

THE SYSTEM OF INFORMATION SUPPLY OF BUSINESS OPERATION

Diyazitdinova A.R., Matveyeva Ye.A., Pronkina M.M.

In article attempt to generate model of system of information supply, and also to connect it with existing methodology of modelling of business processes.

Keywords: *business architecture, business process, dataware, documents circulation, eEPC-diagram.*

Диязитдинова Альфия Радмировна, к.т.н., доцент Кафедры «Экономические и информационные системы» (ЭИС) Поволжского университета телекоммуникаций и информатики (ПГУТИ). Тел. (8-846) 228-00-36. E-mail: dijazitdinova@mail.ru

Матвеева Елена Александровна, к.т.н., доцент Кафедры ЭИС ПГУТИ. Тел. (8-846) 228-00-36. E-mail: helen_matveeva@mail.ru

Пронькина Мария Михайловна, аспирант Кафедры ЭИС ПГУТИ. Тел. (8-846) 228-00-36. E-mail: helen_matveeva@mail.ru

УДК 681.518.3

ВЫБОР КРИТЕРИЯ СОГЛАСИЯ В СЛУЧАЕ ВОССТАНОВЛЕНИЯ ЗАКОНА РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ПРИМЕНИТЕЛЬНО К ЗАДАЧАМ СТАТИСТИЧЕСКОГО ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Димов Э.М., Луковкин С.В., Хабибуллин А.Р., Халимов Р.Р.

В работе представлен сравнительный анализ наиболее мощных критериев согласия эмпирического распределения вероятностей случайной величины с теоретическим. Анализ проведен на основе информации известных литературных источников. Описаны ограничения и мощности критериев при проверке гипотез в рамках задач имитационного моделирования.

Ключевые слова: имитационное моделирование, критерии согласия, критерий хи-квадрат, Андерсона-Дарлинга, типа хи-квадрат, критерий Крамера-Мизеса-Смирнова, омега-квадрат фон-Мизеса, мощность критериев, проверка статистических гипотез.

Постановка задачи

Метод статистического имитационного моделирования (СИМ) [1], позволяющий создавать компьютерные модели объектов управления (ОУ) – процессов, объектов и явлений реального мира, состоящие из оказывающих влияние на ОУ случайных величин и способов их взаимодействия, и осуществлять над ними с целью повышения эффективности управления на основе изучения поведения или состояний ОУ виртуальные эксперименты [1-2], подразумевает использование некоторого определенного массива (выборки) накопленной к текущему моменту статистической информации о моделируемом объекте (процессе, явлении), на основе анализа которой специалист по имитационному моделированию делает выводы о теоретическом законе

распределения вероятностей соответствующих объекту моделирования случайных величин и о значениях его параметров.

Знание параметров и характера распределения вероятностей случайных величин объекта управления необходимо для воспроизведения его функционирования в рамках виртуальной компьютерной модели. Другими словами, адекватность, то есть степень соответствия модели ОУ прототипу – объекту, процессу или явлению реального мира [3] – напрямую и главным образом зависит от того, насколько точно на основе исходной статистической информации об ОУ были определены закон и параметры распределения вероятностей, выделенных в результате анализа ОУ случайных величин. В случае недостижения необходимого уровня адекватности имитационной модели полученные в ходе экспериментов на ней выходные данные оказываются неприменимы к реальному ОУ в силу их искажающего результаты имитационного моделирования характера. Таким образом, результаты проверки статистических гипотез относительно закона распределения по каждой из случайных величин ОУ становятся тем «фундаментом» будущей имитационной модели ОУ, от которого в значительной мере зависит качество результатов моделирования и, следовательно, принимаемых на их основе управленческих решений.

Известно, что нахождение закона распределения вероятностей случайной величины и оценка его параметров осуществляются с использова-

нием специальных методов математической статистики, основанных на применении различных критериев согласия эмпирического (наблюдаемого по выборке исходных данных) распределения с теоретическим (предполагаемым).

В [4-5 и др.] можно найти описание огромного множества критериев согласия и их модификаций для различных целей, в том числе и для проверки гипотез относительно определенного теоретического распределения, например, нормального (критерии нормальности) или экспоненциального (критерии экспоненциальности) – так называемых специальных критериев согласия. Однако далеко не все критерии соответствуют особенностям проверки гипотез в целях имитационного моделирования в силу следующих причин: выборка исходных данных может быть достаточно мала или, напротив, велика; гипотеза о законе распределения является сложной (параметры теоретической функции распределения требуют оценки); критерий должен обладать высокой мощностью, так как именно от того, какой закон распределения для исследуемой случайной величины будет принят за истинный, зависит в конечном счете точность результатов моделирования ее значений, которая для обеспечения адекватности имитационной модели в целом должна быть максимально возможной.

Задача поиска наилучшего в отношении применения метода СИМ критерия согласия эмпирического распределения случайной величины с теоретическим приобретает в свете сказанного весомую актуальность.

Поиск решения

Из совокупности удовлетворяющих вышеуказанные требования, определенные в рамках практики решения задач имитационного моделирования, критериев для дальнейшего анализа авторами были выделены наиболее мощные: критерии типа хи-квадрат (Пирсона, Никулина), Смирнова-Крамера-фон Мизеса и критерий Андерсона-Дарлинга. Для того чтобы определить из оставшихся критериев тот, который в большей степени удовлетворяет интересам СИМ и может быть успешно использован в практике восстановления законов распределения случайных величин при статистическом анализе данных с целью создания имитационной модели, изучим индивидуальные особенности и ограничения рассматриваемых критериев.

В [4] выделенные нами критерии подробно оцениваются и, в частности, отмечается, что при проверке сложной гипотезы:

- распределение критерия хи-квадрат известно лишь приближенно;

- величина критического значения статистики Смирнова-Крамера-фон Мизеса уменьшается, то есть увеличивается количество ошибок второго рода;

- для критерия Андерсона-Дарлинга предыдущее замечание также справедливо, так как он использует нормирование статистики Смирнова-Крамера-фон Мизеса).

Таким образом, в [4] среди множества общих критериев согласия наибольшее предпочтение отдается критерию хи-квадрат, основанному на сравнении теоретической плотности распределения и эмпирической гистограммы, однако в условиях гипотезы какого типа (простой, при известных параметрах теоретического распределения, или сложной, когда параметры оцениваются по исходной выборке) справедливо данное заключение, не уточняется, таким образом, в данном случае имеет место некоторая неопределенность.

Критерий хи-квадрат Никулина исследован в [6], где вывод сводится главным образом к тому, что он мощнее критерия хи-квадрат Пирсона при близких альтернативах, что подтверждается опытными данными в [7], но требует больших вычислительных и аналитических затрат.

В исследованиях [7-8] мощность рассматриваемых критериев анализируется отдельно при проверке сложных и простых гипотез соответственно. Показано (в [8]), что хи-квадрат Пирсона, в случае проверки простой гипотезы и при условии использования в нем асимптотически оптимального группирования (АОГ) [9], действительно имеет мощность, превышающую соответствующее значение для критерия Андерсона-Дарлинга, которое, в свою очередь, выше мощности статистики Крамера-Мизеса-Смирнова.

Однако в [7] указывается, что мощность непараметрических критериев (Андерсона-Дарлинга, Крамера-Мизеса-Смирнова) в условиях проверки сложной гипотезы значительно возрастает, но при этом они теряют свойство «свободы от распределения»; оценка же неизвестных параметров распределения в интересах обеспечения максимальной мощности критериев должна осуществляться методом максимального правдоподобия [10]. По результатам имеющихся в [7] опытных данных, мощность критерия хи-квадрат Пирсона и Никулина в частном случае – при числе интервалов разбиения, равном 15 и АОГ, для выборки относительно малого объема (20 наблюдений) – выше мощности остальных

исследуемых критериев, однако заключительный вывод все же определяет критериям, на основе их мощности, приоритетность в соответствии со следующим порядком: Андерсона-Дарлинга → Крамера-Мизеса-Смирнова → хи-квадрат Никулина (при АОГ) → хи-квадрат Пирсона (АОГ). Выводы [7] о мощности критериев подтверждаются сведениями из [11].

Выводы

В результате анализа приведенных источников авторы данной работы склоняются к выбору для проверки статистических гипотез в целях имитационного моделирования статистик Андерсона-Дарлинга. В основе такого выбора лежит, прежде всего, мощность критерия, превышающая аналогичные значения других критериев, и, следовательно, максимально низкая вероятность совершения ошибок при определении закона распределения случайной величины. Это, однако, не означает, что прочие рассмотрен-

ные критерии не могут быть эффективно применены в рамках озвученных задач. Так, например, критерии типа хи-квадрат можно при наличии соответствующих знаний и практического опыта настроить, подбирая число интервалов и их длину таким образом, чтобы они различали конкурирующие гипотезы максимально эффективным образом.

Замечаниям, содержащимся в [4], о возможных сильных искажениях достоверности статистических выводов, полученных в результате применения критерия Андерсона-Дарлинга и других общих критериев согласия в случае оценки параметров распределения по одной имеющейся выборке, в результате которой, в частности, происходит уменьшение величин критических значений статистик критериев, приводящее к частым ошибкам второго рода, могут быть противопоставлены сведения [12], где данная проблема имеет решение в форме расчета индивидуальных критических значений

Таблица 1. Основные характеристики рассматриваемых критериев согласия

Название критерия	Мощность в случае простой гипотезы	Мощность в случае сложной гипотезы	Основные ограничения
Андерсона-Дарлинга	средняя	высокая	- при проверке сложной гипотезы теряет свойство «свободы от распределения»; - требует расчет индивидуальных критических значений для каждого вида распределения
Смирнова-Крамера-фон Мизеса (ω^2 Мизеса)	средняя	средняя	- в случае сложной гипотезы величина критического значения статистики критерия уменьшается, то есть увеличивается количество ошибок второго рода; - при проверке сложной гипотезы теряет свойство «свободы от распределения»
χ^2 Пирсона	высокая	низкая	- при проверке сложной гипотезы распределение критерия известно лишь приближенно; - мощность критерия существенно зависит от числа интервалов и порядка разбиения исходной выборки
χ^2 Никулина	высокая	средняя	- требует значительных вычислительных и аналитических затрат при расчете; - мощность чувствительна к числу интервалов и порядку разбиения исходной выборки

для каждого из видов распределений (нормального, экспоненциального, Вейбулла и т.д.).

В сводной таблице 1 представлены основные характеристики рассматриваемых критериев.

Аспекты практического применения рассмотренных критериев в задачах имитационного моделирования могут стать основанием для дальнейших исследований.

Литература

1. Димов Э.М., Маслов О.Н., Пчеляков С.Н., Скворцов А.Б. Новые информационные технологии: подготовка кадров и обучение персонала. Часть 2. Имитационное моделирование и управление бизнес-процессами в инфокоммуникациях. Самара: Изд. СамНЦ РАН, 2008. – 350 с.

2. Скворцов А.Б. Имитационное моделирование и технология экспертных систем в управлении инфокоммуникационной компанией. М: Радио и связь, 2002. – 232 с.
3. Богданова Е.А. Алгоритмическая модель управления бизнес-процессом эксплуатации технических средств региональной инфокоммуникационной компании. Дис. к.т.н. Курск: КГТУ, 2009. – 148 с.
4. Кобзарь А.И. Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006. – 816 с.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. – 416 с.
6. Лемешко Б.Ю., Постовалов С.Н., Чимитова Е.В. О распределениях статистики и мощности критерия типа Никулина // Заводская лаборатория. Диагностика материалов. Т. 67, № 3, 2001. – С. 52-58.
7. Лемешко Б.Ю., Лемешко С.Б., Постовалов С.Н. Сравнительный анализ мощности критериев согласия при близких конкурирующих гипотезах. Часть 2. Проверка сложных гипотез // Сибирский журнал индустриальной математики. Октябрь-декабрь, 2008. Т. XI, № 4(36). – С. 52-58.
8. Лемешко Б.Ю., Лемешко С.Б., Постовалов С.Н. Сравнительный анализ мощности критериев согласия при близких конкурирующих гипотезах. Часть 1. Проверка простых гипотез // Сибирский журнал индустриальной математики. Апрель-июнь, 2008. Т. XI, № 2(34). – С. 96-111.
9. ГОСТ Р 50.1.033-2001. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 1. Критерии типа хи-квадрат. М.: Госстандарт России, 2002. <http://www.gosthelp.ru/text/R5010332001Prikladnayasta.html>
10. ГОСТ Р 50.1.037-2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 2. Непаараметрические критерии. М.: Госстандарт России, 2002. – <http://www.gosthelp.ru/text/R5010372002Prikladnayasta.html>
11. Statistical toolkit. Statistics reference guide. Istituto Nazionale di Fisica Nucleare sezione di Genova. <http://www.ge.infn.it/geant4/analysis/HEPstatistics/gof/deployment/userdoc/statistics/index.html>
12. Anderson darling test. U.S. Department of Commerce, National Institute of Standards and Technology's web site. <http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/andedarl.htm>

SELECTION THE CRITERIA OF CONSENT IN CASE OF RESTORATION OF RANDOM VARIABLE DISTRIBUTION LAW APPLIED TO PROBLEMS OF STATISTICAL SIMULATION

Dimov Ad.M., Lukovkin S.V., Khabibullin A.R., Khalimov R.R.

This paper presents a comparative analysis of the most powerful criteria of consent of the empirical distribution of the random variable with a theoretical one. Analysis is based on information of known literary sources. Describes the limitations and capacity of criteria for testing hypotheses in the task of simulation.

Keywords: *simulation modeling, goodness of fit, chi-squared, Anderson-Darling, chi-square type criteria, test of Cramer-von Mises-Smirnov, omega-squared von Mises, power of criteria, testing of statistical hypotheses.*

Димов Эдуард Михайлович, д.т.н., профессор, заведующий Кафедрой «Экономические информационные системы» (ЭИС) Поволжского государственного университета телекоммуникаций и информатики (ПГУТИ). Тел. (8-846) 228-00-36. E-mail: dimov@psati.ru

Луковкин Сергей Владимирович, аспирант Кафедры ЭИС ПГУТИ. Тел. (8-846) 228-00-36. E-mail: slukovkin@gmail.com

Хабибуллин Артур Ринатович, аспирант Кафедры ЭИС ПГУТИ. Тел. (8-846) 228-00-36. E-mail: khalimov@bk.ru

Халимов Руслан Радикович, инженер Кафедры ЭИС ПГУТИ. Тел. (8-846) 228-00-36. E-mail: khalimov@bk.ru