

РЫНОК ТОВАРОВ И УСЛУГ

Е. Д. Соложенцев

докт. техн. наук, зав. лабораторией интегрированных систем автоматизированного проектирования
Института проблем машиноведения РАН (С.-Петербург)

СЦЕНАРНЫЕ ЛОГИКО-ВЕРОЯТНОСТНЫЕ МОДЕЛИ ВЗЯТОК

Истинная логика нашего мира —
это подсчет вероятностей

Д. К. Максвелл

Проблема взяток и коррупции актуальна для многих стран. О неблагоприятном состоянии дел с взятками и коррупцией в России нас информируют на сайте www.vzyatka.ru большое число статей и заметок. Публикации по коррупции и взяткам (Альберт, Венц, Ульямс, 1995; Сатаров, 2004), социальной статистике (Елисеева, Юзбашев, 2004; Neckman, Leamer, 2002) имеют содержательные описания, подробные анализы, большое число примеров, комментариев законов и уголовного кодекса, но не содержат математических моделей взяток.

Для решения социальных и организационных задач (включая проблемы выявления и анализа мошенничеств, взяток и коррупции) нужен, по мнению Джона фон Неймана и Норберта Винера (Соложенцев, 2006), математический аппарат на основе логики, дискретной математики и комбинаторики, более пригодный, чем дифференциальные уравнения.

Такой адекватный математический аппарат разрабатывается и называется «Логико-вероятностная (ЛВ) теория риска с группами несовместных событий» (Соложенцев, 2006; Solojntsev, 2004; Рябинин, 2000). Он апробирован для оценки и анализа: кредитных рисков, риска портфеля ценных бумаг, риска потери эффективности, риска неуспеха менеджмента компаний. ЛВ-модели риска обладают высоким качеством. Например, ЛВ-модели кредитного риска показали в два раза большую точность и в семь раз большую робастность, а также абсолютную прозрачность в распознавании плохих и хороших кредитов, чем известные методики.

В настоящей работе делается попытка использовать ЛВ-подход и ЛВ-исчисление (Соложенцев, 2006; Solojntsev, 2004) для решения актуальной проблемы — оценки и анализа вероятности взяток и коррупции.

Аксиомы теории взяток

Коррупция считается одним из основных видов теневой экономики. Чаще всего под коррупцией подразумевают получение взяток и незаконных денежных доходов государственными бюрократами, которые вымогают их у граждан ради личного обогащения.

Для построения системы и технологии борьбы с взятками и коррупцией приняты следующие положения и аксиомы (Соложенцев, 2006; Solojntsev, 2004):

- каждый способен на мошенничество, если «давят» жизненные обстоятельства, можно на время скрыть факт мошенничества тогда, когда недостаточен контроль за обоснованностью принимаемых решений;

- без количественной оценки и анализа вероятности взяток нельзя эффективно бороться с мошенничеством, взятками и коррупцией;
- каждый коммерческий банк и компания способны на взятки и коррупцию, если нет прозрачности в их бизнесе и отсутствует контроль за их деятельностью;
- за непрозрачностью методик оценки кредитных рисков и рейтингов банков и заемщиков может скрываться мошенничество и взятки;
- сложность организационной структуры учреждения или компании может являться признаком мошенничества и коррупции.

Явления взятки и коррупции близки с явлениями: отказ, авария, катастрофа в технике и явлениями: дефолт, разорение, ущерб в экономике. Их следствием являются финансовые, материальные и моральные потери для общества и государства.

Понятия вероятность взятки и коррупции близки к понятиям надежность и безопасность в технике и риск в экономике, бизнесе и банках. Чаще всего взятки имеют место при получении лицензий (образование, туризм, медицина, строительство), разрешений (ГАИ, таможня), образования (аттестаты, дипломы, экзамены), регистрации (органы МВД, местной власти) и др.

Сценарий и технология взятки различны для министерства, мэрии, учреждения, компании, банка, чиновника, врача, преподавателя и др. Взятка касается двух объектов — взяткодателя и взяткополучателя, каждый из которых имеет свою выгоду. Взяткодатель решает свою проблему быстрее, качественнее, получает привилегии, обходит закон и т. д. Взяткополучатель имеет денежную или материальную выгоду, «откат» и др. При изложении используются термины «вероятность коррупции и взятки», «вероятность успеха и неуспеха», «вероятность наличия и отсутствия взятки», «вероятность хорошего или плохого проекта» (кредита, объекта, чиновника, учреждения) с точки зрения величины вероятности взятки. (Хороший кредит, проект или объект имеют риск взятки меньше допустимого.)

Для количественной оценки и анализа взяток используется логико-вероятностная теория (ЛВ-теория) риска неуспеха с группами несовместных событий (Соложенцев, 2006; Solojentsev, 2004; Соложенцев, Степанова, Карасев, 2005; Соложенцев, Карасев, Соложенцев, 1999) и строятся ЛВ-модели взяток на основе статистических данных. Работа является одной из первых математических работ по вероятности взяток и не претендует на рассмотрение всех аспектов этой сложной проблемы и разработку всех сценариев взяток. В большей мере изложены вопросы описания и построения моделей взяток, оценки и анализа вероятности взятки, и почти не рассматриваются социальные, юридические и организационные проблемы взяток.

ЛВ-теория взяток с группами несовместных событий

События и вероятности. Взятка описывается признаками и их градациями, которые являются случайными величинами и рассматриваются как логические переменные и случайные события-признаки и события-градации, имеющие вероятности. События-признаки связаны логическими связями *OR*, *AND*, *NOT* и могут иметь циклы. События-градации для признака составляют группу несовместных событий (ГНС) (Соложенцев, 2006; Solojentsev, 2004).

Признаки есть характеристики объекта (процесса, проекта), для измерения которых используются шкалы—градации: логическая (истина\ложь, $1\backslash 0$), качественная (высокая и низкая зарплата), числовая (интервалы $[a, b]$, $[b, c]$) и др. В общем случае, градации неупорядочены и нельзя сказать, что градация 3 хуже или лучше градации 4 для итогового события.

Логическая переменная Z_j , соответствующая событию-признаку, равна 0 с вероятностью P_j , если признак j свидетельствует о наличии взятки, и равна 1 с вероятностью $Q_j = 1 - P_j$ при отсутствии взятки. Логическая переменная Z_{j^c}

соответствующая градации r для признака j , равна 0 с вероятностью P_{jr} и равна 1 с вероятностью $Q_{jr} = 1 - P_{jr}$. Вектор $Z(i) = (Z_1, \dots, Z_j, \dots, Z_n)$ описывает объект i из статистики. Для объекта i вместо логических переменных $Z_1, \dots, Z_j, \dots, Z_n$ нужно подставить логические переменные Z_{jr} для градаций признаков именно этого объекта.

Л-функция взятки в общем виде

$$Y = Y(Z_1, Z_2, \dots, Z_n). \tag{1}$$

В-функция взятки в общем виде

$$P_i Y = 1 \mid Z(i) = P(P_1, \dots, P_j, \dots, P_n), \quad i = 1, 2, \dots, N. \tag{2}$$

Для каждого события-градации в ГНС рассматриваются три вероятности (рис. 1): P_{2jr} — относительная частота в статистике; P_{1jr} — вероятность в ГНС; P_{jr} — вероятность, подставляемая в (2) вместо вероятности P_j . Определим эти вероятности для j -й ГНС:

$$P_{2jr} = P_{Z_{jr}=1} = \sum_{r=1}^{N_j} P_{2jr} = 1; \quad r = 1, 2, \dots, N_j; \tag{3}$$

$$P_{1jr} = P_{Z_{jr}=1 \mid Z_j=1} = \sum_{r=1}^{N_j} P_{1jr} = 1; \quad r = 1, 2, \dots, N_j; \tag{4}$$

$$P_{jr} = P_{Z_j=1 \mid Z_{jr}=1} = 1; \quad r = 1, 2, \dots, N_j, \quad j = 1, 2, \dots, n, \tag{5}$$

где n — число признаков; N_j — число градаций в j -м признаке, вертикальная черточка (|) читается «при условии».

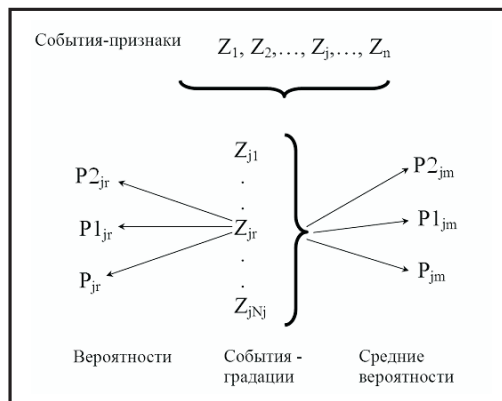


Рис. 1. Вероятности в группе несовместных событий

Средние значения вероятностей P_{2jr} , P_{1jr} и P_{jr} для градаций в ГНС равны:

$$P_{2jm} = 1 / N_j; \quad P_{jm} = \sum_{r=1}^{N_j} P_{jr} P_{2jr}; \quad P_{1jm} = \sum_{r=1}^{N_j} P_{1jr} P_{2jr}. \tag{6}$$

Вероятности P_{jr} будем оценивать при алгоритмическом итеративном обучении (идентификации) В-модели взятки по статистическим данным. Вначале

нужно определить вероятности P_{jr} , удовлетворяющие (4), и перейти от вероятностей P_{jr} к вероятностям P_{jm} . Число независимых вероятностей

$$N_{ind} = \sum_{j=1}^n N_j - n. \quad (7)$$

Вероятности P_{jr} и P_{jm} связаны по формуле Байеса для случая ограниченного количества информации (5, 7) через средние вероятности P_{jm} и P_{jm} :

$$P_{jr} = \frac{P_{jr} P_{jm}}{P_{jm}}; r=1, 2, \dots, N_j; j=1, 2, \dots, n. \quad (8)$$

Идентификация ЛВ-модели взяток по статистическим данным. Задача идентификации (обучения) ЛВ-модели взяток по статистическим данным является одной из основных и сложных в проблеме взяток. Она решается алгоритмическими методами (Соложенцев, 2006; Solojentsev, 2004). Предложена следующая схема решения задачи. Пусть известны в первом приближении вероятности для градаций P_{jr} , $r=1, 2, \dots, N_j$; $j=1, 2, \dots, n$ и вычислены риски P_i , $i=1, \dots, N$ для проектов (объектов) в статистике, каждый из которых мог сопровождаться взятками. В статистике хороших проектов N_g и плохих проектов N_b . Определим допустимый риск P_{ad} так (рис. 2), чтобы принятое нами расчетное число проектов без взяток (хорошие проекты) N_{gc} имело риск меньше допустимого и соответственно число проектов с взятками (плохие проекты) $N_{bc} = N - N_{gc}$ имело риск больше допустимого. На шаге оптимизации изменим так вероятности P_{jr} , $r=1, 2, \dots, N_j$; $j=1, 2, \dots, n$, чтобы число корректно распознаваемых проектов с взятками и без взяток увеличилось. Переменные P_{ad} и N_{gc} связаны однозначно. В алгоритме задачи удобнее задавать N_{gc} и определять допустимый риск P_{ad} .

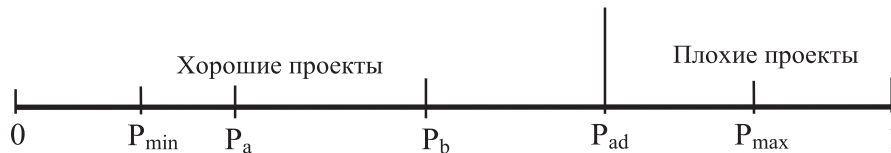


Рис. 2. Схема классификации проектов

Условие $P_i > P_{ad}$ выделяет следующие типы проектов: N_{gg} — хорошие по методике и статистике; N_{gb} — хорошие по методике и плохие по статистике; N_{bg} — плохие по методике и хорошие по статистике; N_{bb} — плохие по методике и статистике. Риски проектов N_{gg} , N_{bg} , N_{gb} , N_{bb} перемещаются относительно P_{ad} при изменении P_{jr} . При переходе одних проектов вправо от P_{ad} по величине риска, такое же число проектов переходит влево. Оптимальными будут такие изменения P_{jr} , которые переводят проекты N_{gb} и N_{bg} через P_{ad} навстречу друг другу.

Задача идентификации В-модели взятки сформулирована следующим образом (Соложенцев, 2006; Solojentsev, 2004). *Заданы:* статистика по взяткам, имеющая N_g хороших и N_b плохих проектов, и В-модель взятки (2). *Требуется определить:* вероятности P_{jr} , $r=1, 2, \dots, N_j$; $j=1, 2, \dots, n$ событий-градаций и допустимый риск P_{ad} , разделяющий проекты на хорошие и плохие. *Целевая функция:* число корректно классифицируемых проектов должно быть максимальным

$$F = N_{bb} + N_{gg} \rightarrow \underset{P_{jr}}{MAX}. \quad (9)$$

Из выражения (9) следует, что точность В-модели взяток в классификации хороших E_g , плохих объектов E_b и в целом E_m равна

$$E_g = N_{gb} / N_g; E_b = N_{bg} / N_b; E_m = (N - F) / N. \quad (10)$$

Ограничения:

1) вероятности P_{jr} должны удовлетворять условию:

$$0 < P_{jr} < 1, j = 1, 2, \dots, n; r = 1, 2, \dots, N_j. \quad (11)$$

2) средние риски проектов по В-модели и статистике должны быть равны; при обучении В-модели риска будем корректировать вероятности P_{jr} на шаге по формуле

$$P_{jr} = P_{jr}(P_{av} / P_m); j = 1, 2, \dots, n; r = 1, 2, \dots, N_j, \quad (12)$$

где $P_{av} = N_b/N$ — средний риск по статистике; P_m — средний риск по модели.

3) допустимый риск P_{ad} определяется при заданном коэффициенте асимметрии распознавания хороших и плохих проектов, равном

$$E_{gb} = N_{gb} / N_{bg}. \quad (13)$$

Формула для идентификации (обучения) ЛВ-модели взятки методом последовательного приближения с использованием случайного поиска следующая

$$\Delta P_{1_{jr}} = K_1 \frac{N_{opt} - N_v}{N_{opt}} K_3 P_{1_{jr}}, j = 1, 2, \dots, n; r = 1, 2, \dots, N_j, \quad (14)$$

где K_1 — коэффициент, равный примерно 0,05; N_{opt} , N_v — число оптимизаций и номер текущей оптимизации; K_3 — случайное число в интервале $[-1, +1]$. В процессе итеративной оптимизации величина $P_{1_{jr}}$ стремится к нулю.

Анализ вероятности взятки. Пусть В-модель взяток обучена и известны вероятности P_{jr} . Определим вклады событий-признаков и событий-градаций в вероятность взятки для проекта и множества проектов, а также в точность ЛВ-модели взятки. Для этого вычислим разности между значениями характеристик для оптимальной модели и при условии (!) придания вероятностям соответствующих событий-градаций нулевых значений (Соложенцев, 2006; Solojntsev, 2004).

Вклад признака j (всех градаций признака) в вероятность взятки для проекта i :

$$\Delta P_j = P(i) - P(i)|_{P_j=0}, j = 1, 2, \dots, n. \quad (15)$$

Вклад признака j в среднюю вероятность взятки P_m множества проектов:

$$\Delta P_{jm} = P_{jm} - P_{jm}|_{P_j=0}, j = 1, 2, \dots. \quad (16)$$

Вклад признака j в целевую функцию F_{max}

$$\Delta F_j = F_{max} - F|_{P_j=0}, j = 1, 2, \dots. \quad (17)$$

Вклады градаций в целевую функцию F_{jr} будем вычислять по аналогии с (10) как ошибки классификации проектов по каждому событию-градации:

$$E_{jrg} = N_{jrgb} / N_{jrg}; E_{jrb} = N_{jrbg} / N_{jrb}; E_{jrm} = (N_{jrgb} + N_{jrbg}) / N_{jr}, \quad (18)$$

где N_{jrgb} , N_{jrbg} , N_{jr} — числа некорректных хороших, плохих и всего проектов с градацией r .

ЛВ-модель взяток в учреждении

Учреждение принимает решения по неким проектам (делам граждан). Проектов много и для каждого проекта известно, был ли он успешным или неуспешным. Неуспешность проектов объясняется необоснованностью выдачи разрешения или ресурсов из-за взяток.

Элементами сценария и ЛВ-модели вероятности (риска) взяток являются функциональные подразделения $Z_1, \dots, Z_j, \dots, Z_n$, каждое из которых имеет N_j чиновников, принимающих решения.

В общем случае, элементы $Z_1, \dots, Z_j, \dots, Z_n$ связаны логическими связями *OR*, *AND*, *NOT* и могут иметься циклы. Чиновники в j -ом подразделении $Z_{j1}, \dots, Z_{jr}, \dots, Z_{jN_j}$ составляют ГНС. Чиновник, принимающий решение, оставляет «визу» на соответствующем документе. Построение ЛВ-модели взяток заключается в вычислении вероятностей $P_{jr}, j = 1, 2, \dots, n; r = 1, 2, \dots, N_j$, с которыми чиновники берут взятки, на основе статистики из N успешных и неуспешных проектов.

Рассмотрим ЛВ-модель взятки на примере условного банка. Используется статистика об успехах кредитов. Причины неуспеха кредитов объясняются взятками.

Пусть банк имеет пять функциональных групп чиновников, которые принимают решения о выдаче кредитов. Поставим этим группам в соответствие логические переменные Z_1, Z_2, Z_3, Z_4, Z_5 . Каждая из групп имеет соответственно N_1, N_2, N_3, N_4, N_5 чиновников, принимающих решения. Число чиновников в группе соответствует числу градаций в признаке.

Выданные кредиты оказались успешными (градация 1) или неуспешными (градация 0). Выдача кредитов оформлена документами, где зафиксированы визы чиновников и результат об успехе или неуспехе кредита.

Возможное число разных комбинаций прохождения клиента через банк и взятку равно

$$N_{\max} = N_1 N_2 N_3 N_4 N_5. \quad (19)$$

Логическая функция взяток в совершенной дизъюнктивной нормальной форме (СДНФ) содержит N_{\max} логических слагаемых и запишется так:

$$Y = Z_1 Z_2 Z_3 Z_4 Z_5 \vee \overline{Z_1} Z_2 Z_3 Z_4 Z_5 \vee Z_1 \overline{Z_2} Z_3 Z_4 Z_5 \vee \dots \vee \overline{Z_1} \overline{Z_2} \overline{Z_3} \overline{Z_4} \overline{Z_5}. \quad (20)$$

В каждый конъюнкт (20) входит каждая логическая переменная из Z_1, Z_2, Z_3, Z_4, Z_5 или ее отрицание (черта над логической переменной). Все конъюнкты попарно ортогональны, т. е. совершенная дизъюнктивная нормальная форма является ортогональной. При вычислении вероятности события Y вместо событий Z_1, Z_2, Z_3, Z_4, Z_5 следует подставить вероятности этих событий, а знак логической операции «OR» заменить на знак «+»

СДНФ является самой громоздкой записью логической функции. В действительности логическая модель взяток может быть записана проще, если можно учесть структуру подразделений банка и их связи. Для определенности примем, что сценарий и структуру модели взяток определяет «мостик», приведенный на рис. 3.

Здесь чиновники из Z_1 и Z_2 проверяют обеспечение кредитов, а чиновники из Z_3 и Z_4 принимают решения о размере и сроках кредита. Чиновники (начальники) из Z_5 руководят процессом. Клиент идет к одному из начальников, который за взятку направляет к чиновникам из групп $Z_1 - Z_4$, которые берут взятки.

Логическая модель (Л-модель) взяток в дизъюнктивной нормальной форме (ДНФ) (записи логической функции без скобок) на основе кратчайших путей успешного функционирования (Рябинин, 2000) системы с взятками запишется так:

$$Y = Z_1 Z_3 \vee Z_2 Z_4 \vee Z_1 Z_5 Z_4 \vee Z_2 Z_5 Z_3. \quad (21)$$

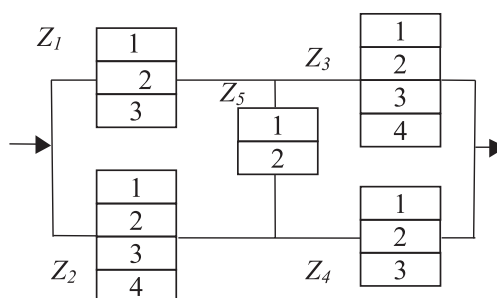


Рис. 3. Структурная модель взяток

Вероятностная модель (В-модель, В-полином) взяток, полученная после ортогонализации логической функции (21), запишется так:

$$P = p_2 p_4 + p_1 p_3 + q_1 p_2 p_3 q_4 p_5 + p_1 q_2 q_3 p_4 p_5 - p_1 p_2 p_3 p_4. \quad (22)$$

Идентификация модели взяток заключается в определении вероятностей P_{jr} , $j = 1, \dots, n$; $r = 1, \dots, N_j$ (с которыми чиновники берут взятки) по статистике фактов взяток, установленных судом.

Пример 1. Для идентификации В-модели взяток в банке использовалась статистика из 1000 кредитов, из которых 700 были хорошими и 300 плохими, т. е. средний риск взятки равен $P_{av} = 300/1000 = 0,3$. Пять признаков имеют от 4 до 10 градаций; всего 40 градаций.

Таблица 1

Средние вероятности взяток для групп чиновников

Группы, j	Вероятности, P_{jm}	Вероятности, $P1_{jm}$	Число чиновников, N_j
1	0,478113	0,249540	4
2	0,348310	0,075949	10
3	0,298833	0,133823	5
4	0,388857	0,116348	11
5	0,291868	0,091775	10

Таблица 2

Вероятности взяток чиновников

Номера градаций	Вероятности, $P1_r$	Вероятности, $P11_r$	Частоты, $P21_r$
Группа Z_1			
1	1,0	0,522300	0,274
2	0,596084	0,311103	0,269
3	0,248278	0,129579	0,063
4	0,070927	0,037017	0,394
Группа Z_2			
1	0,0	0,0	0,0
2	0,687703	0,149933	0,014
3	0,227359	0,0495688	0,002
4	1,0	0,218209	0,054
5	0,510577	0,111316	0,017
6	0,704722	0,153643	0,086
7	0,570149	0,124304	0,057
8	0,448856	0,097859	0,224
9	0,434821	0,094799	0,187
10	0,001675	0,000365	0,359

В результате обучения определены вероятности P_{jr} и $P1_{jr}$ для всех чиновников и подсчитаны характеристики ЛВ-модели взяток. Значение целевой функции в результате идентификации $F_{\max} = 720$ и допустимое значение риска взятки $P_{ad} = 0,3094$.

Результаты исследований приведены в табл. 1 и 2. Вероятности $P2_{jr}$ и $P1_{jr}$ градаций, хотя и составляют в сумме 1 в ГНС, существенно отличаются между собой (табл. 2). Вероятности взяток у чиновников (вероятности P_{jr}) отличаются более чем в 10 раз. Наибольшие средние вероятности P_{jm} имеют группы чиновников 1 и 4. Эти же группы вносят наибольшие вклады в средний риск P_m . Вероятности взяток групп чиновников P_{jm} отличаются почти в два раза.

ЛВ-модель взяток на основе описания поведения чиновника

Взятки — это преступление, не выставляемое напоказ. Не возникает вопроса о наличии преступления при ограблении банка, когда свидетелями являются служащие и клиенты. Взятки же отличается от других видов преступлений сложностью выявления самого факта их совершения. Однако взятки носят массовый характер, и по ним имеется достаточно много данных в судебных и контролирующих органах.

Для каждого типа взяток можно выделить признаки (Альберт, Венц, Ульямс, 1995; Соложенцев, 2006), ассоциирующие с подобным преступлением. Каждый из признаков имеет как минимум две градации. В-модель взяток может быть обучена по статистическим данным.

Расследование взятки нужно проводить в том случае, когда с большой вероятностью можно полагать, что она имела место. Серьезность подозрения во взятке можно оценить количественно по вероятности взятки. О взятках чиновника (врача, учителя) свидетельствуют, например, следующие признаки личности и необычного поведения:

- возраст;
- продолжительность работы в учреждении или компании;
- приобретение дома, квартиры, дачи, машины и т. д. со стоимостью, не соответствующий уровню заработной платы;
- наличие долгов;
- финансовые запросы;
- пристрастие к азартным играм;
- выходящий за привычные рамки образ жизни;
- необычное поведение;
- наличие жалоб;
- неясное или уголовное прошлое;
- отсутствие разделения обязанностей;
- отсутствие независимых проверок;
- отсутствие соответствующих документов и записей;
- пренебрежение существующими правилами;
- ненадлежащая система документооборота и т. д.

Элементами сценария и ЛВ-модели взятки являются названные выше признаки $Z_1, \dots, Z_j, \dots, Z_n$, каждый из которых имеет несколько градаций. Градации для j -го признака $Z_{j1}, \dots, Z_{jr}, \dots, Z_{jN_j}$ составляют ГНС. *Сценарий взятки чиновника описывается так: взятка имеет место, если имеет место любое одно событие-признак, или любые два события-признака, или...все события-признаки.* Сценарий взятки в виде графа представлен на рис. 4.

Л-функция (ЛВ-модель) взяток в ДНФ (Соложенцев, 2006; Solojentsev, 2004):

$$Y = Z_1 \vee Z_2 \vee \dots \vee Z_j \vee \dots \vee Z_n. \quad (23)$$

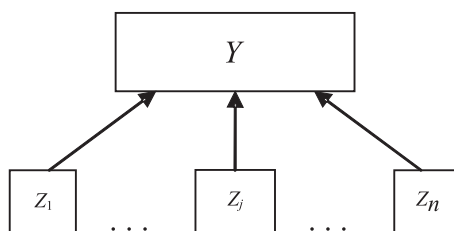


Рис. 4. Структурная модель взяток (— логическая связь ИЛИ)

Л-функция риска взяток в ортогональной форме (ОДНФ) после ортогонализации (23):

$$Y = Z_1 \vee Z_2 \overline{Z_1} \vee Z_3 \overline{Z_2} \overline{Z_1} \vee \dots \tag{24}$$

В-функция (модель, полином) взяток, записанная по ОДНФ (24):

$$P = p_1 + p_2 q_1 + p_3 q_1 q_2 + \dots \tag{25}$$

Идентификация ЛВ-модели взяток заключается в вычислении вероятностей P_{jr} , $j = 1, 2, \dots, n$; $r = 1, 2, \dots, N_j$ (с которыми чиновник берет взятки) по статистике подозрений на взятки, рассмотренных судом.

«Арифметика» в В-модели взяток такова, что вероятность взяток находится в пределах $[0,1]$ при любых значениях вероятностей инициирующих событий. Для каждого события-градации в ГНС используются три ранее введенные вероятности (рис. 1): P_{2jr} , P_{1jr} , P_{jr} .

Наибольшее число разных комбинаций взяток равно:

$$N_{\max} = N_1 N_2 \dots N_j \dots N_n, \tag{26}$$

где $N_1, \dots, N_j, \dots, N_n$ — число градаций в событиях-признаках.

Если число признаков $n = 20$ и каждый имеет $N_j = 5$ градаций, то число разных комбинаций взяток (или конъюнктов в СДНФ) равно астрономическому числу $N_{\max} = 5^{20}$, что объясняет трудности борьбы с взятками и коррупцией.

ЛВ-модель взяток (23—25) описывает все возможные взятки и является самой точной. Однако в ряде случаев не нужно учитывать все возможные взятки. Например, из статистики известно, что взятка имела место при появлении одного и не более двух событий из Z_1, Z_2, \dots, Z_n . Тогда следует использовать Л-модель взяток при ограниченном числе разных взяток (Соложенцев, 2006).

Пусть имеем логическую модель взяток из четырех элементов

$$Y = Z_1 \vee Z_2 \vee Z_3 \vee Z_4. \tag{27}$$

Для ограниченного числа разных взяток, сценарий формулируется так: взятка может иметь место, когда происходит любое одно или любые два события. Л-модель риска неуспеха запишется так:

$$Y = Z_1 \overline{Z_2} \overline{Z_3} \overline{Z_4} \vee Z_2 \overline{Z_1} \overline{Z_3} \overline{Z_4} \vee Z_3 \overline{Z_1} \overline{Z_2} \overline{Z_4} \vee Z_4 \overline{Z_1} \overline{Z_2} \overline{Z_3} \vee Z_1 Z_2 \overline{Z_3} \overline{Z_4} \vee Z_1 Z_3 \overline{Z_2} \overline{Z_4} \vee Z_1 Z_4 \overline{Z_2} \overline{Z_3} \vee Z_2 Z_3 \overline{Z_1} \overline{Z_4} \vee Z_2 Z_4 \overline{Z_1} \overline{Z_3} \vee Z_3 Z_4 \overline{Z_1} \overline{Z_2}. \tag{28}$$

В Л-модели риска (28) все логические слагаемые попарно ортогональны, что позволяет сразу же записать В-модель (В-полином) риска неуспеха:

$$p\{Y\} = p_1q_2q_3q_4 + p_2q_1q_3q_4 + p_3q_1q_2q_4 + p_4q_1q_2q_3 + p_1p_2q_3q_4 + p_1p_3q_2q_4 + p_1p_4q_2q_3 + p_2p_3q_1q_4 + p_2p_4q_1q_3 + p_3p_4q_1q_2. \quad (29)$$

Пример 2. Фактических данных о взятках, установленных судами по уголовным делам, у автора не было. В качестве статистических данных использовались модельные данные. Из 1000 подозреваемых во взятках, на которых возбуждены уголовные дела, были осуждены только 300, а остальные 700 были признаны невиновными. Таким образом, средний риск взятков равен $P_{av} = 300/1000 = 0,3$. Подозреваемые описывались $n = 20$ признаками, которые имели в сумме 94 градации.

Идентификация В-модели взятков (25) заключается в определении вероятностей P_{jr} , $r = 1, 2, \dots, N_j$; $j = 1, 2, \dots, n$ событий-градаций. На шаге оптимизации вычисляется вероятность взятков для каждого подозреваемого и сравнивается с допустимой вероятностью P_{ad} . Подозреваемого относят к плохим или хорошим. Целевая функция формулировалась так: число корректно классифицируемых подозреваемых должно быть максимальным.

Вклады событий-градаций в точность ЛВ-модели взятков рассмотрим на примере событий-градаций (табл. 3) признаков Z_2 и Z_{13} для оптимально обученной ЛВ-модели взятков ($F_{\max} = 826$). В табл. 3 приведены частоты градаций для всех P_{2jr} , плохих P_{20jr} и хороших P_{21jr} подозреваемых; вероятности событий-градаций P_{1jr} и P_{jr} ; ошибки распознавания по градациям для всех E_{jr} , плохих E_{0jr} и хороших E_{1jr} подозреваемых.

Вклады событий-признаков в вероятность взятки подозреваемого пропорциональны вероятностям P_j , $j = 1, 2, \dots, n$, равным вероятностям событий-градаций P_{jr} . Вероятности P_{jr} событий-градаций признаков отличаются более чем в 10 раз. Ошибки градаций E_{jr} в классификации взятков отличаются почти в 5 раз.

ЛВ-анализ модели взятков выполнен с использованием выражений (15–18). Для каждого j -го признака определялись (табл. 4): средние значения вероятностей P_{1jm} и P_{jm} , а также уменьшение числа распознаваемых хороших и плохих подозреваемых ΔF_j при исключении этого признака из модели риска. ЛВ-модель взятков после такого изменения переобучалась. Уменьшение числа распознаваемых подозреваемых определено по отношению к оптимально обученной модели взятков со всеми признаками.

Таблица 3

Вероятности и ошибки распознавания для градаций признаков подозреваемых

P_{2jr}	P_{20jr}	P_{21jr}	P_{1jr}	P_{jr}	E_{jr}	E_{1jr}	E_{0jr}
Признак Z_2							
0,014	0,007	0,007	0,010	0,019	0,214	0,429	0,0
0,002	0,001	0,001	0,070	0,014	0,500	1,0	0,0
0,054	0,032	0,022	0,194	0,038	0,278	0,682	0,0
0,017	0,005	0,012	0,159	0,031	0,412	0,5	0,2
0,086	0,038	0,048	0,145	0,028	0,256	0,417	0,053
0,057	0,019	0,038	0,095	0,019	0,228	0,289	0,105
0,224	0,066	0,158	0,067	0,013	0,169	0,196	0,106
0,167	0,056	0,131	0,053	0,010	0,203	0,183	0,250
0,359	0,076	0,283	0,016	0,003	0,114	0,081	0,237
Признак Z_{13}							
0,0190	0,080	0,110	0,283	0,027	0,237	0,345	0,087
0,511	0,142	0,369	0,233	0,021	0,186	0,201	0,148
0,248	0,065	0,183	0,093	0,008	0,113	0,082	0,200
0,028	0,007	0,021	0,346	0,032	0,178	0,238	0,0
0,023	0,006	0,017	0,044	0,004	0,217	0,117	0,5

Наибольший вклад в точность распознавания вносят признаки: $Z_1, Z_2, Z_4, Z_5, Z_6, Z_3, Z_{13}$. Нулевой вклад вносят признаки $Z_{11}, Z_{12}, Z_{17}, Z_{18}, Z_{19}$. Исключение признаков 11, 12, 17 и 18 уменьшает число распознаваемых подозреваемых всего на 4.

Таблица 4

Вклады признаков в точность модели взяток

Признаки, j	Число градаций, N_j	$P1_{jm}$	P_{jm}	F_j
1	4	0,272384	0,020226	-64
2	10	0,063346	0,012359	-27
3	5	0,098475	0,009327	-18
4	11	0,090820	0,020927	-26
5	10	0,080377	0,017593	-20
6	5	0,272148	0,022466	-20
7	5	0,206945	0,018549	-6
8	4	0,266619	0,017736	-6
9	4	0,183897	0,014253	-10
10	3	0,318015	0,018295	-10
11	4	0,251871	0,018974	0
12	4	0,247375	0,017166	0
13	5	0,206718	0,018900	-16
14	3	0,235637	0,014733	-2
15	3	0,261648	0,017591	-8
16	4	0,341959	0,021975	-2
17	4	0,289853	0,018739	0
18	2	0,482499	0,017417	0
19	2	0,508613	0,018138	0
20	2	0,750896	0,018326	-2

Точность ЛВ-модели взяток изменяется при изменении числа градаций в признаке. Выполнены исследования для признака Z_2 , который в исходном варианте имеет 10 градаций. После переобучения модели взяток получены следующие результаты: при отсутствии признака $F_{max} = 800$; при двух градациях в признаке $F_{max} = 808$; при четырех градациях $F_{max} = 812$; при десяти градациях $F_{max} = 824$; при ста градациях (имеется семьдесят пустых градаций, которые не используются в статистике) $F_{max} = 828$.

Строились графики вероятностей взяток для 1000 подозреваемых до и после их сортировки по величине вероятности. Примерно 15% подозреваемых имели маленькие вероятности взяток, т. е. были очень хорошими, и 15% подозреваемых имели большие вероятности взяток, т. е. были очень плохими. Это говорит о том, что подозреваемых следует классифицировать по величине вероятности взяток не на два, а на четыре класса.

ЛВ-модель взяток на основе анализа параметров обслуживания

Оценим вероятность взяток, используя статистику параметра обслуживания, например, времени решения чиновником (учреждением) проблемы клиента от поступления заявки до принятия окончательного решения, времени протезирования протезистом от начала до конца этого процесса и др.

Пусть для N клиентов имеется статистика по времени обслуживания $Y_i, i = 1, 2, \dots, N$. Если построить нормальный закон распределения параметра Y_i и задать его средним значением и дисперсией, то это приведет к существенному уменьшению точности оценки взяток, так как нормальный закон распределения встречается крайне редко.

Параметр обслуживания может иметь непрерывные или дискретные значения. В обоих этих случаях с целью повышения адекватности модели взятки и использования аппарата ЛВ-исчисления будем строить дискретное распределение на выбранных интервалах разбиения значений параметра. Каждому интервалу со средним значением параметра на нем присваивается номер градации. Градации составляют группу несовместных событий (ГНС). Вероятности событий-градаций определяются по формуле

$$P_j = N_j / N, \quad (30)$$

где N_j — число значений параметра в статистике с данной градацией; N — число значений параметра в статистике.

Параметр обслуживания имеет допустимое значение Y_{ad} (рис. 5). Вероятность $P\{Y < Y_{ad}\}$ назовем риском взятки. *Сценарий взятки формулируется так: если параметр обслуживания меньше (больше) допустимого значения, то имеется подозрение на взятку.*

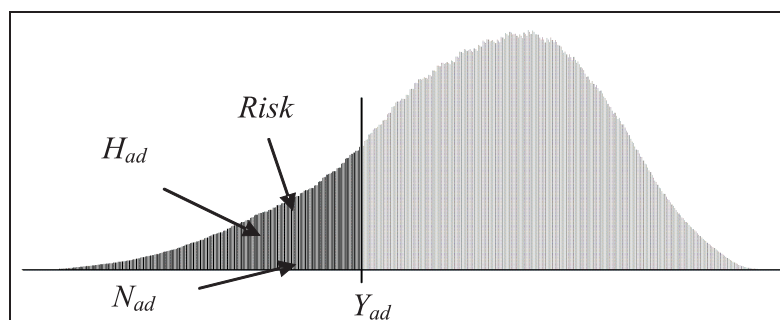


Рис. 5. Дискретное распределение параметра обслуживания

Таким образом, для параметра обслуживания при заданном риске $Risk$ могут быть вычислены: допустимое значение Y_{ad} , число значений параметра в «хвосте» распределения N_{ad} , энтропия вероятностей параметра в «хвосте» распределения H_{ad} .

Во многих работах по VaR (*Value-at-Risk*) (Соложенцев, 2006; Solojentsev, 2004) исследуются «хвосты» распределений, для которых вводят различные типы распределений и условные вероятности. В ЛВ-теории риска этого не нужно делать, ибо используется произвольный закон распределения значений параметра, который задается дискретным рядом.

Пример 3. Обслуживались $N = 700$ клиентов. Параметр Y_1 определяет продолжительность обслуживания в днях и имеет $N_1 = 30$ дней-градаций, для которых вычислены вероятности P_{1r} , $r = 1, 2, \dots, 30$ по (30). Допустимое значение параметра $Y_{1ad} = 10$ и значение риска $Risk1 = 0,2$. При условии $Y_1 < Y_{1ad}$ вызывается подозрение во взятке.

Пусть имеется еще один параметр обслуживания Y_2 , имеющий $N_2 = 20$ градаций, допустимое значение Y_{2ad} и риск $Risk2$. Параметрам обслуживания соответствуют логические переменные, которые могут быть зависимыми, но не изначально, а потому, что они содержатся в определенной логической формуле, которая и определяет зависимость между ними.

Для случая двух параметров обслуживания Y_1 и Y_2 будем иметь $N = N_1 N_2 = 30 \times 20 = 600$ комбинаций обслуживания. Л-функции для двух разных комбинаций обслуживания $Y = Y_1 \wedge Y_2$ ортогональны (произведение логических функций равно нулю), так как эти комбинации содержат разные градации для Y_1 и Y_2 , принадлежащие одним и тем же ГНС.

Свойство ортогональности разных комбинаций обслуживания позволяет перейти от Л-функций к алгебраическим выражениям для вероятностей, т. е. Л-переменные заменить на вероятности и знаки «или» на знаки «плюс».

Легко вычислить число комбинаций, удовлетворяющих условию взятки

$$P\{(Y_1 \leq Y_{1ad}) \vee (Y_2 \leq Y_{2ad})\}, \quad (31)$$

и вероятность взятки.

Заключение

Задачи обучения ЛВ-моделей взяток по статистическим данным, оценки и анализа вероятностей взяток на основе ЛВ-теории риска с группами несовместных событий, как следует из выражений (19, 25), имеют исключительно высокую вычислительную сложность и могут решаться только с использованием компьютеров и специальных *Software*. Разработан комплекс программ для идентификации модели, оценки и анализа взяток. Подробнее сведения об этих программных средствах можно получить в работах (Соложенцев, 2006; Solojentsev, 2004; Соложенцев, Карасев, Соложенцев, 1999), на вебсайтах www.inorisklab.com, www.ipme.ru/ipme/labs/iisad/sapr.htm, и по E_mail: risk@sapr.ipme.ru.

ЛВ-модели взяток и соответствующие *Software* предназначены для службы «Экономических преступлений» города с целью выявления взяток по статистическим данным.

Основные результаты настоящей работы следующие:

1. Предложено использовать ЛВ-теорию риска с ГНС для разработки ЛВ-моделей взяток с целью выявления, оценки и анализа вероятности взятки по статистическим данным.
2. Построение ЛВ-модели взяток с ГНС включает в себя:
 - представление Л-модели взяток в СДНФ для оценки числа возможных комбинаций взяток и вычислительной сложности алгоритмов;
 - запись Л-модель взяток в ДНФ, используя сценарий взяток в виде высказываний или с помощью графа, или в виде кратчайших путей, или для ограниченного множества событий;
 - преобразование Л-модели взяток из ДНФ в ОДНФ;
 - запись В-модели взяток по ОДНФ;
 - идентификация (обучение) В-модели взяток по статистике;
 - анализ В-модели взяток с вычислением вкладов признаков и их градаций в вероятность взятки, среднюю вероятность взяток и точность ЛВ-модели.
3. Разработаны сценарии и ЛВ-модели для взяток с ГНС:
 - в учреждении по результатам (параметрам) его функционирования;
 - чиновников на основе описания их поведения параметрами и признаками;
 - в учреждении и чиновников на основе анализа параметров обслуживания.
4. Приведены примеры оценки и анализа вероятностей взяток на основе идентификации ЛВ-модели взяток с ГНС по статистическим данным.
5. Предложенные ЛВ-модели взяток могут использоваться для учреждения как по отдельности, так и в совокупности.
6. Созданы *Software* для идентификации (обучения) ЛВ-моделей взяток, оценки и анализа вероятностей взяток, предназначенные для службы «Экономических преступлений» города с целью выявления взяток по статистическим данным.

Источники

- Альберт С., Венц Дж., Ульямс Т.* Мошенничество. Луч света на темные стороны бизнеса / Пер. с англ. СПб., 1995.
- Елисеева И. И., Юзбашев М. М.* Общая теория статистики: Учебник. М., 2004.
- Рябинин И. А.* Надежность и безопасность структурно-сложных систем. СПб., 2000.
- Сатаров Г. А.* Антикоррупционная политика. Учебное пособие, М., 2004.
- Соложенцев Е. Д.* Сценарное логико-вероятностное управление риском в бизнесе и технике. 2-е изд. СПб., 2006.
- Соложенцев Е. Д., Карасев В. В., Соложенцев В. Е.* Логико-вероятностные модели риска в банках, бизнесе и качестве. СПб., 1999.
- Соложенцев Е. Д., Степанова Н. В., Карасев В. В.* Прозрачность методик оценки кредитных рисков и рейтингов. СПб., 2005.
- Heckman J. J., Leamer Ed.* Handbook of Econometrics. 2002. Vol. 5.
- Solojentsev E. D.* Scenario Logic and Probabilistic Management of Risk in Business and Engineering. Springer, 2004.