного релиза, имеет распределение Пуассона с математическим ожиданием λ:

$$x_2 \in \text{Пуас}(\lambda).$$

Также предположим, что имеются статистические данные по количеству дефектов за 10 релизов, предшествующих текущему релизу (табл. 1).

Таблица 1

Статистические данные по количеству дефектов

Номер релиза i	Количество дефектов Z_i
1	2
2	6
3	5
4	7
5	5
6	4
7	4
8	4
9	2
10	3
∑	42

Оценим параметр λ, используя метод максимального правдоподобия [7]:

$\hat{\lambda} = \bar{z} = \frac{42}{10} = 4{,}2$ – оценка математического ожидания случайной величины x_2.

В рамках рассматриваемого нами примера предположим, что отдел тестирования состоит на 20% из специалистов с опытом работы меньше года, 70% – с опытом работы от 1 года до 3 лет, 10% – свыше 3 лет. Ответственный за тестирование релиза выбирается случайным образом среди свободных ресурсов отдела.

Определим дискретную случайную величину x_3 в соответствии с табл. 2.

Таблица 2

Распределение случайной величины x_3

Опыт работы специалиста по тестированию	Значение x_3	Вероятность p
> 3 лет	1	0,1
1–3 года	2	0,7
< 1 года	3	0,2

Пусть также по результатам опытно-промышленной эксплуатации предыдущих релизов собраны данные по пропущенным дефектам (табл. 3).

Для того чтобы избежать трудоемких расчетов, в рамках примера будем исследовать зависимость ИТ-риска только от одной техники тестирования, а именно от использования автоматизации тестирования. Для этого введем переменную x_4, которая принимает значение 1, если в рамках релиза применяются средства автоматизированного тестирования, и 0 – в ином случае.

Таблица 3

Данные по пропущенным дефектам

Значение x_3	Найдено дефектов в процессе тестирования	Пропущено дефектов по результатам опытно-промышленной эксплуатации	Дефектов всего
1	8	2	10
2	9	6	15
3	6	14	20

В силу ограниченного количества лицензий на средства автоматизированного тестирования только 30% доработок программного обеспечения могут быть протестированы с использованием данных средств. Отсюда

$$P(x_4 = 0) = 0{,}7\ ;\ P(x_4 = 1) = 0{,}3.$$

Имеется экспертная оценка, согласно которой при прочих равных условиях использование автоматизированного тестирования снижает вероятность пропуска дефектов на 0,1.

Отсюда вероятность $p(x_3, x_4)$ пропуска дефекта может быть представлена в следующем виде:

$$p(x_3, x_4) = \begin{cases} 0{,}2 \text{ при } x_3 = 1, x_4 = 0, \\ 0{,}1 \text{ при } x_3 = 1, x_4 = 1, \\ 0{,}4 \text{ при } x_3 = 2, x_4 = 0, \\ 0{,}3 \text{ при } x_3 = 2, x_4 = 1, \\ 0{,}7 \text{ при } x_3 = 3, x_4 = 0, \\ 0{,}6 \text{ при } x_3 = 3, x_4 = 1. \end{cases}$$

Будем считать, что количество пропущенных дефектов x_1 имеет биномиальное распределение со следующими параметрами:

$$x_1 \in \text{Бин}(n = x_2, p = p(x_3, x_4)).$$

На рис. 2 изображена форма задания условных вероятностей для случайной величины x_1 в среде AgenaRisk.

Предположим, что потери от каждого отдельного инцидента подчинены гамма-распределению [7] с параметрами α и β. Тогда совокупные потери от ИТ-риска за релиз имеют следующее распределение:

$$y \in x_1 \cdot \text{Гамма}(\alpha, \beta).$$

X3 (Testers...	> 3 years		1-3 years		< 1 years	
X4 (Automa...	False	True	False	True	False	True
Expressions	Binomial(x2,0.2)	Binomial(x2,0.1)	Binomial(x2,0.4)	Binomial(x2,0.3)	Binomial(x2,0.7)	Binomial(x2,0.6)

Рис. 2. Настройка таблицы условных вероятностей в AgenaRisk

При наличии данных (табл. 4) о потерях можно оценить параметры гамма-распределения.

Таблица 4

Статистические данные по ущербу от ИТ-инцидентов

Номер инцидента	Ущерб, тыс. руб.
1	11,645
2	4,341
3	6,302
4	12,248
5	15,689
6	10,432
7	21,886
8	6,290
9	46,409
10	11,967
11	12,668
12	12,097

Для того чтобы оценить параметры $\hat{\alpha}$ и $\hat{\beta}$ гамма-распределения методом моментов [7], выполним скрипт (рис. 3) на языке программирования RStudio [8].

```
Console ~/
> library("fitdistrplus")
> data<-c(11.645, 4.341, 6.302, 12.248, 15.689, 10.432, 21.886, 6.290, 4
6.409, 11.967, 12.668,  12.097)
> fitdist(data, distr="gamma", method="mme")
Fitting of the distribution ' gamma ' by matching moments
Parameters:
       estimate
shape 1.8162792
rate  0.1267363
>
```

Рис. 3. Оценка параметров гамма-распределения в среде RStudio методом моментов

Получим:

$$\hat{\alpha} = 1{,}816\,; \quad \hat{\beta} = 0{,}127\,.$$

Таким образом, для каждой из случайных величин y, x_1, x_2, x_3 и x_4 описан соответствующий закон распределения вероятностей.

Данной информации достаточно для окончательного построения байесовской сети в AgenaRisk (рис. 4).

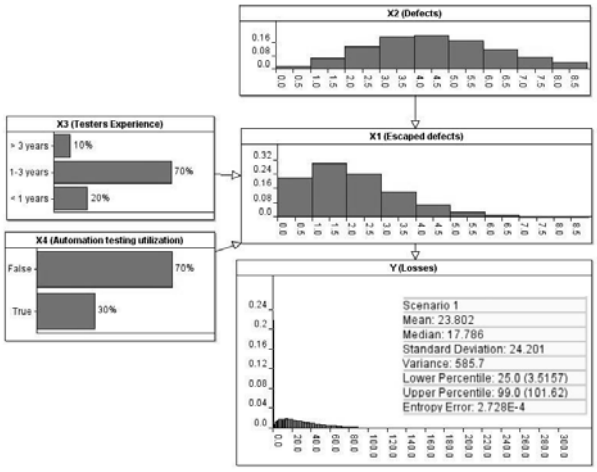

Рис. 4. Байесовская сеть операционного ИТ-риска с заданными таблицами условных вероятностей

На рис. 4 также можно видеть рассчитанные программой математическое ожидание и 99-ю перцентиль для случайной величины y:

$$M(y) = 23{,}802\,, \; y_{0,99} = 101{,}62\,.$$

Приведем экономическую интерпретацию величины $y_{0,99}$. Для этого введем понятие стоимостной меры операционного риска (operational value at risk – OpVar). Определим стоимостную меру операционного риска ИТ-релиза t как значение потерь L от инцидентов операционного рис-

ка в данном релизе, которое не будет превышено с вероятностью α [2]:

$$OpVar_{\alpha}(release_t) = sup\{u \mid P(L_t \le u) \le \alpha\}.$$

В вышеприведенной формуле множество $\{u, u \ge 0\}$ – это множество всех возможных значений ущерба u в рассматриваемом релизе. Исходя из данного определения величина *OpVar* для рассматриваемого релиза t_0 совпадает с 99-й перцентилью:

$$OpVar_{0,99}(release_{t_0}) = y_{0,99} = 101{,}62.$$

Таким образом, ожидаемая величина ущерба от ИТ-риска в текущем релизе составляет 23,802 тыс. рублей, также с вероятностью 0,99 ущерб не превысит 101,62 тыс. рублей.

Задача прогнозирования в байесовской сети операционного ИТ-риска

Задачу определения вероятности события при наблюдаемых причинах принято называть задачей прогнозирования или прямым байесовским выводом [1. – С. 69].

Рассмотрим сценарий, когда опыт работы специалиста по тестированию составляет менее 1 года. Для этого в среде AgenaRisk зададим для величины x_3 фиксированное значение, равное 3 (рис. 5).

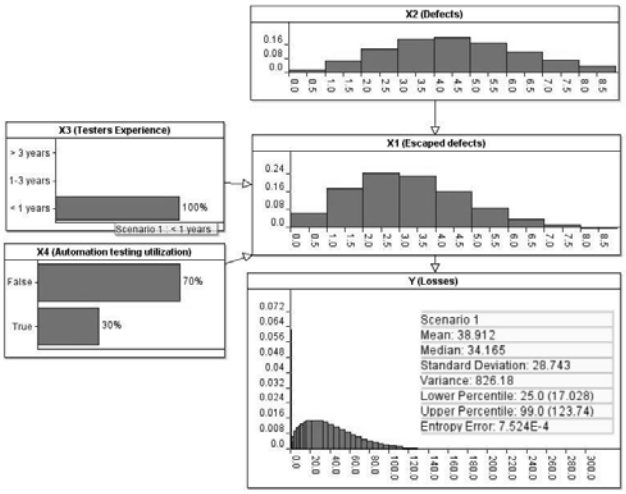

Рис. 5. Разыгрывание сценария $x_3 = 3$ (опыт работы специалиста по тестированию менее 1 года)

На рис. 5 можно видеть более «тяжелые хвосты» распределения ущерба по сравнению с рис. 4. Это свидетельствует о большей подверженности операционному риску в случае, если априори известно, что $x_3 = 3$.

Также можно вычислить абсолютные изменения ожидаемого ущерба и величины *OpVar* после того, как стало известно, что $x_3 = 3$:

$$\Delta M(y) = 38{,}912 - 23{,}802 = 15{,}110 \text{ тыс. руб.},$$

$$\Delta OpVar_{0,99}(release_{t_0}) =$$

$$= 123{,}74 - 101{,}62 = 22{,}12 \text{ тыс. руб.}$$

Анализ чувствительности операционного ИТ-риска

Байесовские сети являются удобным инструментом анализа чувствительности [5].

Используя AgenaRisk, осуществим анализ чувствительности математического ожидания потерь $M(y)$ при изменении величины x_3 (рис. 6).

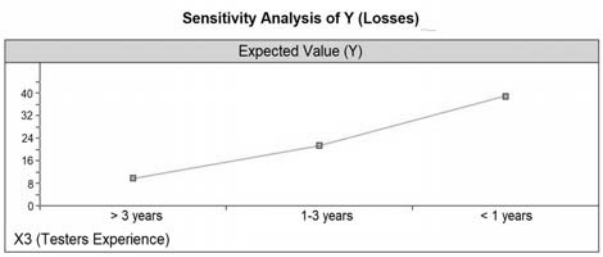

Рис. 6. График зависимости математического ожидания потерь от квалификации специалистов по тестированию ПО

Также при помощи AgenaRisk составим таблицу (табл. 5).

Таблица 5

Изменение математического ожидания потерь и величины *OpVar* в зависимости от x_3 и x_4

x_3	x_4	$M(y)$	$OpVar_{0,99}(release_{t_0})$
1	1	5,805	50,17
2	1	17,415	81,93
3	1	34,854	115,539
1	0	11,606	67,239
2	0	23,212	93,487
3	0	40,657	125,84

Пусть у руководителя проекта стоит задача минимизировать стоимость f тестирования релиза:

$$f = c(x_3) + 25x_4 \rightarrow \min,$$

где функция $c(x_3)$ имеет следующий вид:

$$c(x_3) = \begin{cases} 10, \text{ если } x_3 = 1, \\ 20, \text{ если } x_3 = 2, \\ 30, \text{ если } x_3 = 3. \end{cases}$$

Также имеется ограничение на стоимостную меру операционного риска:

$$OpVar_{0,99}(release_t) \le 90.$$

Решением данной оптимизационной задачи является вектор x^* ($x_3 = 1$, $x_4 = 0$).

Это означает, что оптимальной стратегией для руководителя является привлечение к тестированию релиза сотрудника с опытом более 3 лет, не используя при этом автоматизации тестирования. В рамках данной оптимальной стратегии

$$f(x^*) = 30 \text{ тыс. руб.},$$
$$OpVar_{0,99}(release_t)|_{x^*} = 67{,}239 \text{ тыс. руб.},$$
$$M(y)|_{x^*} = 11{,}606 \text{ тыс. руб.}$$

Определение наиболее вероятных причин рискового события

Предположим, что потери за релиз известны и составили 5 тыс. рублей ($y^0 = 5$). Введя данную информацию в параметры байесовской сети в AgenaRisk (рис. 7), при помощи последовательного применения теоремы Байеса [5] можно получить следующие апостериорные вероятности:

$$P(x_1 = 0) = 0,\ P(x_1 = 1) = 0{,}977,\ P(x_1 = 2) = 0{,}023;$$
$$P(x_3 = 1) = 0{,}107,\ P(x_3 = 2) = 0{,}771,$$
$$P(x_3 = 3) = 0{,}122;\ P(x_4 = 0) = 0{,}682,$$
$$P(x_4 = 1) = 0{,}318.$$

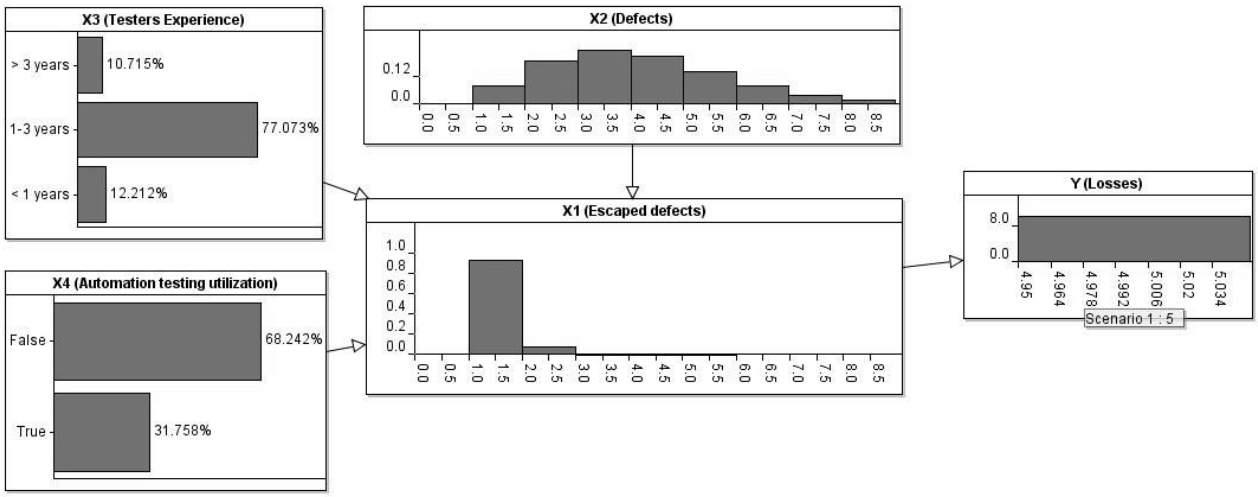

Рис. 7. Разыгрывание сценария $y^0 = 5$

Данный процесс определения вероятности причины при наблюдаемых следствиях называют диагностированием или обратным байесовским выводом [1. – С. 69].

Полученную информацию удобно использовать при определении первопричин ИТ-инцидентов. Например, в рамках рассматриваемого примера можно сделать следующий вывод: при известной величине ущерба $y^0 = 5$ тыс. рублей с наибольшей вероятностью был пропущен один дефект командой тестирования с опытом работы 1–3 года, а в процессе тестирования не использовались средства автоматизации.

Таким образом, можно сделать вывод, что байесовские сети являются эффективным инструментом для анализа операционных ИТ-рисков, а также средством поддержки принятия решений для руководителей проектов.

Список литературы

1. *Ешин С. В.* Разработка и внедрение систем менеджмента качества на основе использования байесовских сетей : монография. – Брянск : БГТУ, 2013.
2. *Петросян Г. С.* Методы анализа операционных рисков при управлении релизами банковских информационных систем // Фундаментальные исследования. – 2017. – № 11-1. – С. 108–113.
3. *Bourque P.* Guide to the Software Engineering Body of Knowledge. – USA : IEEE Computer Society Press, 2014.
4. *Cruz M. G.* Modeling, Measuring and Hedging Operational Risk. – UK : John Wiley & Sons, Inc., 2002.
5. *Fenton N.* Risk Assessment and Decision Analysis with Bayesian Networks. – USA : CRC Press, 2012.
6. *Lewis Nigel Da Costa.* Operational Risk with Excel and VBA: Applied Statistical Methods for Risk Management. – USA : John Wiley & Sons, Inc., 2004.
7. *Ross Sh. M.* Introduction to Probability and Statistics for Engineers and Scientists. – 5th ed. – USA : Associated Press, 2009.
8. *Wickham H.* R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. – Canada : O'Reilly Media, 2016.

References

1. Eshin S. V. Razrabotka i vnedrenie sistem menedzhmenta kachestva na osnove ispol'zovaniya bayesovskikh setey, monografiya [Development and Implementation of Quality Control System Based on Bayesian Networks, monograph]. Bryansk, Bryansk State Technical University, 2013. (In Russ.).
2. Petrosyan G. S. Metody analiza operatcionnykh riskov pri upravlenii relizami bankovskikh informatcionnykh sistem [Methods for Software Risk Analysis in Release Management of Banking Software]. *Fundamental'nye issledovaniya* [Fundamental Research], 2017, No. 11-1, pp. 108–113. (In Russ.).
3. Bourque P. Guide to the Software Engineering Body of Knowledge. USA, IEEE Computer Society Press, 2014.
4. Cruz M. G. Modeling, Measuring and Hedging Operational Risk. UK, John Wiley & Sons, Inc., 2002.
5. Fenton N. Risk Assessment and Decision Analysis with Bayesian Networks. USA, CRC Press, 2012.
6. Lewis Nigel Da Costa. Operational Risk with Excel and VBA: Applied Statistical Methods for Risk Management. USA, John Wiley & Sons, Inc., 2004.
7. Ross Sh. M. Introduction to Probability and Statistics for Engineers and Scientists. 5th ed. USA, Associated Press, 2009.
8. Wickham H. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Canada, O'Reilly Media, 2016.

Сведения об авторе

Грант Саркисович Петросян
аспирант кафедры информатики
РЭУ им. Г. В. Плеханова.
Адрес: ФГБОУ ВО «Российский экономический университет имени Г. В. Плеханова», 117997, Москва, Стремянный пер., д. 36.
E-mail: grant-petrosyan-s@yandex.ru

Information about the author

Grant S. Petrosyan
Post-Graduate Student of the Department for Informatics of the PRUE.
Address: Plekhanov Russian University of Economics, 36 Stremyanny Lane, Moscow, 117997, Russian Federation.
E-mail: grant-petrosyan-s@yandex.ru