

Keywords: *E-learning management, training programs, creation of electronic textbooks, distance learning, object text processing.*

Прохоров Сергей Антонович, д.т.н., профессор, заведующий Кафедрой «Информационные системы и технологии» Самарского государственного аэрокосмического университета им. акад. С.П. Королева. Тел. (8-846) 267-46-72. E-mail: sp@smr.ru

Тимченко Максим Сергеевич, экономист ООО «Самарские контейнерные линии». Тел. 8-927-206-27-17; (8-846) 230-01-78. E-mail: mairon@mail.ru

УДК 519.6

ИСПОЛЬЗОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СЛУЧАЙНЫХ ЧИСЕЛ ДЛЯ ОЦЕНКИ СТАТИСТИЧЕСКИХ СВОЙСТВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Колесова Н.А.

Предложена и обоснована методика построения репрезентативной выборки из генеральной совокупности, которая позволяет более точно оценивать статистические свойства генеральной совокупности и может быть успешно использована для повышения эффективности методов статистического анализа данных.

Ключевые слова: последовательность случайных чисел, равномерность, генеральная совокупность, выборка, репрезентативность, статистический анализ.

Введение

В настоящее время статистическая информация часто используется для вывода «обоснованных предположений» о будущем. Неоднократно замечено, что тенденции, имевшие место в прошлом и настоящем, сохраняются и в дальнейшем. Хотя на основе прошлого опыта не всегда можно предсказать будущее, все же данные о прошлом часто используются для этого. В этих случаях на основе известных значений наблюдаемого показателя можно определить его значения для некоторого времени в будущем. Этот процесс получил название прогнозирования, или анализа тенденций. Такой анализ может применяться для предсказания изменений курса акций, выполнения клинических тестов для определения границ применения нового лекарства или, например, определения средних результатов спортивных команд [1].

Современные подходы к статистическому анализу сложились в период, когда стали доступны для обработки большие наборы данных, а применение ЭВМ позволило быстро находить связь, обрабатывать, а также представлять их в удобной для человека форме.

При проведении статистического анализа широко используются следующие величины: математическое ожидание (МО), дисперсия (Д) и

среднеквадратическое отклонение (СО). Для их вычисления обычно применяется выборочный метод. Он состоит в том, что свойства генеральной совокупности устанавливаются путем изучения тех же свойств на некоторой выборке из этой совокупности [2].

Можно, например, оценивать дневную загруженность локальной сети по величине трафика, измеренного через определенные промежутки времени. Если в качестве генеральной совокупности принять значения объема трафика, измеренного через каждую минуту, то выборкой может служить массив чисел, отражающих объем трафика, проходящего по сети – например, каждую десятую минуту.

Выборка может совпадать с генеральной совокупностью, и в этом случае она будет исчерпывающей. Если выборка меньше генеральной совокупности, то результат будет неточен, так как исследуются не все элементы совокупности. Для уменьшения ошибки выборка должна обладать свойством репрезентативности, то есть отражать основные особенности генеральной совокупности [3].

Постановка и решение задачи

Необходимым условием построения репрезентативной выборки является равная вероятность включения в нее каждого элемента генеральной совокупности, то есть выборка должна обладать свойством случайности с точки зрения равномерности.

Проверка на случайность выбираемых элементов, образующих числовую последовательность, может быть осуществлена с помощью методики, описанной в [4].

Согласно данной методике влияние результатов различных проверок последовательности чисел на общий уровень ее качества может быть представлено в виде ориентированного трехуров-

невого графа G , имеющего одну корневую вершину и не содержащего петель и горизонтальных ребер в пределах одного уровня иерархии:

$$G = \langle \{F_i\}; \{D_{ij}\} \rangle,$$

где $\{F_i\}$ – множество вершин графа; $\{D_{ij}\}$ – множество дуг, соединяющих i -ю и j -ю вершины; K_0 – корневая вершина, отвечающая интегральному критерию. Дуги расположены так, что началу дуги соответствует вершина нижнего уровня иерархии (ранга), а концу дуги – вершина ранга, на единицу меньшего. На втором (нижнем) уровне расположены тесты T_i , используемые для проверки различных характеристик числовой последовательности. На уровень выше находятся обозначенные через K_i основные характеристики случайности: равномерность, стохастичность и независимость. И, наконец, корневой вершине нулевого уровня соответствует комплексный критерий K_0 оценки качества проверяемой последовательности.

Далее на полученный граф необходимо наложить систему весов или отношений предпочтения одних критериев над другими по степени их влияния на заданный элемент следующего уровня иерархии [4]. Для комплексной оценки качества числовой последовательности производится агрегирование данных, собранных в рамках иерархии G .

При этом агрегирование совершается по направлению дуг графа иерархии, где при переходе со второго уровня на первый применяется аддитивная свертка, а при переходе с первого уровня на нулевой – мультипликативная. Таким образом, пройдя последовательно снизу вверх по всем уровням иерархии G , можно путем комплексного агрегирования данных выработать суждение о качественном уровне показателя на каждой ступени иерархии (вплоть до K_0).

Пример расчета

Был измерен объем входящего и исходящего трафика, переданного по сети организации, за 24 часа с шагом в 1 мин. Эти данные были взяты в качестве генеральной совокупности, которую можно охарактеризовать с помощью МО, Д и СО.

Пусть объем выборки составляет 10% от объема генеральной совокупности. Покажем, что случайная (с точки зрения свойства равномерности) выборка является более репрезентативной, чем неслучайная. Для этого вычислим отклонение указанных параметров (МО, Д и СО),

рассчитанных для генеральной совокупности, от параметров, рассчитанных для выборок. Для проверки свойства равномерности будем использовать программный продукт, описанный в [5]. Результаты расчетов приведены в таблицах 1-2.

Таблица 1. Отклонения, полученные при расчете статистических характеристик исходящего трафика

	Статистическая характеристика выборки	Неслучайные числовые последовательности	Случайные числовые последовательности
Среднее отклонение, %	МО	16,15	1,23
	Д	20,15	1,37
	СО	18,75	2,34
Минимальное отклонение, %	МО	6,24	0,08
	Д	12,13	0,01
	СО	9,60	0,02
Максимальное отклонение, %	МО	25,38	2,56
	Д	27,94	2,48
	СО	26,31	4,86

Таблица 2. Отклонения, полученные при расчете статистических характеристик входящего трафика

	Статистическая характеристика выборки	Неслучайные числовые последовательности	Случайные числовые последовательности
Среднее отклонение, %	МО	10,37	1,96
	Д	26,77	3,95
	СО	26,17	3,23
Минимальное отклонение, %	МО	6,42	0,06
	Д	6,57	0,03
	СО	7,05	0,09
Максимальное отклонение, %	МО	14,37	3,25
	Д	41,53	6,31
	СО	44,39	6,13

Результаты наглядно показывают, что минимальные отклонения рассчитанных параметров, полученные при использовании неравномерных выборок, значительно превышают максимальные отклонения, которые могут быть получены при использовании равномерных выборок.

Заключение

Показано, что использование случайных числовых последовательностей при составлении выборки из генеральной совокупности данных позволяет более точно оценить статистические свойства генеральной совокупности.

Таким образом, предложенная методика оценки качества ПСЧ, а также разработанный на ее основе программный продукт могут быть успешно использованы для повышения эффективности методов статистического анализа данных.

Литература

1. Статистический анализ. [http://valera.asf.ru / delphi/struct /ststis.html](http://valera.asf.ru/delphi/struct/ststis.html)
2. Математическая статистика. <http://michae1983.narod.ru/t/8.htm>
3. Теоретические основы анализа данных. Выборочный метод. [http://www.statsoft.ru /home/portal/dataan/selection/repres.htm](http://www.statsoft.ru/home/portal/dataan/selection/repres.htm)
4. Колесова Н.А., Ажмухамедов И.М. Методика оценки качества последовательности случайных чисел // Вестник АГТУ. Серия: «Управление, вычислительная техника и информатика». №2, 2010. – С.141-148.
5. Свид. о регистрации программы для ЭВМ №2010614210. Программа для комплексной оценки качества последовательностей случайных чисел // Колесова Н.А., Ажмухамедов И.М. Зарег. в реестре программ для ЭВМ 30.06.2010.

USING SEQUENCES OF RANDOM NUMBERS FOR ASSESSING THE STATISTICAL PROPERTIES OF THE GENERAL AGGREGATE

Kolesova N.A.

Proposed and justified method of constructing a representative sample of the aggregate, which allows more accurately assessing the statistical properties of the general aggregate and can be successfully used to improve the efficiency of statistical data analysis.

Keywords: *sequence of random numbers, uniformity, general aggregate, sampling, representative ness, statistical analysis.*

Колесова Наталья Александровна, ассистент Кафедры «Информационная безопасность» Астраханского государственного технического университета. Тел. (8512)-61-45-10. E-mail: sangreal_88@inbox.ru

Поздравляем!

Авторов и читателей журнала ИКТ с наступающим Новым Годом! Пусть сбудутся все Ваши мечты и желания, пусть хватит творческой энергии, сил и времени, чтобы воплотить их в жизнь! Будем рады публиковать на страницах ИКТ Ваши новые материалы. Удачи и счастья в 2012 году!

Редакционный совет и редколлегия журнала ИКТ