

УДК 330:519.257

АКТУАРНЫЙ АНАЛИЗ СТАТИСТИЧЕСКОЙ БАЗЫ В МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРОЦЕССА ЗАБОЛЕВАНИЯ САХАРНЫМ ДИАБЕТОМ**Валитов Д.И.***ФГБОУ ВО «Башкирский государственный университет», Уфа, e-mail: damirvalitov@yandex.ru*

В основе статьи лежат актуарные расчеты на базе статистических данных. Актуарные расчёты – это расчёты страховых ставок на базе математической статистики. За статистическую базу взяты данные заболеваемости сахарным диабетом. На основе реальных данных, предоставленных в Медицинском информационно-аналитическом центре при министерстве здравоохранения Республики Башкортостан, получены необходимые статистические совокупности, построена математическая модель, поставлены две задачи (прямая и обратная). В прямой задаче важна правильная постановка самой задачи, решение же имеет тривиальный вид, так как вся структура модели и необходимые данные для расчета заведомо известны. А вот при решении обратной задачи уже необходимо определить некоторые неизвестные параметры модели по известным характеристикам процесса, порой даже каких-то эмпирических данных, поэтому для их решения использовались методы вероятностно-статистического анализа, при построении модели применялась теория марковских процессов, графов и интервального анализа, а в основе поиска обратного решения лежит принцип двойственности в линейном программировании. Аналогичные процессы рассматривались во многих научных исследованиях ученых, занимающихся актуарной математикой, моделированием процессов, связанных со страхованием, работы которых анализировались при формировании текущего исследования. Написана программа для решения обратных задач определения интенсивностей переходов в системе многих состояний процесса. Используя результаты исследования и на основе вероятностно-статистического моделирования полученных результатов можно предсказывать, управлять и оценивать риски, специалистам по медицинскому страхованию - рассчитывать тарифы. Это дает определенный импульс в анализе рисков при расчете границ величин страховых взносов в задачах профилактики и лечения сахарного диабета.

Ключевые слова: статистические данные, страхование, актуарные расчеты, марковская модель, интенсивности перехода, система уравнений Колмогорова – Чепмэна, сахарный диабет, граф состояний

ACTUARIAL ANALYSIS OF STATISTICAL BASE IN THE MATHEMATICAL MODEL OF THE PROCESS OF DIABETES**Valitov D.I.***Federal State Budgetary Educational Institution of Higher Education Bashkir State University, Ufa, e-mail: damirvalitov@yandex.ru*

The article is based on actuarial calculations based on statistical data. Actuarial calculations are calculations of insurance rates based on mathematical statistics. For the statistical base taken data on the incidence of diabetes. On the basis of real data provided by the Medical Information and Analytical Center under the Ministry of Health of the Republic of Bashkortostan, the necessary statistical aggregates were obtained, a mathematical model was constructed, 2 tasks were set (direct and inverse). In the direct problem, the correct formulation of the problem itself is important, the solution has a trivial view, since the entire structure of the model and the necessary data for the calculation are known. But when solving the inverse problem, it is already necessary to determine some unknown parameters of the model from the known characteristics of the process, sometimes even some empirical data, therefore, methods for probabilistic-statistical analysis were used to solve them, and the theory of Markov processes, graphs and interval analysis was used to build the model, and the basis of the search for the reverse solution is the duality principle in linear programming. Similar processes were considered in many scientific studies of scientists involved in actuarial mathematics, modeling insurance-related processes, whose work was analyzed during the formation of the current study. A program has been written for solving inverse problems of determining transition intensities in a system of many states of a process. Using the results of the study and on the basis of probabilistic statistical modeling of the obtained results, it is possible to predict, manage and evaluate risks, health insurance specialists can calculate the rates. This gives a certain impetus to risk analysis when calculating the boundaries of insurance premiums in the tasks of prevention and treatment of diabetes.

Keywords: statistical data, insurance, actuarial calculations, Markov model, transition intensity, system of Kolmogorov – Chapman equations, diabetes mellitus, state graph

В последнее время начали широко применяться актуарные исследования в анализе рисков. Существует большое количество моделей, и одной из них является стохастическая модель Уилки. Такие модели описывают случайные процессы, и их «случайность» описывается при помощи теории вероятности. Немаловажную роль в их построении также играет и статистическое моделирование. В основе регулирования

страховых отношений между страховщиком и страхователем лежат методы, опирающиеся на математические и статистические закономерности. Они и называются актуарными расчетами.

Большинство популярных приложений используют принцип, использующий марковские процессы при моделировании условий поведения системы. Это задачи с начальными статистическими данными,

например задачи системам массового обслуживания: почта, мастерские, заправочные станции, магазины и т.д. Так, для представления состояния страхователя используется тот же подход.

В результате математического анализа получаем две противоположные задачи: прямая, где необходимо найти вероятности нахождения индивида в том или ином состоянии при известных параметрах модели, и обратная, где необходимо найти параметры модели, используя известные характеристики из эксперимента.

Обратные задачи для марковских процессов начали развиваться сравнительно недавно в отличие от методов решения прямых задач. На примере модели страхования в медицине можно видеть, что интенсивности переходов есть не что иное, как количество людей, находящихся в разных «состояниях» системы в определенный момент времени. То есть это заболевающие, выздоравливающие и умершие [1].

Цель работы: определение метода решения обратных задач процесса заболеваемости сахарным диабетом на основе статистической базы в математических модели случайного процесса.

Задачи исследования:

- статистическое исследование базы данных;
- получение математической модели медицинского страхования;
- написание программы на языке программирования C++ для решения обратных задач;
- приближенное решение для математической модели;
- на основе вероятностно-статистического моделирования полученных результатов предсказание и оценка рисков, расчет тарифов при страховании.

Материалы и методы исследования

Вероятностный процесс в той или иной системе с определенным числом состояний и дискретным временем, а также подчиняющийся правилу, что вероятность наступления любого состояния модели в будущем влияет только ее состояние в настоящем, будет называться марковским. В случае с непрерывным временем в марковском процессе основы были заложены А.Н. Колмогоровым. Он показал, что марковское моделирование дает возможность исследователям видеть изменения системы в каждой из дискретных «точек» времени.

Используя статистические данные, можно получить интенсивности переходов между состояниями цепи, они и являются начальными данными самой цепи. Человек, вернее его нахождение в определенном состоянии в определенный момент времени, является основным звеном анализируемой цепи. Самый простой случай это всего два состояния (рис. 1).



Рис. 1. Схема из двух состояний

В данном случае индивид может перейти только из состояния «жив» в состояние «мертв». При разных схемах страхования страховые пособия выплачиваются по-разному. Например, периодические равные платежи идут пока клиент находится в состоянии 1, а после перехода в состояние 2 – прекращаются. Либо производятся при переходе в состояние 2 и т.д. Все эти схемы просты и легко вычисляются.

Рассматривая схему с тремя возможными вариантами: «здоров», «болен», «мертв» – видно, что в этом случае взносы будут выплачиваться при нахождении страхователя в первом состоянии, а выплаты будут осуществляться при нахождении во втором состоянии.

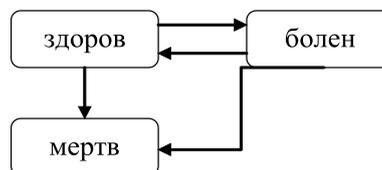


Рис. 2. Схема из трех состояний

Сложность актуарных расчетов в этом случае обуславливается тем, что человек может переходить между состояниями туда и обратно по нескольку раз. Именно поэтому такие переходы при расчетах «блокируются» [2].

Марковская цепь при непрерывном времени выглядит:

Пусть $X(t)$ состояние человека в возрасте t ($t \geq 0$). Зададим случайный процесс $\{X(t), t \geq 0\}$. Количество состояний будем предполагать счетным $\{1, 2, \dots, n\}$. Тогда $\{X(t), t \geq 0\}$ – марковский процесс при любых $s, t \geq 0$ и $i, j, x(u) \in \{1, 2, \dots, n\}$,

$$\Pr\{X(s+t) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\} = \Pr\{X(s+t) = j \mid X(s) = i\}.$$

Будущее этого процесса зависит от «прошлого» (в момент времени s) только через «настоящее» (до момента времени s).

Функция вероятности перехода:

$$p_{ij}(s, s+t) \equiv \Pr\{X(s+t) = j | X(s) = i\}, i, j \in \{1, 2, \dots, n\},$$

предполагаем $\sum_{j=1}^n p_{ij}(s, s+t) = 1$ при любых $t \geq 0$ и $\mu_{ij}(t) = \lim_{h \rightarrow 0} \frac{p_{ij}(t, t+h)}{h}$, $i, j \in \{1, 2, \dots, n\}, i \neq j$.

При $i \neq j$ μ_{ij} – это интенсивность перехода из состояния i в состояние j при $s, t, u \geq 0$.

Уравнение Колмогорова – Чепмэна имеет следующий вид:

$$p_{ij}(s, s+t+u) = \sum_{l=1}^n p_{il}(s, s+t)p_{lj}(s+t, s+t+u), i, j \in \{1, 2, \dots, n\}. \quad (1)$$

Для получения вероятностей перехода решаем систему дифференциальных уравнений. А полученные вероятности переходов связаны с интенсивностями переходов в прямой и обратной задачах [3].

$$\frac{\partial}{\partial t} p_{ij}(s, s+t) = \sum_{l=1}^n p_{il}(s, s+t)\mu_{lj}(s+t), \quad (2)$$

$$\frac{\partial}{\partial t} p_{ij}(s, s+t) = -\sum_{l=1}^n \mu_{li}(s)p_{lj}(s, s+t). \quad (3)$$

Граничные условия:

$$p_{ij}(s, s) = \delta_{ij},$$

где $\delta_{ij} = \begin{cases} 1, & \text{если, } i = j \\ 0 & \text{иначе.} \end{cases}$

В случае когда при $\mu_{ij}(t) = \mu_{ij}$ можно определить функции вероятностей перехода для любого t , мы имеем дело с так называемым однородным марковским процессом относительно времени.

Если в рассматриваемом процессе интенсивности известны, то решение дифференциальных уравнений Колмогорова и будет прямой задачей. А модель с неизвестными интенсивностями (в случае обратной задачи) ведет к оцениванию интенсивностей по имеющимся статистическим исходным данным.

При моделировании схемы марковского процесса необходимо ввести числовую оценку показателя качества, находящуюся внутри определенного интервала. Другими словами – интенсивности переходов между состояниями принадлежат определенному промежутку, у которого границы заданы условиями самого процесса.

Таким образом проверка на попадание величины интенсивности в сам интервал есть решение обратной задачи. Возникает

задача регулирования качества, решаемая путем анализа экспериментальных данных. Критерием контроля качества будет являться возможность описания входных данных в рамках модели.

$p_{ij}(t)$ известны нам из статистической базы. Зададим характеристику точности измерения:

$$|p_i^{ct} - p_i^p| \leq \epsilon_p, \quad (4)$$

где p_i^{ct} – табличные данные по вероятности; p_i^p – расчетные вероятности (1).

Система имеет решение в пределах некоторого интервала неопределенности m_{ij} , характеризуемого некоторым диапазоном вариации D_{ij} (4).

$$D_{ij} = [\min m_{ij}, \max m_{ij}]. \quad (5)$$

Задачу по поиску интервала, покрывающего область возможных значений, которые можно отнести к характеристикам изучаемого параметра (5) при удовлетворении неравенству (4) ставил еще в 1962 г. основоположник линейного программирования Л.В. Канторович [4].

Результаты исследования и их обсуждение

Исследования проводились при анализе процесса, в котором модель многих состояний, описывала состояния застрахованного лица с заболеванием сахарного диабета. Статистические данные по заболеванию сахарным диабетом получены в Медицинском информационно-аналитическом центре при министерстве здравоохранения Республики Башкортостан.

Рассматриваемая модель состоит из четырех состояний: «здоров», «болен сахарным диабетом, инсулинонезависимый тип», «болен сахарным диабетом, инсулинозависимый тип», «мертв», где λ_{ij} – интенсивности переходов между состояниями.

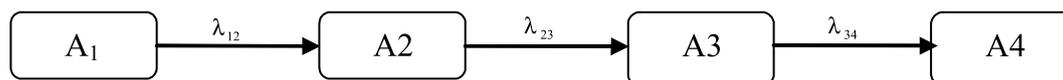


Рис. 3. Схема состояний

По представленной на рис. 3 схеме составим систему дифференциальных уравнений Колмогорова для поиска вероятности нахождения индивида в каждом из состояний:

$$\begin{cases} \frac{dp_1(t)}{dt} = -\lambda_{12}p_1(t) \\ \frac{dp_2(t)}{dt} = \lambda_{12}p_1(t) - \lambda_{23}p_2(t) \\ \frac{dp_3(t)}{dt} = \lambda_{23}p_2(t) - \lambda_{34}p_3(t) \\ \frac{dp_4(t)}{dt} = \lambda_{34}p_3(t) \end{cases}, \quad (6)$$

$p_i(t)$ ($i = \overline{1, 4}$) – вероятность состояния A_i .

Изначально предполагаем, что индивид находится в состоянии «здоров», т.е. задаем следующие начальные условия:

$$\begin{aligned} p_1(0) = 1, \quad p_2(0) = 0, \quad p_3(0) = 0, \\ p_4(0) = 0. \end{aligned} \quad (7)$$

В любой момент времени выполняется нормировочное условие, т.е. индивид находится в каком-то из четырех состояний:

$$p_1(t) + p_2(t) + p_3(t) + p_4(t) = 1. \quad (8)$$

Решаем прямую и обратную задачи. Соответственно ищем решение уравнений Колмогорова при известных интенсивностях переходов, а при неизвестных интенсивностях переходов – по статистическим данным оцениваем интенсивности переходов.

Имея средние значения интенсивностей, которые были найдены на базе статистических данных, решена обратная задача методом Рунге – Кутты 4 порядка, реализованного в программе, написанной на языке программирования C++. Вычисления были реализованы на суперкомпьютере математического факультета БашГУ.

Для вероятности p_1 из описанной схемы выражаем интенсивность перехода

$$\lambda_{12} = -\frac{\ln\left(\frac{p_1}{0,998458}\right)}{t}.$$

Другие интенсивности выражаем таким же образом. Анализируя средние значения показателей интенсивностей, выводим расчетные значения

вероятностей нахождения в состояниях A_1 – «здоров», A_2 – «болен сахарным диабетом инсулинонезависимый тип», A_3 – «болен сахарным диабетом инсулинозависимый тип», A_4 – «мертв».

На основании сопоставимости расчетных и экспериментальных данных можно заключить, что рассматриваемая модель применима на практике для задач определения, проверки, прогноза необходимого диапазона средств на борьбу с болезнью [5].

Определяем из системы (4) области неопределенности по интенсивностям λ_{ij} , взяв расчетные и экспериментальные вероятности. Для получения интервала необходимых денежных средств используем интервал изменения интенсивностей [6].

В данный момент лечение сахарного диабета в большинстве случаев сводится к компенсации болезни, кроме того сахарный диабет второго типа неизлечим, лечение в этом случае не может устранить саму причину заболевания, и оно сводится лишь к исключению симптомов и осложнений, которых на сегодняшний день существует огромное множество. Сама болезнь до конца не изучена, от нее самой как таковой не умирают, а умирают от вызванных ею осложнений. Эта болезнь уже сейчас считается мировой проблемой, которая из года в год многократно растет. Выделяемые средства на борьбу с сахарным диабетом могут варьироваться в достаточно большом интервале от 22 000 руб. до 156 000 руб. в зависимости от осложнений. Например, в 2013 г. зарегистрировано больных сахарным диабетом в России по данным ГосРегистра примерно 3 800 000 человек. По данным International Diabetes Federation (IDF) на лечение одного больного тратится в среднем 22750 рублей в год.

Величина расходов лежит в интервале $[8,3 \cdot 10^9; 5,89 \cdot 10^{11}]$, полученном на основе математического ожидания. Интервал изменения средств $[5,3 \cdot 10^{10}; 8,2 \cdot 10^{10}]$ найден из интервала изменения интенсивности перехода. Видим, что интервал изменения средств включает полученный нами интервал. То есть рассмотренная методология поиска интервала изменения необходимых денежных средств применима для реальных задач, возникающих в страховании и не обязательно медицинском.

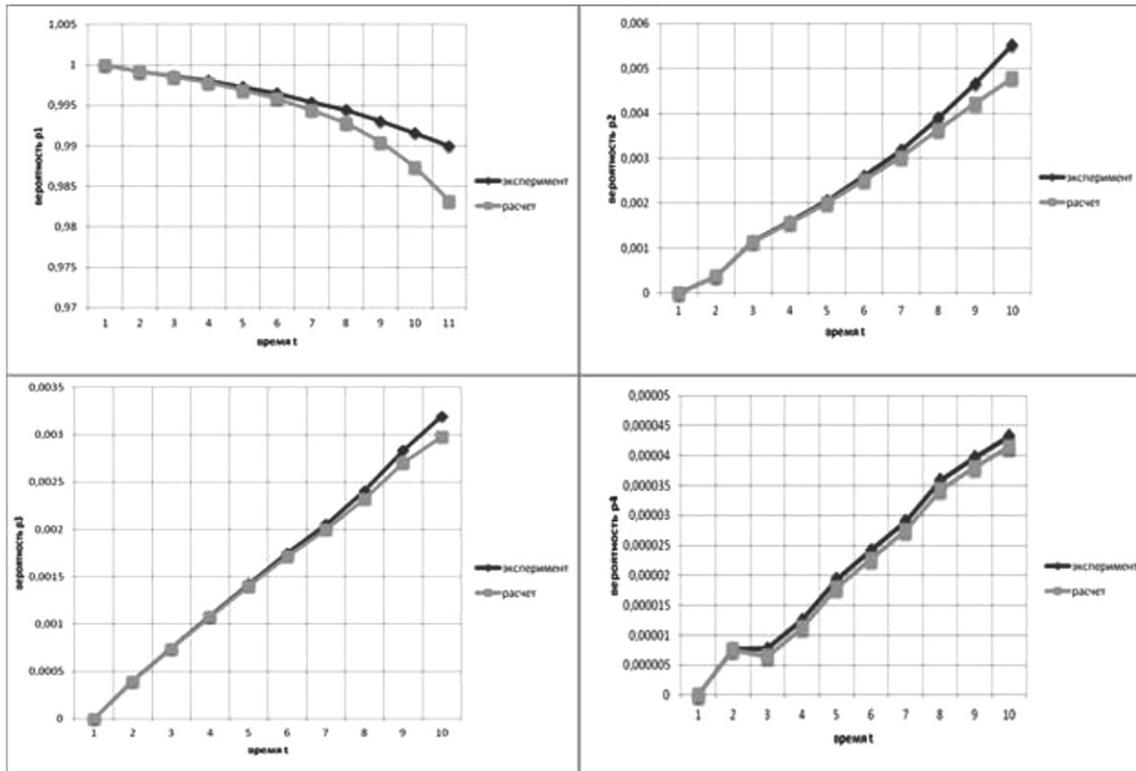


Рис. 4. Графики сравнения расчета и эксперимента

Заключение

Проведенные исследования на основе вероятностно-статистического моделирования, опирающиеся на реальные данные, полученные в Медицинском информационно-аналитическом центре при министерстве здравоохранения Республики Башкортостан, показали, что разработанная схема-модель, а также компьютерная программа могут быть использованы для предсказания и управления, а полученные результаты могут служить основой для дальнейших исследований в случаях отличных от рассматриваемого процесса заболевания сахарным диабетом. Проведенный анализ дает страховщикам в области медицины реальный механизм расчета страховых ставок.

Список литературы

1. Валитов Д.И., Спивак С.И., Райманова Г.К. Актуарные расчеты при медицинском страховании процесса заболевания сахарным диабетом // Управление риском. 2016. № 2. С. 26–34.
2. Фалин Г.И., Фалин А.И. Введение в математику финансов и инвестиций для актуариев. М.: МАКС Пресс, 2017. 327 с.
3. Спивак С.И., Райманова Г.К., Абдюшева С.Р. Обратные задачи для марковских моделей медицинского страхования // Страховое дело. 2008. № 9. С. 36–42.
4. Канторович Л.В. О некоторых новых подходах к вычислительным методам // Сибирский математический журнал. 1962. Т. 3. № 5. С. 701–709.
5. Спивак С.И., Райманова Г.К., Валитов Д.И. Актуарные расчеты при медицинском страховании процесса заболевания сахарным диабетом // Известия Уфимского научного центра РАН. 2017. № 1. С. 17–22.
6. Спивак С.И., Райманова Г.К. Математическая модель процесса заболевания туберкулезом // Системы управления и информационные технологии. 2009. № 36. С. 293–297.