

УДК 621.31.83.52

## **ПРИМЕНЕНИЕ ПОСЛЕДОВАТЕЛЬНЫХ СТАТИСТИЧЕСКИХ ТЕСТОВ В ИНФОРМАЦИОННЫХ СИСТЕМАХ**

студент Ефименкова К.Г.

*Научный руководитель – канд. техн. наук, доцент Бородуля А.В.*

Белорусский национальный технический университет

Минск, Беларусь

В настоящее время возникает высокая необходимость в распознавании различных типов данных, поступающих извне. Огромные массивы информации требуют обработки. В частности, последовательно поступающие одинарные данные. Такие данные можно трактовать как случайный процесс. Возникает вопрос, к какому конкретно распределению относится данный процесс. И даже если тип распределения ясен, возникает необходимость выяснить параметры распределения. Именно в таких случаях применяется последовательный тест Вальда.

Последовательный анализ – раздел математической статистики. Последовательный анализ является методом статистического исследования, характерная черта которого заключается в том, что количество наблюдений, необходимых в процессе испытания, заранее не определено. Решение об окончании эксперимента зависит на каждой данной стадии эксперимента от результатов предыдущих наблюдений.

Достоинство данного метода, применительно к проверке статистических гипотез, заключается в том, что он позволяет сконструировать такую методику проверки, которая требует, в среднем, существенно меньшего числа наблюдений, чем равная ей по надежности проверка, основанная на заранее определенном количестве наблюдений. Последовательный анализ позволяет 3 сокращать число необходимых наблюдений в среднем в два-три раза по сравнению с обычными методами, фиксирующими число наблюдений заранее.

Проблема последовательного анализа возникла в группе статистических исследований Колумбийского университета в связи с некоторыми замечаниями, высказанными капитаном Шульцем из Артиллерийского управления Министерства военно-морского флота. Милтон Фридман и В. Аллен Валлис увидели большие возможности

и далеко идущие последствия, которые могло иметь применение последовательного анализа к дальнейшему развитию теоретической статистики. В частности, они предположили, что можно сконструировать такую методику последовательного анализа, для которой возможные ошибки, связанные с принятием ложных решений, были бы точно такими же, как и у лучшей из существующих методик проверок, основанных на заранее определенном количестве наблюдений, и которая в то же время требовала бы, в среднем, существенно меньшее количество наблюдений, чем фиксированное количество наблюдений, необходимое при существующей методике проверки. Фридман и Валлис предложили также несколько примеров модификаций существующей методики проверки в смысле последовательного анализа, получив в результате в ряде случаев увеличение эффективности критерия.

Именно это предложение стало началом исследований Абрахама Вальда. В последствии было сделано множество экспериментов и публикаций на протяжении 1940-х годов. Работа Вальда была высоко оценена и отмечена научным сообществом, а его книга «Последовательный анализ» стала основополагающим трудом всего направления.

Интерес представляет модификация последовательных тестов с целью оптимизации. Возможно применить множество методов и провести ряд испытаний и проверок на эффективность. Суть исследования в том, чтобы построить классический тест Вальда и провести испытания. Далее сроятся различные модификации и проводится сравнительный анализ результатов.

Модификация состоит в том, чтобы на каждой итерации изменять границы  $A$  и  $B$ . Границы можно менять в различных вариациях. В первом случае линейно меняются лишь параметры, на основании которых строятся границы. Во втором случае строится линейная зависимость от номера итерации и значения границ, построенных на первой итерации. Строятся случаи как сужения границ, так и расширения. Применяется множество комбинаций параметров, строятся сравнительные графики и таблицы.

На практике, при применении последовательного теста на реальных данных, часто возникает проблема искажений. В последовательности наблюдаемых данных возникают наблюдения, которые не впи-

сываются в общую последовательность. Такие наблюдения называются выбросами. Возникает задача изучить модели появления таких выбросов и влияние искажений на результаты последовательного теста.

Для изучения данного вопроса строятся модели искажений при помощи генераторов. Генерируются случайные величины в заданном диапазоне с применением повышающих(понижающих) коэффициентов и применяются к исходной выборке с заданной долей вероятности.

Существуют методы сглаживания выбросов, они основываются на сравнении приращений результирующей суммы для каждой гипотезы. Это увеличивает точность определения, однако в след за этим растет и время теста.

Как правило, тесты удобно исследовать использовать язык программирования Python. В нем предусмотрены функции имитирующие различные типы распределений с заданием соответствующих параметров. Так же легко строится сравнительная визуализация результатов.

### *Литература*

1. Вальд, А. Последовательный анализ / А. Вальд // Под редакцией В.А. Севастьянова. – Москва : Наука. – 1960. – 327 с
2. Гнеденко Б. В. Курс теории вероятностей / Б. В. Гнеденко, И.Н.Коваленко - 5-е изд. – М.: Наука, 1966. – 449 с.
3. Kharin A.Y. Performance and robustness evaluation in sequential hypotheses testing // Communications in Statistics – Theory and Methods. 2016. Vol. 45, No.6 P. 1663-1709
4. Компьютерная математика Mathematica электронный учебник. [Электронный ресурс]. - Режим доступа : [http://atomas.ru/mat/Book\\_Mat/GL14/Index1.htm](http://atomas.ru/mat/Book_Mat/GL14/Index1.htm). - Дата доступа : 15.09.2015