Алгоритмизация процедуры оценки качества управленческого решения по критерию Неймана–Пирсона

Е. Н. Десятирикова[1], Л. П. Мышовская
Воронежский государственный технический университет
Воронеж, Россия
[1]science2000@ya.ru

В. И. Лютин
ВУНЦ ВВС «ВВА им. проф. Н.Е. Жуковского и Ю.А. Гагарина»
Воронеж, Россия

В. Е. Магер
Санкт-Петербургский политехнический университет Петра Великого
Санкт-Петербург, Россия

Ю. В. Хрипунов
Орловский государственный университет
Орёл, Россия

Аннотация. **Алгоритм различения гипотез с использованием критерия Неймана–Пирсона реализован в процедуре численного контроля качества принимаемых ЛПР решений.**

Ключевые слова: Байесовский риск; критерий Неймана–Пирсона; теория принятия решений; автоматизированная обучающая система

I. Введение

Качество процесса управления во многом определяется своевременностью принятия решений, квалификацией органов управления [1, 2]. В связи с этим ставится задача объективной оценки квалификации ЛПР. Для этого необходимо разработать автоматическое устройство, которое определяет оптимальное решение исходя из состояния системы и предыдущих наблюдений за ней [3, 4]. В работе приведены теоретические основания расчётов данного исследования, базирующихся на методике, предложенной в работах [5, 6], где установлены соотношения, при которых минимизируется риск и показано, что принятие решений проводится на основе сравнения отношения правдоподобия с порогом, зависящим от экономических потерь при принятии правильных или ложных решений и априорных вероятностей гипотез.

Целью работы является разработка алгоритмической процедуры оценки качества управленческого решения по критерию Неймана–Пирсона и методики расчёта экономического риска при ошибочном решении. Объектом исследования является хозяйственный риск при принятии управленческих решений в деятельности предприятия. Предметом исследования является качество принятия управленческих решений и способы прогнозирования обоснованных решений, уменьшающих риск. Методами исследований являются:

- вероятностно-статистический синтез алгоритма различения гипотез с применением критерия Неймана-Пирсона, в соответствии с которым задаётся наибольшее допустимое значение вероятности нежелательных ошибок, а именно вероятности ложных решений.
- математическое моделирование по методу Монте–Карло управленческих решений.

II. Алгоритм расчёта рисков и потерь при принятии решения

Будем считать, что ставится задача синтеза вероятностно-статистического алгоритма принятия решений в сложной экономической обстановке [7] и разработка методики анализа качества принимаемых решений в случае равенства средних значений и неравенства дисперсий результатов наблюдений при условии, что неизвестны априорные вероятности гипотез и заданной является вероятность ложной тревоги.

В работах [5, 6] установлены соотношения, при которых минимизируется риск и показано, что принятие решений проводится на основе сравнения отношения правдоподобия с порогом, зависящим от экономических потерь при принятии правильны или ложных решений и априорных вероятностей гипотез. При моделировании определяются вероятности принятия решений об одной из гипотез при условии, что известна истинная гипотеза, соответствующая наблюдению.

Суть моделирования по методу Монте–Карло заключается в проведении большого числа экспериментов, исход которых зависит от случайных факторов, определяющих наблюдаемые явления. В случае моделирования решения о выборе гипотез случайным фактором является выборка наблюдений. Если известно, которой из гипотез соответствует наблюдаемая выборка, то можно подсчитать число правильных решений об этой гипотезе и найти вероятность правильного решения как относительную частоту появления этого события.

Применить рассмотренный алгоритм моделирования можно при известных априорных вероятностях и матрице потерь. Рассмотрим случай, когда не представляется возможным назначить априорные вероятности и матрицу потерь. При этом задаётся одна из условных вероятностей, как правило, это вероятность ложной тревоги F – принять решение о гипотезе H_1, когда истинной является гипотеза H_0: $F = P(H_1|H_0)$. Этот случай соответствует моделированию по методу Монте–Карло с принятием решений по критерию Неймана-Пирсона. При неизвестной плотности распределения вероятностей достаточной статистики $w_0(\xi \mid H_0)$ продуктивный путь решения задачи заключается в моделировании по методу Монте–Карло и построении гистограммы достаточной статистики при условии справедливости гипотезы H_0 с целью определения по гистограмме порога Π по заданной вероятности ложной тревоги F.

Машинное моделирование алгоритма различения гипотез по критерию Неймана–Пирсона с применением метода Монте–Карло выполняется в несколько этапов.

1. Определение числа испытаний N и числа зон квантования M значений достаточной статистики.

2. Формирование массива значений достаточной статистики ξ_{0n}, $n = \overline{1,N}$ при условии справедливости гипотезы H_0.

3. Определение максимального $\xi_{0\max}$ и минимального $\xi_{0\min}$ значений достаточной статистики и расчёт ширины зоны квантования $\Delta z_0 = \dfrac{\xi_{0\max} - \xi_{0\min}}{M}$.

4. Расчёт числа попаданий значений достаточной статистики в каждую из зон и определение вероятностей попадания достаточной статистики в каждую из зон как относительной частоты $P_{0m} = \dfrac{N_{0m}}{N}$, $m = \overline{1,M}$, построение гистограммы и определение порога сравнения достаточной статистики.

5. Моделирование при найденном пороге Π по гипотезе H_0, в ходе которого определяются $P(H_0|H_0)$ – вероятность принятия гипотезы H_0 или вероятность правильного решения о гипотезе H_0, если она истинна, и $P(H_1|H_0)$ – вероятность принятия гипотезы H_1, если истинна гипотеза H_0 – вероятность ошибки первого рода или вероятность ложной тревоги.

6. Уточнение порога сравнения достаточной статистики путём небольших приращений порога в обе стороны и получения максимального приближения вероятности $P(H_1|H_0)$ к заданной вероятности ложной тревоги F.

7. Моделирование при найденном пороге по гипотезе H_1, в ходе которого определяются $P(H_1|H_1)$ – вероятность принятия гипотезы H_1 или вероятность правильного решения о гипотезе H_1, если она истинна, и $P(H_0|H_1)$ – вероятность принятия гипотезы H_0, если истинна гипотеза H_1 – вероятность ошибки второго рода или вероятность пропуска сообщения.

Гистограмма распределения вероятностей значений достаточной статистики строится в виде вплотную пристыкованных прямоугольников с основанием, равным ширине зоны квантования, и с высотой, равной вероятности попадания значений достаточной статистики в соответствующую зону квантования, как показано на рис. 1. Примем, что зоны нумеруются с 1 до M. Разделительный символ – стрела – расположен в точке искомого порога Π. Пунктиром показан график огибающей плотности вероятности $w_0(\xi \mid H_0)$, проведённый через середины вершин прямоугольников. Площадь расположенной справа от стрелы заштрихованной области под графиком плотности распределения вероятностей $w_0(\xi \mid H_0)$ численно равна вероятности ложной тревоги F.

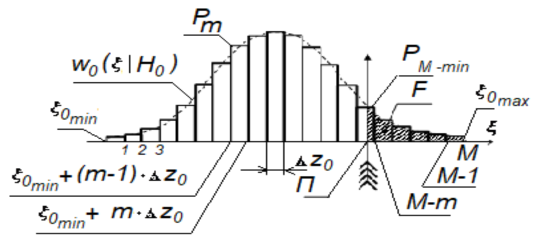

Рис. 1. Гистограмма распределения вероятностей достаточной статистики

Порог сравнения Π определяется таким образом, чтобы сумма площадей прямоугольников, расположенных справа от стрелы, включая площадь части рассечённого стрелой прямоугольника, равнялась произведению вероятности ложной тревоги на ширину зоны квантования. Вначале определяется номер зоны квантования, в которой размещается порог. Для этого суммируются вероятности в порядке убывания номера зоны квантования, начиная с последнего, до тех пор, пока сумма не станет большей или равной вероятности ложной тревоги $m = \arg\min\limits_{m} \left\{ m = \overline{1,M} \,\middle|\, S = \sum\limits_{k=M+1-m}^{M} P_k > F \right\}$.

Эта запись означает, что надо выбрать наименьшее значение *m*, при котором частичная сумма вероятностей S превышает вероятность ложной тревоги F.

Пусть при одном испытании наблюдается реализация из L результатов наблюдений. Например, каждодневные наблюдения в течение месяца курса валюты, акций или других ценных бумаг. По гипотезе H_1 реализация имеет среднее m и дисперсию σ_1^2, соответственно по гипотезе H_0 реализация имеет среднее m и дисперсию σ_0^2. Представим наблюдение вектором $\vec{Y} = [y_1, y_2, y_3, \ldots, y_L]^T$. Каждый компонент вектора является случайной величиной. Но у всех компонентов есть общее. Это среднее значение и дисперсия. Среднее значение результатов наблюдений определяется по формуле $m = \frac{1}{L}\sum_{k=1}^{L} y_k$. Дисперсия результатов наблюдений равна $d = \sigma^2 = \frac{1}{L-1}\sum_{k=1}^{L}(y_k - m)^2$.

На рис. 2 приведён алгоритм расчёта массива значений достаточной статистики ξ_{0n}, $n = \overline{1, N}$ при условии справедливости гипотезы H_0. В блоке 01 проводится ввод исходных данных, в блоке 02 вычисляются константы, необходимые для моделирования. В блоке 03 устанавливается начальное значение счётчика испытаний. В блоке 04 устанавливается начальное значение счётчика объёма выборки наблюдений. В блоке 05 вырабатываются значения выборки по гипотезе H_0. В блоке 06 проверяется полнота объёма выборочных значений до величины L. В блоке 07 содержимое счётчика объёма выборки наблюдений увеличивается на 1. В блоке 08 вычисляются значения достаточной статистики по полному объёму выборки. В блоке 09 проверяется полнота объёма моделирования до величины N. В блоке 10 содержимое счётчика объёма моделирования увеличивается на 1. В блоке 11 выводится массив значений достаточной статистики.

Алгоритм расчёта массива значений достаточной статистики ξ_{1n}, $n = \overline{1, N}$ при условии справедливости гипотезы H_1 имеет такой же вид, только с заменой индекса гипотезы «0» на «1».

На рис. 3 приведён алгоритм расчёта гистограммы достаточной статистики. В блоке 01 проводится ввод исходных данных – массив значений достаточной статистики и его объём, число зон квантования, минимальное и максимальное значения достаточной статистики. В блоке 02 определяется ширина зоны квантования, и обнуляются счётчики попаданий значений достаточной статистики в каждую из зон квантования. В блоке 03 устанавливается начальное значение счётчика значений достаточной статистики. В блоке 04 устанавливается начальное значение счётчика зон квантования. В блоке 05 проверяется очередное значение достаточной статистики на попадание в выбранную зону квантования. В блоке 06 увеличивается на 1 содержимое счётчика числа попаданий значений статистики в проверяемую зону квантования. В блоке 07 проверяется содержимое счётчика зон квантования, и если оно не достигло предельного значения, то увеличивается на 1 в блоке 08 и вновь проверяется условие блока 05. В блоке 09 проверяется полнота объёма моделирования до величины N.

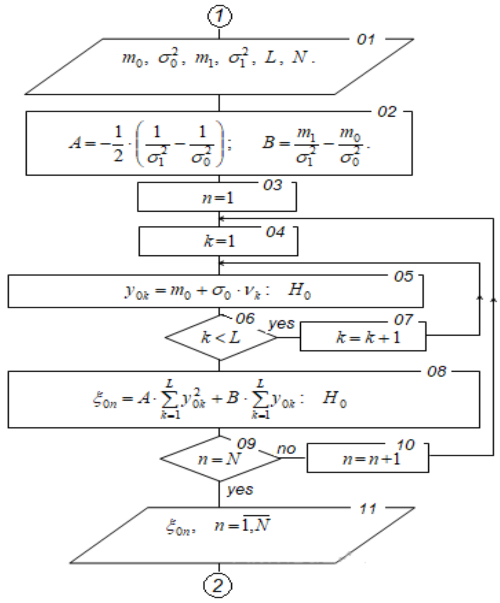

Рис. 2. Алгоритм расчёта массива достаточной статистики по гипотезе H_0

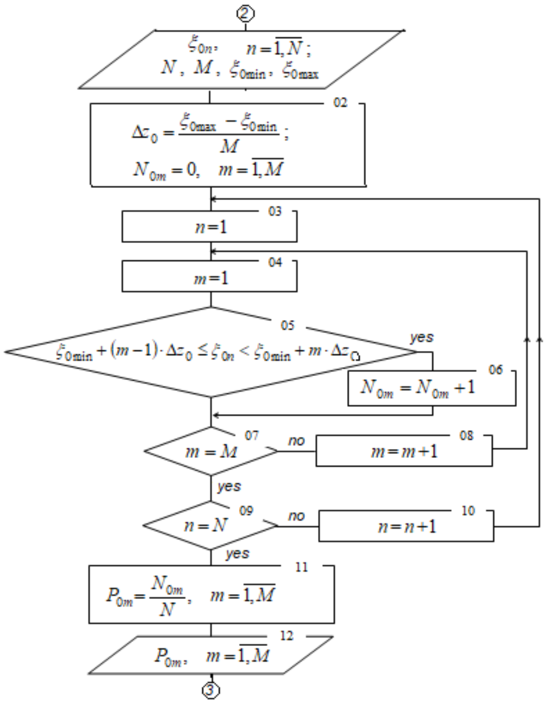

Рис. 3. Алгоритм расчёта гистограммы достаточной статистики

В блоке 10 содержимое счётчика объёма моделирования увеличивается на 1, и вновь устанавливается начальное значение счётчика зон квантования в блоке 04. Вычисления производятся до тех пор, пока содержимое обоих счётчиков не достигнет установленного значения. В блоке 11 вычисляются и в блоке 12 выводятся вероятности попадания значений достаточной статистики в каждую из зон квантования.

III. Пример реализации алгоритма

Модельная задача представляет расчёт рисков и потерь предприятия от операций на рынке ценных бумаг. Исходные данные к расчёту следующие.

1. Среднеквадратическое отклонение курса акций равно:

- по гипотезе H_0 «покупать акции» σ_0 =28.6900 у.е.
- по гипотезе H_1 «продавать акции» σ_1=16.5300 у.е.

2. Среднее значение курса акций предприятия m определяется по результатам наблюдений за месяц.

3. Потери при правильных решениях равны:

- при выборе H_0, когда правильная H_0: Π_{00}=41400.0 у.е.
- при выборе H_1, когда правильная H_1: Π_{11}=28200.0 у.е.

Потери при неправильных решениях равны:

- при выборе H_0, при верной H_1: Π_{01}= .411000E+07 у.е.
- при выборе H_1, при верной H_0: Π_{10}= .286000E+07 у.е.

4. Вероятность выбора гипотезы H_1, когда верная гипотеза H_0 (вероятность ложной тревоги), равна: $F = P(H_1|H_0)$= .100

5. Результаты наблюдений курса акций y_k за месяц:

k	1	2	3	4	5
y_k	971.000	985.900	1010.60	1017.10	999.500
k	6	7	8	9	10
y_k	1001.70	981.200	994.200	974.800	1000.50
k	11	12	13	14	15
y_k	999.100	993.800	1004.10	1003.50	992.100
k	16	17	18	19	20
y_k	970.900	993.000	959.000	1010.90	1009.10
k	21	22	23	24	25
y_k	1002.30	958.300	997.800	1001.60	998.500
k	26	27	28	29	30
y_k	996.300	977.100	987.800	986.300	993.600

Число значений наблюдений равно $L = 30$.

Отметим, что $m_1 = m_0 = m$, $\sigma_1^2 < \sigma_0^2$.

1. Среднее значение курса акций предприятия $m = \frac{1}{L}\sum_{k=1}^{L} y_k$ =992.387.

2. Дисперсия результатов наблюдений

$$\sigma^2 = \frac{1}{L-1}\sum_{k=1}^{L}(y_k - m)^2 .$$

Среднеквадратическое отклонение курса акций

$$\sigma = \sqrt{\sigma^2} = 14.7505.$$

3. Порог сравнения достаточной статистики из [5], формула (5): $\Pi = \Phi^{-1}(F, L, \sigma_0)$= 16955.6.

4. Достаточная статистика равна $\xi = \sum_{k=1}^{L}(y_k - m)^2$ =6309.77.

5. Правила принятия решений при равенстве средних значений $m_1 = m_0$ в зависимости от соотношения дисперсий $\sigma_1^2 > \sigma_0^2$ или $\sigma_1^2 < \sigma_0^2$ имеют вид (3), см. [5]. В ходе решения установлено, что достаточная статистика меньше порога $\xi < \Pi$, поэтому при $m_1 = m_0$ и $\sigma_1^2 < \sigma_0^2$ принятое решение: 1 – гипотеза H_1 «продавать акции».

6. Порог сравнения отношения правдоподобия из [5], формула (6): h =65.7585.

7. Априорная вероятность гипотезы H_1 из [5], формула (7): P_1=0.0103919.

8. Априорная вероятность гипотезы H_0 из из [5], формула (1): $P_0 = 1 - P_1$=0.989608.

9. Вероятность правильного решения о гипотезе H_0 равна

$$P(H_0|H_0) = 1 - F = 0.90.$$

10. Вероятность неправильного решения о гипотезе H_1 в принципе задана – это вероятность ложной тревоги

$$P(H_1|H_0) = F = 0.10.$$

11. Вероятность правильного решения о гипотезе H_1 из (8) (для расчёта интегральной функции χ^2-распределения с порогом Π= 16955.6, с числом степеней свободы L=30, со среднеквадратическим отклонением σ_1=16.5300 и при $\sigma_1^2 < \sigma_0^2$ искомая вероятность равна

$$P(H_1|H_1) = \Phi(\Pi, L, \sigma_1) = 0.999485.$$

12. Вероятность неправильного решения о гипотезе H_0 при $m_1 > m_0$: $P(H_0|H_1) = 1 - P(H_1|H_1)$=0.514984E–03.

13. Средние потери или средний Байесовский риск из [5], формула (2): R = 320216.0 у.е.

Список литературы

[1] E.N. Desyatirikova, L.P. Myshovskaya, A.N. Desyatirikov "Automatic Control of the Sustainability of Digital Transformation Processes in the Digital Economy" in Proc. 2019 Int. Conf. IT&QM&IS, Sochi, Russia, 23-27 Sept. 2019. DOI: 10.1109/ITQMIS.2019.8928446

[2] E.N. Desyatirikova, I.M. Gubkin, A.V. Chernenkii "Optimal Control of Open System with Asymmetric Information" in Proc. 2019 IEEE EIConRus, Saint Petersburg and Moscow, Russia, 28-31 Jan. 2019. DOI: 10.1109/EIConRus.2019.8656809

[3] V.I. Lutin, V.E. Mager, E.N. Desyatirikova, "The Processing of Signals from Sensors to Observe Objects in Various Physical Fields", in Proc. 2018 IEEE ElConRus, St. Petersburg and Moscow, Russia, Jan.29 – Febr.1, 2018. DOI: 10.1109/EIConRus.2018.8317285

[4] V.I. Lutin, Yu.V. Khripunov, E.N.Desyatirikova "Automatic Quality Control of Processes in the Online Educational Environment", in Proc. 2019 Int. Conf. IT&QM&IS, Sochi, Russia, 23-27 Sept. 2019. DOI: 10.1109/ITQMIS.2019.8928311

[5] V.N. Volkova, A.V. Loginova, E.N. Desyatirikova "Simulation modeling of a technological breakthrough in the economy" in Proc. 2018 IEEE ElConRus, St. Petersburg and Moscow, Russia, Jan.29 – Febr.1, 2018, DOI: 10.1109/EIConRus.2018.8317332

[6] L.V. Chernenkaya, E.N. Desyatirikova "Optimal planning of distributed control systems with active elements", in Proc. *IEEE II Int. Conf. on Control in Technical Systems*, St. Petersburg, Russia, 25-27 Oct. 2017. DOI: 10.1109/CTSYS.2017.8109482

[7] E.N. Desyatirikova, V.E. Belousov, S.P. Fedosova, A.A. Ievleva, "DSS design for risk management of projects", in Proc. 2017 Int. Conf. IT&QM&IS, St. Petersburg, Russia, 24-30 Sept. 2017. DOI: 10.1109/ITMQIS.2017.8085869