ВЕСТНИК ЮГОРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2023 г. Выпуск 1 (69). С. 17–21

УДК 519.852
DOI: 10.18822/byusu20230217-21

ОПРЕДЕЛЕНИЕ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ КУСОЧНО-ЛИНЕЙНОЙ ФУНКЦИИ РИСКА МЕТОДОМ СМЕШАННОГО ОЦЕНИВАНИЯ

Носков Сергей Иванович
доктор технических наук, профессор,
профессор кафедры «Информационные системы и защита информации»,
Иркутский государственный университет путей сообщения,
Иркутск, Россия
E-mail: sergey.noskov.57@mail.ru

Предмет исследования: задача определения параметров кусочно-линейной функции риска.

Цель исследования: применить аппарат линейно-булевого программирования для решения этой задачи.

Методы и объекты исследования: объектом исследования является формализация содержательной постановки проблемы минимизации риска нежелательных последствий функционирования анализируемой системы, методами – регрессионный анализ и аппарат математического программирования.

Основные результаты исследования: описан подход к определению оценок параметров кусочно-линейной функции риска посредством применения метода смешанного оценивания, что позволяет свести эту задачу к задаче линейно-булевого программирования. Решен численный пример.

Ключевые слова: кусочно-линейная функция риска, метод наименьших модулей, метод антиробастного оценивания, метод смешанного оценивания, функция потерь, задача линейного-булевого программирования, подвыборка.

DETERMINATION OF UNKNOWN PARAMETERS OF A PIECEWISE LINEAR RISK FUNCTION BY THE METHOD OF MIXED ESTIMATION

Sergey I. Noskov
Doctor of Technical Sciences, Professor
Professor of the Department of Information Systems and Information Security,
Irkutsk State Transport University,
Irkutsk, Russia
E-mail: sergey.noskov.57@mail.ru

Subject of study: the problem of determining the parameters of a piecewise linear risk function.

The purpose of the study: apply the apparatus of linear-Boolean programming to solve this problem.

Methods and objects of research the object of the study is the formalization of a meaningful statement of the problem of minimizing the risk of undesirable consequences of the functioning of the analyzed system, the methods are regression analysis and the apparatus of mathematical programming.

The main results of the study an approach to determining estimates of the parameters of a piecewise linear risk function by using the mixed estimation method is described, which allows us to reduce this problem to a problem of linear Boolean programming. Numerical example solved.

Keywords: piecewise linear risk function, method of least modules, anti-robust estimation method, mixed estimation method, loss function, linear Boolean programming problem, subsample.

Введение

Комплексный анализ сложных систем самого различного характера часто предполагает необходимость изучения некоторых аспектов их функционирования, связанных с риском. Весьма эффективны при этом методы математического моделирования. Так, в работе [1] представлена новая байесовская структура вывода, которая обеспечивает «подгонку» математической модели под задачу оптимизации эффективности прогнозирования в отношении несбалансированных затрат на ошибочную классификацию в медицине. Статья [2] посвящена построению и оценке байесовского индекса для измерения риска информационной безопасности (ИБ) предприятий. Интегрируя мнения экспертов по ИБ, строится количественная модель этого байесовского индекса. Такой подход позволяет предприятиям осознать свой риск ИБ и принимать более эффективные решения для его снижения. С помощью метода Дельфи и подробных интервью с экспертами в предметной области факторы риска сгруппированы в пять категорий, охватывающих 29 элементов. В [3] с помощью стохастического подхода производится оценка экономического риска при покупке электрогенератора для обеспечения энергией в пиковое время. В [4] показано, как можно применить математический аппарат теории управляемых марковских полей для моделирования катастрофических рисков, вызванных природными явлениями или террористическими угрозами. Приведены примеры постановок задач долгосрочного инвестирования в сферу безопасности. Предлагается обзор методов решения стохастических задач оптимального управления, к которым сводятся такие постановки. В статье [5] анализ устойчивости суверенного долга выполняется с использованием стохастических коррелированных факторов риска, что позволяет фиксировать так называемые хвостовые эффекты. Публикация [6] посвящена построению экономико-математических моделей для оценки рисков при выборе стратегии развития предприятия на основе анализа его прибыли и издержек в условиях неопределённости. В [7] описаны модели оценки уровней привнесенного и собственного риска производственного звена, а также модель оценки действенности мероприятий, направленных на снижение риска. В качестве теоретической базы исследования применяются вероятностный и нечетко-множественный подходы. В статье [8] представлена методика построения количественных оценок риска и уязвимости с использованием прямого, обратного моделирования и некоторых методов теории чувствительности. В [9] описана математическая модель вычисления оценок информационных рисков транспортировки и распределении ресурсов в условиях неопределенности. Под информационными рисками при этом понимается опасность увеличения убытков или ущерба в результате применения соответствующих информационных технологий.

Результаты и обсуждение

В работах [10-12] описана кусочно-линейная регрессионная модель риска:

$$y_k = \max\left\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, ..., \alpha_m x_{km}\right\} + \varepsilon_k,\ k = \overline{1,n}, \tag{1}$$

где k – номер наблюдения, y и x_i, $i = \overline{1,m}$ – соответственно, зависимая и независимые переменные, значения которых известны, α_i, $i = \overline{1,m}$ – подлежащие определению параметры, ε_k – ошибки приближения, n – длина выборки.

В качестве переменных y и x_i могут использоваться негативные для объекта исследования факторы, в частности убытки, уровень травматизма, уязвимость, технические сбои, ущерб и т.д.

В [10-12] представлен способ оценивания параметров модели риска (1) с помощью метода наименьших модулей (МНМ), сводящийся в этом случае к решению задачи линейно-булевого программирования (ЛБП). В настоящей работе мы рассмотрим случай определения этих оценок на основе применения метода смешанного оценивания (МСО). В работах [13-15] он описан для решения задачи идентификации параметров линейного регрессионного уравнения:

$$y_k = \sum_{i=1}^{m} \alpha_i x_{ki} + \varepsilon_k, \ k \in P = \{1, 2, ..., n\}. \quad (2)$$

Суть МСО в простейшем случае состоит в следующем. Пусть исходная выборка данных с номерами наблюдений из индексного множества P из соображений формального или содержательного (что бывает значительно чаще) характера разделена на две непересекающиеся части (подвыборки) с номерами $N_1 = \{1, 2, ..., \bar{n}\}$ и $N_2 = \{\bar{n}+1, \bar{n}+2, ..., n\}$. При этом на подвыборке N_1 МСО «работает» как МНМ, а на N_2 - как метод антиробастного оценивания (МАО) [16-18]. Это означает, что МСО позволяет решать задачу минимизации двухкритериальной функции потерь:

$$J^1(\alpha) = \sum_{k \in N_1} |\varepsilon_k| \to \min, \quad (3)$$

$$J^2(\alpha) = \max_{k \in N_2} |\varepsilon_k| \to \min. \quad (4)$$

Применение метода смешанного оценивания при решении задачи определения параметров кусочно-линейного регрессионного уравнения (1) на основе использования результатов, описанных в работах [10-18], позволяет свести ее к следующей задаче ЛБП:

$$z_k + u_k - v_k = y_k, k = \overline{1,n}, \quad (5)$$

$$z_k \ge \alpha_i x_{ki}, \ i = \overline{1,m}, \ k = \overline{1,n}, \quad (6)$$

$$\alpha_i x_{ki} - z_k \ge (\sigma_{ki} - 1) M, \ k = \overline{1,n}, \ i = \overline{1,m}, \quad (7)$$

$$\sum_{i=1}^{m} \sigma_{ki} = 1, \ k = \overline{1,n}, \quad (8)$$

$$u_k + v_k - r \le 0, \ k \in N_2, \quad (9)$$

$$u_k \ge 0, \ v_k \ge 0, \ z_k \ge 0, \ k = \overline{1,n}, \quad (10)$$

$$\sigma_{ki} \in \{0,1\}, \ k = \overline{1,n}, \ i = \overline{1,m}, \quad (11)$$

$$\sum_{k \in N_1} (u_k + v_k) / \bar{n} + r \to \min. \quad (12)$$

Здесь M – наперед заданное большое положительное число. Полученные значения z_k, $k = \overline{1,n}$ будут представлять собой расчетные (т.е. вычисленные по модели) значения зависимой переменной y.

Задача ЛБП (5)-(12) содержит m+3n+1 вещественных и mn булевых переменных и 2n(m+1)+n-$\bar{n}$ ограничений без условий неотрицательности и булевости.

Пример

Пусть выборка данных имеет вид:

$$X = \begin{pmatrix} 2 & 5 \\ 9 & 3 \\ 7 & 4 \\ 1 & 8 \end{pmatrix}, \ y = \begin{pmatrix} 7 \\ 3 \\ 9 \\ 6 \end{pmatrix}.$$

Сформируем индексные множества N_1 и N_2 следующим образом:

$$N_1 = \{1,2\}, N_2 = \{3,4\}.$$

Вначале построим модель (1) с помощью метода наименьших модулей в соответствии с описанным в работах [10-12] алгоритмом. В результате получим:

$$y_k = \max\{0.333x_{k1}, 1.0x_{k2}\} + \varepsilon_k,\ k = \overline{1,4}. \tag{13}$$

При этом вектора z расчетных значений у и ошибок аппроксимации ε, а также максимальная ошибка r на подвыборке N_2 соответственно равны:

$$z = (5,3,4,8), \varepsilon = (2,0,5,-2), r = 5.$$

Теперь решим задачу ЛБП (5)-(12). Она имеет 15 вещественных и 8 булевых переменных, а также 26 ограничений. В результате получим:

$$y_k = \max\{0.416x_{k1}, 1.25x_{k2}\} + \varepsilon_k,\ k = \overline{1,4}, \tag{14}$$

$$z = (6.25,3.75,5,10), \varepsilon = (0.75,-0.75,4,-4), r = 4.$$

Нетрудно видеть, что параметры кусочно-линейных моделей риска (14) и (15) существенно различаются. При этом при переходе от первой модели ко второй сумма модулей ошибок аппроксимации, соответствующая МНМ, увеличилась на 0.25, а максимальная ошибка на подвыборке N_2, напротив, уменьшилась на единицу.

Заключение и выводы

В работе на основе применения полученных ранее результатов автора описан алгоритмический способ определения оценок параметров кусочно-линейной регрессионной функции риска методом смешанного оценивания, сводящийся к решению задачи линейно-булевого программирования допустимой при исследовании реальных практических проблем размерности.

Литература

1. Karapanagiotis, S. Tailored Bayes: a risk modeling framework under unequal misclassification costs / S. Karapanagiotis, U. Benedetto, S. Mukherjee, P. D. W. Kirk, P. J. Newcombe // Biostatistics. – 2023. – V. 24. – № 1. – P. 85–107.
2. Chien-Lung, Chan. Information Security Risk Modeling Using Bayesian Index Arrow // The Computer Journal. – 2011. – V. 54. – № 4. – P. 628–638.
3. Zaroni, H. Monte Carlo Simulation approach for economic risk analysis of an emergency energy generation system / H. Zaroni, L. B. Maciel, D. B. Carvalho, Edson de O. Pamplona // Energy. – 2019. – V. 172, № 1. – P. 498–508.
4. Haivoronskyy, O. O. Mathematical Modeling of Distributed Catastrophic and Terrorist Risks / O. O. Haivoronskyy, Yu. M. Ermoliev, P. S. Knopov, V. I. Norkin // Cybernetics and Systems Analysis. – 2015. – V. 51. – P. 85–95.
5. Consiglio, A. Stochastic debt sustainability analysis for sovereigns and the scope for optimization modeling / A. Consiglio, S. A. Zenios // Optimization and Engineering volume. – 2017. – № 18. – P. 537–558.
6. Ковалева, А. В. Экономико-математическая модель оценки стратегического риска при выборе стратегии развития промышленного предприятия / А. В. Ковалева. – Текст : непосредственный // Инженерный вестник Дона. – 2012. – № 1 (19). – С. 356–364.
7. Строев, С. П. Математические модели управления риском несостоятельности промышленного предприятия / С. П. Строев. – Текст : непосредственный // Continuum. Математика. Информатика. Образование. – 2021. – № 2 (22). – С. 89–98.

8. Пененко, В. В. Математические модели для изучения рисков загрязнения природной среды / В. В, Пененко, Е. А. Цветова. – Текст : непосредственный // Прикладная механика и техническая физика. – 2004. – Т. 45, № 2 (264). – С. 136–146.

9. Коробейников, А. Г., Проектирование математических моделей расчета оценки рисков перемещения материальных грузов на железнодорожных узлах с использованием лингвистических переменных / А. Г. Корабейников, А. Г. Зыков, В. И. Поляков, Д. Ю. Ашевский, С. А. Алексанин. – Текст : непосредственный // Вестник Ростовского государственного университета путей сообщения. – 2015. – № 2 (58). – С. 68–73.

10. Носков, С. И. Идентификация параметров кусочно-линейной функции риска / С. И. Носков. – Текст : непосредственный // Транспортная инфраструктура Сибирского региона. – 2017. – Т. 1. – С. 417–421.

11. Носков, С. И., Применение функции риска для моделирования экономических систем / С. И. Носков, А. А. Хоняков. – Текст : непосредственный // Южно-Сибирский научный вестник. – 2020. – № 5 (33). – С. 85–92.

12. Носков, С. И. Идентификация параметров комбинированной кусочно-линейной регрессионной модели / С. И. Носков. – Текст : непосредственный // Вестник Югорского государственного университета. – 2022. – № 4 (67). – С. 115–119.

13. Носков, С. И. О методе смешанного оценивания параметров линейной регрессии / С. И. Носков. – Текст : непосредственный // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 1. – С. 41–45.

14. Носков, С. И. Эмпирический анализ некоторых свойств метода смешанного оценивания параметров линейного регрессионного уравнения / С. И. Носков, К. С. Перфильева. – Текст : непосредственный // Наука и бизнес: пути развития. – 2020. – № 6. – С. 62–66.

15. Носков, С. И. Метод смешанного оценивания параметров линейной регрессии для данных с интервальной неопределенностью / С.И. Носков. – Текст : непосредственный // Известия Тульского государственного университета. Технические науки. – 2022. – № 9. – С. 274–277.

16. Носков, С. И. Метод антиробастного оценивания параметров линейной регрессии: число максимальных по модулю ошибок аппроксимации / С. И. Носков. – Текст : непосредственный // Южно-Сибирский научный вестник. – 2020. – № 1. – С. 51–54.

17. Носков, С. И. Выбор метода оценивания параметров линейной регрессии на основе выявления аномальных наблюдений / – Текст : непосредственный // Вестник Воронежского государственного технического университета. – 2021. – т. 17. – № 2. – С. 24–29.

18. Носков, С. И. Сравнительная оценка значимости предикторов при использовании различных методов идентификации параметров регрессионной модели / С. И. Носков. – Текст : непосредственный // Известия Тульского государственного университета. Технические науки. – 2021. – № 9. – С. 228–230.