БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УДК 519.2

ТОН
Тхат Ту

ЭФФЕКТИВНОСТЬ И РОБАСТНОСТЬ ПОСЛЕДОВАТЕЛЬНЫХ СТАТИСТИЧЕСКИХ ТЕСТОВ ПРОВЕРКИ ГИПОТЕЗ О ПАРАМЕТРАХ НЕОДНОРОДНЫХ НАБЛЮДЕНИЙ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук
по специальности 01.01.05 — теория вероятностей
и математическая статистика

Минск, 2019

Работа выполнена в **Белорусском государственном университете.**

НАУЧНЫЙ РУКОВОДИТЕЛЬ -

Харин Алексей Юрьевич,
доктор физико-математических наук, доцент,
заведующий кафедрой теории вероятностей
и математической статистики
Белорусского государственного университета.

ОФИЦИАЛЬНЫЕ ОППОНЕНТЫ:

Малинковский Юрий Владимирович,
доктор физико-математических наук, профессор,
профессор кафедры фундаментальной
и прикладной математики
УО «Гомельский государственный
университет им. Ф. Скорины»;
Волошко Валерий Анатольевич,
кандидат физико-математических наук,
старший научный сотрудник
НИЛ математических методов защиты
информации Учреждения Белорусского
государственного университета
«НИИ прикладных проблем математики
и информатики».

ОППОНИРУЮЩАЯ ОРГАНИЗАЦИЯ –

ГНУ «Институт математики
НАН Беларуси».

Защита состоится **07 июня 2019 года** в ***10.00*** на заседании совета по защите диссертаций Д 02.01.07 при Белорусском государственном университете по адресу: *Минск, ул. Ленинградская 8 (корпус юридического факультета), ауд. 407.* Телефон ученого секретаря 209-57-09.
Почтовый адрес: пр-т Независимости 4, Минск, 220030.

С диссертацией можно ознакомиться в Фундаментальной библиотеке Белорусского государственного университета.

Автореферат разослан *«   » мая* 2019 года.

Ученый секретарь
совета по защите диссертаций
кандидат физ.-мат. наук доцент Е.М. Радыно

ВВЕДЕНИЕ

Последовательный подход, предложенный американским статистиком А. Вальдом[1], широко применяется во многих задачах построения статистических тестов (решающих правил). Этот подход часто используется по двум причинам: 1) Для некоторых схем наблюдений естественно следовать последовательному подходу к построению вероятностной модели; 2) Последовательные процедуры имеют важные для практики оптимальные свойства, например, последовательный критерий (тест) отношения вероятностей (ПКОВ) минимизирует математическое ожидание случайного числа наблюдений[2], необходимых для обеспечения требуемой точности (заданных малых значений условных вероятностей ошибочных решений).

Для адекватного применения этого подхода важен анализ характеристик эффективности последовательных статистических тестов: условных вероятностей ошибочных решений и математических ожиданий случайного числа наблюдений. Однако, в общем случае имеется аналитическая сложность при расчете этих характеристик.

Для независимых одинаково распределенных наблюдений разработаны некоторые способы оценивания этих значений. В частности, А. Вальд предложил метод аппроксимации этих характеристик, основанный на игнорировании величины приращения в момент выхода тестовой статистики за пороги теста. В. И. Лотов[3] получил аппроксимации в виде рядов, из которых можно получить остаточные члены и скорость сходимости. Кроме того, доказано, что оперативная характеристика и функции математического ожидания числа наблюдений удовлетворяют интегральным уравнениям Фредгольма второго рода[4]. Другой подход к вычислению характеристик эффективности, основанный на аппроксимации случайной последовательности статистик логарифмического отношения правдоподобия поглощающей цепью Маркова, развит в работах А. Ю. Харина[5].

В рамках модели независимых неодинаково распределенных наблюдений И. Лю и С. Р. Лю[6] построили численные алгоритмы нахождения харак-

---

1 Wald, A. Sequential analysis / A. Wald. — New York: John Wiley and Sons, 1947. — 212 p.

2 Wald, A. Optimum character of the sequential probability ratio test / A. Wald, J. Wolfowitz // Ann. Math. Statist. — 1948. — Vol. 19, № 3. — P. 326–339.

3 Lotov, V. I. Asymptotic expansions in a sequential likelihood ratio test / V. I. Lotov // Theory Probab. Appl. — 1987. — Vol. 32, № 1. — P. 57–67.

4 Basseville, M. Detection of abrupt changes: Theory and Application / M. Basseville, I. V. Nikiforov. — New Jersey: Prentice Hall, 1993. — 528 p.

5 Харин, А. Ю. Робастность байесовских и последовательных статистических решающих правил / А. Ю. Харин. — Минск: БГУ, 2013. — 207 с.

6 Liu, Y. Performance analysis of sequential probability ratio test / Y. Liu, X. R. Li // Sequential Analysis. — 2013. — Vol. 32, № 4. — P. 469–497.

теристик эффективности для некоторых частных случаев.

На практике из-за условий эксперимента или ошибок регистрации некоторые наблюдения могут отсутствовать; это приводит к проблеме обработки отсутствующих значений в последовательном анализе. Кроме того, по некоторым причинам, связанным с ограничением ресурсов, времени или условий эксперимента, максимально допустимое количество наблюдений может быть зафиксировано заранее. Это приводит к появлению усеченного по числу наблюдений ПКОВ (УПКОВ), предложенного Вальдом. В случае независимых одинаково распределенных наблюдений, когда максимально допустимое количество наблюдений велико, можно использовать нормальную аппроксимацию для получения верхних оценок вероятностей ошибочных решений. Обобщенный вариант этого результата получен в работе З. Говиндараюлу[7]. Методы, предложенные Л. А. Арояном[8] и И. Лю[6], могут быть использованы для расчета характеристик эффективности в некоторых частных случаях.

В прикладных задачах гипотетические модельные предположения часто нарушаются, и это приводит к неустойчивости статистических тестов. Именно по этой причине необходимо провести анализ робастности (от англ. robust – крепкий, стойкий)[9] для последовательного теста, а для обработки данных при искажениях должны быть построены робастные тесты. Результаты анализа робастности последовательного теста проверки двух простых гипотез о параметре вероятностного распределения наблюдений (т.е. гипотез, каждая из которых однозначно определяет значение этого параметра) для модели независимых одинаково распределенных наблюдений представлены в работах Ф. С. Куанга[10] и А. Ю. Харина[11]. Для некоторых меделей данных построены робастные последовательные тесты проверки простых гипотез[5].

Большинство результатов, указанных выше, получено для независимых одинаково распределенных наблюдений, и имеется необходимость расширения этих результатов для более общего случая неоднородных (т.е. неодинаково распределенных) независимых наблюдений. В диссертации рассматривается проблема последовательной статистической проверки двух простых гипотез о значениях параметров распределений вероятностей неоднородных наблюдений. Разработаны методы вычисления характеристик эффективности, а также исследована робастность последовательных статистических тестов

[7] Govindarajulu, Z. Sequential statistics / Z. Govindarajulu. — Singapore: World Scientific, 2004. — 336 p.

[8] Aroian, L. A. Sequential analysis, direct method / L. A. Aroian // Technometrics. — 1968. — Vol. 10, № 1. — P. 125–132.

[9] Huber, P. Robust statistics / P. Huber. — New York: John Wiley and Sons, 1981. — 308 p.

[10] Quang, P. X. Robust sequential testing / P. X. Quang // Ann. Statist. — 1985. — Vol. 13, № 2. — P. 638–649.

[11] Kharin, A. Performance and robustness evaluation in sequential hypotheses testing / A. Kharin // Communications in Statistics – Theory and Methods. — 2016. — Vol. 45, № 6. — P. 1693–1709.

для модели временных рядов с трендом в частности и модели независимых разнораспределенных наблюдений в общем случае. Построен последовательный тест для модели временных рядов с трендом при наличии пропущенных наблюдений и выполнено аналитическое сравнение характеристик эффективности в этом случае с тестом Вальда. Исследован также усеченный последовательный тест, как в общем случае, так и для модели временных рядов с трендом. Кроме того, получен ряд результатов для задачи последовательной проверки многих (более двух) простых гипотез в модели временных рядов с трендом.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с крупными научными программами (проектами) и темами

Работа выполнена на кафедре теории вероятностей и математической статистики Белорусского государственного университета. Тема диссертации соответствует направлениям «Междисциплинарные исследования» и «Информационно-коммуникационные и авиакосмические технологии», определенным Перечнями приоритетных направлений научных исследований и научно-технической деятельности в Республике Беларусь на 2016–2020 годы. Результаты диссертационных исследований использованы при выполнении в Белорусском государственном университете (БГУ) следующих научно-исследовательских работ (НИР):

1. НИР по ГПНИ «Информатика и космические исследования» (2011–2018 гг.): «Разработка методов, алгоритмов и программного обеспечения мониторинга систем и процессов на основе статистических данных большого объема» (номер госрегистрации 20161355).

2. НИР в рамках Белорусского государственного университета (2016–2018 гг.): «Разработка методов и алгоритмов исследования случайных процессов, полей и их приложения» (номер госрегистрации 20162501).

Цель и задачи исследования

Целью диссертационной работы является разработка методов вычисления и анализа характеристик эффективности, а также исследования робастности последовательных статистических тестов проверки простых гипотез относительно значений параметров модели неоднородных независимых наблюдений.

Для достижения поставленной цели в диссертации решаются следующие **основные задачи**:

P1. Установить условия конечности теста, существования и конечности моментов случайного числа необходимых наблюдений, построить асимптотические разложения характеристик эффективности для последовательных тестов проверки двух простых гипотез о параметрах временных рядов с трендом.

P2. Установить условия конечности теста, существования и конечности моментов случайного числа необходимых наблюдений для M-арного и матричного последовательного теста проверки $M \geq 2$ простых гипотез о параметрах временных рядов с трендом.

P3. Построить последовательные тесты проверки двух простых гипотез о параметрах временных рядов с трендом при пропусках наблюдений и оценить уклонения характеристик эффективности для этих тестов.

P4. Оценить вероятности ошибочных решений, вероятность завершения теста до момента усечения, и доказать оптимальность усеченного последовательного теста проверки двух простых гипотез для модели временных рядов с трендом и модели неоднородных независимых наблюдений.

P5. Разработать метод вычисления характеристик эффективности для последовательных и усеченных последовательных тестов проверки двух простых гипотез в рамках модели неоднородных независимых наблюдений.

P6. Провести асимптотический анализ робастности последовательных и усеченных последовательных тестов проверки двух простых гипотез для модели неоднородных независимых наблюдений при наличии искажений.

Объектом исследования являются последовательные статистические тесты.

Предмет исследования — характеристики эффективности и робастность последовательных статистических тестов.

Научная новизна

В диссертации разработаны новые методы вычисления характеристик эффективности последовательных и усеченных последовательных тестов проверки гипотез в рамках модели неоднородных независимых наблюдений и проведён асимптотический анализ робастности этих тестов при искажении распределений вероятностей и пропусках наблюдений.

Положения, выносимые на защиту

1. Вероятностный метод аппроксимации характеристик эффективности последовательного критерия отношения вероятностей для проверки двух простых гипотез о параметрах временных рядов с трендом.

2. Новые последовательные тесты для проверки двух простых гипотез о параметрах временных рядов с трендом при пропусках наблюдений и неравенства, позволяющие исследовать уклонения характеристик эффективности для этих тестов.

3. Неравенства для характеристик эффективности усеченных последо-

вательных тестов в рамках модели неоднородных независимых наблюдений и доказательство оптимальности правила принятия решения в момент усечения последовательного теста для модели временных рядов с трендом.

4. Численный метод аппроксимации характеристик эффективности последовательных и усеченных последовательных тестов для модели неоднородных независимых наблюдений и асимптотические разложения, позволяющие анализировать робастность этих процедур.

Личный вклад соискателя ученой степени

Основные результаты в диссертации получены соискателем самостоятельно. Вклад научного руководителя в совместных работах заключается в постановке задачи и обсуждении результатов.

Апробация диссертации

Основные результаты диссертационной работы были представлены и обсуждались на следующих научных конференциях и семинарах: 73-я научная конференция студентов и аспирантов БГУ (Минск, Беларусь, 2016), Международная конференция по робастной статистике (Женева, Швейцария, 2016), 11-я международная конференция «Компьютерный анализ данных и моделирование» (Минск, Беларусь, 2016), 12-я Белорусская математическая конференция (Минск, Беларусь, 2016), Международный конгресс по информатике: «Информационные системы и технологии» (Минск, Беларусь, 2016), Международная конференция по робастной статистике (Вуллонгонг, Австралия, 2017), 11-я международная конференция: «Применение многомерного статистического анализа в экономике и оценке качества»(НИУ ВШЭ, Москва, Россия, 2018).

Результаты диссертации внедрены в учебный процесс в БГУ (1 акт о практическом использовании результатов).

Опубликованность результатов диссертации

Основные результаты диссертации опубликованы в 11 научных работах, из них 4 статьи в рецензируемых изданиях в соответствии с пунктом 18 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь (общим объемом 4.1 авторских листа), 3 статьи в сборниках трудов научных конференций и 4 тезисов.

Структура и объем диссертации

Диссертационная работа состоит из перечня используемых обозначений, введения, общей характеристики работы, четырёх глав, заключения, библиографического списка и одного приложения. Полный объем диссертации составляет 107 страниц, в том числе 7 рисунков (на 7 страницах), 14 таблиц (на 11 страницах) и 1 приложение (на 2 страницах). Библиографический список содержит 118 именований, включая 11 собственных публикаций соискателя.

ОСНОВНОЕ СОДЕРЖАНИЕ

В **главе 1** дается аналитический обзор литературы по теме диссертационного исследования.

Глава 2 посвящена решению задач **P1** и **P2** по исследованию ПКОВ для проверки двух простых гипотез о параметрах временных рядов с трендом. В **разделе 2.1** введена вероятностная модель временного ряда с трендом:

$$x_t = \theta^T \psi(t) + \xi_t,\ t \in \mathbb{N} = \{1, 2, 3, ...\}, \tag{1}$$

где $\psi(t) = (\psi_1(t), ..., \psi_m(t))^T \in \mathbb{R}^m, t \in \mathbb{N}, \psi_1(t), ..., \psi_m(t)$ – заданные неслучайные линейно независимые базисные функции тренда, $\theta = (\theta_1, ..., \theta_m)^T \in \mathbb{R}^m$ – неизвестный вектор параметров, $\{\xi_t, t \geq 1\}$ – последовательность независимых одинаково распределенных (н.о.р.) случайных велечин, $\xi_t \sim N(0, \sigma^2)$, $\sigma > 0$ – заданное значение. Рассматриваются две простые гипотезы:

$$\mathcal{H}_0 : \theta = \theta^0, \mathcal{H}_1 : \theta = \theta^1, \tag{2}$$

где $\theta^0, \theta^1 \in \mathbb{R}^m$ – заданные векторы, $\theta^0 \neq \theta^1$. ПКОВ после n наблюдений определяется следующим решающим правилом проверки гипотез (2):

$$d = \mathbf{1}_{[C_+,+\infty)}(\Lambda_n) + 2 \cdot \mathbf{1}_{(C_-,C_+)}(\Lambda_n), \tag{3}$$

$$\Lambda_n = \Lambda_n(x_1, ..., x_n) = \sum_{t=1}^{n} \lambda_t, \tag{4}$$

где $\lambda_t = \lambda_t(x_t) = \ln(p_t(x_t, \theta^1)/p_t(x_t, \theta^0))$, $p_t(x, \theta)$ – плотность распределения вероятностей (п.р.в.) величины x_t, $\mathbf{1}_D(\cdot)$ – индикаторная функция множества D. Решение $d = 2$ соответствует требованию следующего наблюдения. Решение $d = 0$ ($d = 1$) означает остановку процесса наблюдения и принятие гипотезы $\mathcal{H}_0(\mathcal{H}_1)$. В (3) C_-, C_+ – параметры теста (пороги), вычисленные по формулам: $C_+ = \ln((1 - \beta_0)/\alpha_0), C_- = \ln(\beta_0/(1 - \alpha_0))$, где α_0, β_0 – заданные допустимые значения вероятностей ошибок I и II рода соответственно.

Раздел 2.2 содержит результаты о вспомогательной последовательности дискретных случайных величин.

В **разделе 2.3** анализируется гипотетическая модель (1). Примем обозначения: $\Gamma = (\theta^0 - \theta^1)(\theta^0 - \theta^1)^T$, $\mathbf{P}_k\{\cdot\}$ и $\mathbf{E}^{(k)}(\cdot)$ – вероятностная мера и условное математическое ожидание при верной гипотезе $\mathcal{H}_k$; $\mathbf{Cov}(\cdot, \cdot)$ – ковариационная матрица случайного вектора; $tr(\cdot)$ – след матрицы; $\mu_n^{(k)} = (-1)^{k+1}\sigma_n^2/2$, $\sigma_n^2 = ((\theta^0 - \theta^1)^T\psi(n))^2/\sigma^2$, $s_n^2 = \sum_{t=1}^n \sigma_t^2, m_n^{(k)} = \sum_{t=1}^n \mu_t^{(k)}$, $H_n = \sum_{i=1}^n \psi^T(i)\psi(i), T_n = (\Lambda_1, ..., \Lambda_n)^T$, $\mu_{T_n}^{(k)} = \mathbf{E}^{(k)}(T_n), \Sigma_{T_n} = \mathbf{Cov}(T_n, T_n)$, $k \in \{0, 1\}, n \geq 1$; $n_i(\cdot, \mu, \Sigma)$ – i-мерная плотность гауссовского распределения вероятностей с параметрами μ и Σ, $i \geq 1$; $N = \inf\{n \in \mathbb{N} : \Lambda_n \notin (C_-, C_+)\}$ – случайное число наблюдений, необходимых для принятия решения в пользу $\mathcal{H}_0$ или $\mathcal{H}_1$. Эффективность последовательного теста принято характеризовать величинами: $\alpha = \mathbf{P}_0\{\Lambda_N \geq C_+\}, \beta = \mathbf{P}_1\{\Lambda_N \leq C_-\}$ – фактические значения вероятности ошибок I и II рода, $\mathbf{E}^{(k)}(N), k = 0,1$ – условные математические ожидания случайного числа наблюдений.

Теорема 2.3.[2] *Если $tr(\Gamma H_n) \to +\infty$ при $n \to +\infty$, то последовательный тест (3), (4) завершается за конечное число наблюдений с вероятностью 1.*

Теорема 2.4.[2] *В условиях теоремы 2.3 справедливы следующие выражения для характеристик эффективности последовательного теста (3), (4):*

$$\mathbf{E}^{(k)}(N) = 1 + \sum_{i=1}^{+\infty} \int\limits_{C_-}^{C_+} ds_i \int\limits_{C_-}^{C_+} ds_{i-1} ... \int\limits_{C_-}^{C_+} n_i(s, \mu_{T_i}^{(k)}, \Sigma_{T_i}) ds_1,\ k \in \{0, 1\}, \tag{5}$$

$$\alpha = \int\limits_{C_+}^{+\infty} n_1(s_1, \mu_1^{(0)}, \sigma_1^2) ds_1 + \sum_{i=2}^{+\infty} \int\limits_{C_+}^{+\infty} ds_i \int\limits_{C_-}^{C_+} ds_{i-1} ... \int\limits_{C_-}^{C_+} n_i(s, \mu_{T_i}^{(0)}, \Sigma_{T_i}) ds_1, \tag{6}$$

$$\beta = \int\limits_{-\infty}^{C_-} n_1(s_1, \mu_1^{(1)}, \sigma_1^2) ds_1 + \sum_{i=2}^{+\infty} \int\limits_{-\infty}^{C_-} ds_i \int\limits_{C_-}^{C_+} ds_{i-1} ... \int\limits_{C_-}^{C_+} n_i(s, \mu_{T_i}^{(1)}, \Sigma_{T_i}) ds_1. \tag{7}$$

Использовать формулы (5) – (7) для вычисления характеристик теста затруднительно из-за интегралов растущей кратности в правых частях этих равенств.

Теорема 2.5.[2] *Пусть $h(t) = (\theta^0 - \theta^1)^T \psi(t), \forall t \geq 1$. Если существует целое число $T \geq 1$ такое, что $h(t+T) = h(t), \forall t = 1, 2, ...$, то случайное число необходимых наблюдений N для теста (3), (4) имеет конечные моменты произвольного порядка.*

Следствие 2.2.[2] *Если функция $\psi(t)$ периодична на множестве $\mathbb{N}$, то случайное число необходимых наблюдений N имеет конечные моменты произвольного порядка.*

В **разделе 2.4** рассмотрен случай, когда $h(t) \equiv const \neq 0, \forall t \in \mathbb{N}$; тогда $\{\lambda_n, n \geq 1\}$ – последовательность н.о.р. случайных величин, и для вычисления характеристик эффективности использован численный метод решения интегральных уравнений Фредгольма второго рода[4].

В **разделе 2.5** построен метод вычисления характеристик эффективности ПКОВ на основе использования свойств вспомогательной случайной последовательности. Пусть $K \in \mathbb{N}$ – фиксированное число. Обозначим $A_0 = (-\infty, C_-), A_i = [C_{i-1}, C_i), i = \overline{1, K}, A_{K+1} = [C_+, +\infty), C_- = C_0 < C_1 < ... < C_K = C_+, C_i = C_- + ih, h = \frac{C_+ - C_-}{K}, i = \overline{1, K}$; $f_{C_-}^{C_+}(x) = \left(\left[\!\left[\frac{x - C_-}{h}\right]\!\right] + 1\right) \cdot \mathbf{1}_{(C_-, C_+)}(x) + (K+1) \cdot \mathbf{1}_{[C_+, +\infty)}(x), x \in \mathbb{R}$, где $[\![x]\!]$ – целая часть значения x. Последовательность статистик $\{\Lambda_n, n \geq 1\}$ аппроксимирована последовательностью дискретных случайных величин $\{Z_n, n \geq 1\}$:

$$Z_1 = f_{C_-}^{C_+}(\Lambda_1), Z_n = \begin{cases} 0, & Z_{n-1} = 0, \\ K+1, & Z_{n-1} = K+1, (n \geq 2) \\ f_{C_-}^{C_+}(\Lambda_n), & \text{иначе}. \end{cases}$$

Примем обозначения:

$$P^{(n)}(\theta^k) = \left(\begin{array}{c|c} I_2 & \mathbf{O}_{2\times K} \\ \hline R_n(\theta^k) & Q_n(\theta^k) \end{array} \right), n \geq 1, k \in \{0,1\},$$

$$Q_n(\theta^k) = \{q_{ij}^{(n)}(\theta^k)\}_{K\times K}, R_n(\theta^k) = \{r_{ij}^{(n)}(\theta^k)\}_{K\times 2}, n \geq 1, k \in \{0,1\},$$

$$q_{ij}^{(n)}(\theta^k) = \frac{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) \int_{A_j} n_1(x, y+\mu_n^{(k)}, \sigma_n^2) dx dy}{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) dy}, 1 \leq i, j \leq K,$$

$$r_{i1}^{(n)}(\theta^k) = \frac{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) \int_{A_0} n_1(x, y+\mu_n^{(k)}, \sigma_n^2) dx dy}{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) dy}, 1 \leq i \leq K,$$

$$r_{i2}^{(n)}(\theta^k) = \frac{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) \int_{A_{K+1}} n_1(x, y+\mu_n^{(k)}, \sigma_n^2) dx dy}{\int_{A_i} n_1(y, m_{n-1}^{(k)}, s_{n-1}^2) dy}, 1 \leq i \leq K,$$

$$S(\theta^k) = I_K + \sum_{i=1}^{+\infty} \prod_{j=1}^{i+1} Q_j(\theta^k), B(\theta^k) = R_2(\theta^k) + \sum_{i=2}^{+\infty} \prod_{j=1}^{i} Q_j(\theta^k) R_{i+1}(\theta^k),$$

$\mathbf{O}_{2\times K}$ – $2\times K$-матрица с нулевыми элементами, $B_{(j)}(\cdot)$ – j-й столбец матрицы $B(\cdot)$, $Q_1(\theta^k) = I_K$ – единичная матрица размера K, $t(\theta^k) = \mathbf{E}^{(k)}(N)$, $\pi(\theta^k)$ – распределение вероятностей случайной величины Z_1 при справедливой гипотезе $\mathcal{H}_k$, $k \in \{0,1\}$, $\mathbf{1}_K$ – K-вектор-столбец, все элементы которого равны 1.

Теорема 2.6.[2] *Если* $\inf_n tr(\Gamma\psi(n)\psi^T(n)) \geq C, C = const > 0$, *то характеристики теста* (3), (4) *удовлетворяют при* $h \to 0$ *разложениям:*

$$\alpha = (\pi(\theta^0))' B_{(2)}(\theta^0) + \pi_{K+1}(\theta^0) + O(h),$$

$$\beta = (\pi(\theta^1))' B_{(1)}(\theta^1) + \pi_0(\theta^1) + O(h),$$

$$t(\theta^k) = 1 + (\pi(\theta^k))' S(\theta^k) \cdot \mathbf{1}_K + O(h), \; k \in \{0,1\},$$

где $(\pi(\theta^k))' = (\pi_1(\theta^k), ..., \pi_K(\theta^k))$.

В **разделе 2.6** исследована робастность теста при искажениях, заданных моделью $\bar{x}_t = \theta^T\psi(t) + (1-\varepsilon)\xi_t + \varepsilon\tilde{\xi}_t$, где $\{\tilde{\xi}_t, t \geq 1\}$ – последовательность независимых одинаково распределенных случайных величин, $\tilde{\xi}_t \sim N(0, \tilde{\sigma}^2)$, $\tilde{\sigma} > 0$ – заданное значение, ξ_t и $\tilde{\xi}_t$ независимы для всех t, и $\varepsilon \in [0, 1/2)$ – уровень засорения. Обозначим характеристики теста, вычисленные аналогино $\alpha, \beta, t(\theta^k)$ заменой $\lambda_t = \lambda_t(x_t)$ на $\bar{\lambda}_t = \lambda(\bar{x}_t)$ через $\bar{\alpha}, \bar{\beta}, \bar{t}(\theta^k), k \in \{0,1\}$. В **теореме 2.7** доказано, что $\bar{\alpha} = \alpha + O(\varepsilon) + O(h), \bar{\beta} = \beta + O(\varepsilon) + O(h), \bar{t}(\theta^k) =$

$t(\theta^k) + O(\varepsilon) + O(h), k \in \{0, 1\}$, при $\varepsilon \to 0$, $h \to 0$.

В **разделе 2.7** решена задача **P2** для проверки многих простых гипотез о параметрах модели (1) с использованием M-арного ПКОВ[12] и матричного ПКОВ[13]. Для каждого теста получены достаточное условие для завершения теста за конечное число наблюдений и существования конечных моментов любого порядка для случайного числа необходимых наблюдений.

В **разделе 2.8** приводятся результаты компьютерных экспериментов, иллюстрирующие теоретические результаты.

В **главе 3** решаются задачи **P3**, **P4**. **Раздел 3.1** посвящен построению последовательного теста для модели (1) при пропусках наблюдений. Пусть $k \geq m$ – фиксированное натуральное число и значение x_{k+1} не наблюдается (m – размерность вектора параметров θ). Для построения теста будем использовать $\tilde{x}_{k+1} = \hat{\theta}^T \psi(k+1)$ вместо пропущенного наблюдения, а вместо Λ_n статистику:

$$\tilde{\Lambda}_n = \begin{cases} \Lambda_n, & n \leq k, \\ \Lambda_n + \tilde{\lambda}_{k+1} - \lambda_{k+1}, & n > k, \end{cases} \tag{8}$$

где $\tilde{\lambda}_{k+1} = \lambda_{k+1}(\tilde{x}_{k+1})$, $\hat{\theta}$ – оценка метода наименьших квадратов (ОМНК) для параметра θ, вычисленная по наблюдениям $\{x_t, t = \overline{1, k}\}$.

Обозначим: $A = [\psi(1), ..., \psi(k)]^T$ – матрица со строками $\psi^T(i), i = \overline{1, k}$, $\tilde{N} = \inf\{n \in \mathbb{N} : \tilde{\Lambda}_n \notin (C_-, C_+)\}, \tilde{\alpha} = \mathbf{P}_0\{\tilde{\Lambda}_{\tilde{N}} \geq C_+\}, \tilde{\beta} = \mathbf{P}_1\{\tilde{\Lambda}_{\tilde{N}} \leq C_-\}, a = C_+ - C_-, \Phi(\cdot)$ – функция распределения стандартного нормального закона.

Теорема 3.1.[1] *Если в рамках модели (1) $rank(A) = m$, $k \geq m$, $tr(\Gamma H_n) \to +\infty$ при $n \to +\infty$, то с вероятностью 1 одна из гипотез (2) принимается тестом (3), (8) на основе конечного числа наблюдений.*

Теорема 3.3.[1] *В условиях теоремы 3.1 для уклонений характеристик эффективности последовательных тестов (3), (4) и (3), (8) имеют место следующие неравенства ($l \in \{0, 1\}$):*

$$|\tilde{\alpha} - \alpha| \leq \mathbf{P}_0\{N > k\}\Big(1 + \Phi\big(\mu_{k+2}^{(0)}/\sigma_{k+2}\big) +$$
$$+ \sum_{i=k+3}^{+\infty} \int_0^{+\infty} \int_{-a}^{a} n_1(x, m_{i-1}^{(0)} - m_{k+1}^{(0)}, s_{i-1}^2 - s_{k+1}^2) n_1(y, x + \mu_i^{(0)}, \sigma_i^2) dx dy\Big),$$
$$|\tilde{\beta} - \beta| \leq \mathbf{P}_1\{N > k\}\Big(1 + \Phi\big(-\mu_{k+2}^{(1)}/\sigma_{k+2}\big) +$$
$$+ \sum_{i=k+3}^{+\infty} \int_{-\infty}^{0} \int_{-a}^{a} n_1(x, m_{i-1}^{(1)} - m_{k+1}^{(1)}, s_{i-1}^2 - s_{k+1}^2) n_1(y, x + \mu_i^{(1)}, \sigma_i^2) dx dy\Big);$$

---

[12] Baum, C. A sequential procedure for multihypothesis testing / C. Baum, V. V. Venugopal // IEEE Trans. Inf. Theory. — 1994. — Vol. 40, № 6. — P. 1994–2007.

[13] Tartakovsky, A. Sequential Analysis: Hypothesis Testing and Changepoint Detection / A. Tartakovsky, I. Nikiforov, M. Basseville. — Florida: CRC Press, 2015. — 603 p.

$$|\mathbf{E}^{(l)}(\tilde{N}-N)| \leq \mathbf{P}_l\{N>k\}\Big(\Phi\big((a-\mu_{k+2}^{(l)})/\sigma_{k+2}\big)+\Phi\big((a+\mu_{k+2}^{(l)})/\sigma_{k+2}\big)+$$
$$+\sum_{i=k+3}^{+\infty}\int_{-a}^{a}\int_{-a}^{a} n_1(x, m_{i-1}^{(l)}-m_{k+1}^{(l)}, s_{i-1}^2-s_{k+1}^2)n_1(y, x+\mu_i^{(l)}, \sigma_i^2)dxdy\Big).$$

На практике встречается следующая ситуация: значение x_{k+1} наблюдается с вероятностью $q \in (0,1)$. Пусть v – случайная величина, $v \in \{0,1\}$, $\mathbf{P}\{v=0\}=1-q, \mathbf{P}\{v=1\}=q$, v не зависит от последовательности $\{\xi_t, t \geq 1\}$. Построим тест вида (3), использующий вместо Λ_n статистику

$$\bar{\Lambda}_n = \begin{cases} \Lambda_n, & v=1, \\ \tilde{\Lambda}_n, & v=0. \end{cases} \qquad (9)$$

Обозначим: $\overline{N} = \inf\{n \in \mathbb{N} : \bar{\Lambda}_n \notin (C_-, C_+)\}, \bar{\alpha} = \mathbf{P}_0\{\bar{\Lambda}_{\overline{N}} \geq C_+\}, \bar{\beta} = \mathbf{P}_1\{\bar{\Lambda}_{\overline{N}} \leq C_-\}$.

Теорема 3.4.[1] *В условиях теоремы 3.1 для теста* (3), (9) *имеют место следующие неравенства* ($l \in \{0,1\}$)*:*

$$|\bar{\alpha}-\alpha| \leq (1-q)\mathbf{P}_0\{N>k\}\Big(1+\Phi\big(\mu_{k+2}^{(0)}/\sigma_{k+2}\big)+$$
$$+\sum_{i=k+3}^{+\infty}\int_{0}^{+\infty}\int_{-a}^{a} n_1(x, m_{i-1}^{(0)}-m_{k+1}^{(0)}, s_{i-1}^2-s_{k+1}^2)n_1(y, x+\mu_i^{(0)}, \sigma_i^2)dxdy\Big),$$

$$|\bar{\beta}-\beta| \leq (1-q)\mathbf{P}_1\{N>k\}\Big(1+\Phi\big(-\mu_{k+2}^{(1)}/\sigma_{k+2}\big)+$$
$$+\sum_{i=k+3}^{+\infty}\int_{-\infty}^{0}\int_{-a}^{a} n_1(x, m_{i-1}^{(1)}-m_{k+1}^{(1)}, s_{i-1}^2-s_{k+1}^2)n_1(y, x+\mu_i^{(1)}, \sigma_i^2)dxdy\Big),$$

$$|\mathbf{E}^{(l)}(\overline{N}-N)| \leq (1-q)\mathbf{P}_l\{N>k\}\times$$
$$\times\Big(\Phi\big((a-\mu_{k+2}^{(l)})/\sigma_{k+2}\big)+\Phi\big((a+\mu_{k+2}^{(l)})/\sigma_{k+2}\big)+$$
$$+\sum_{i=k+3}^{+\infty}\int_{-a}^{a}\int_{-a}^{a} n_1(x, m_{i-1}^{(l)}-m_{k+1}^{(l)}, s_{i-1}^2-s_{k+1}^2)n_1(y, x+\mu_i^{(l)}, \sigma_i^2)dxdy\Big).$$

Раздел 3.2 посвящен исследованию свойств усеченного последовательного теста как для модели неоднородных независимых наблюдений в общем случае, так и для частного случая – модели временных рядов с трендом.

Рассмотрим последовательность независимых случайных величин $\{x_t, t \geq 1\}$ с семейством п.р.в. $\{p_t(x,\theta), t \geq 1\}$, зависящих от неизвестного значения параметра θ, и задачу проверки гипотез (2) тестом (3). Пусть M – максимально допустимое число наблюдений. Если процедура не приводит к решению в пользу одной из гипотез (2) при $n \leq M-1$, то решение на основании M-го наблюдения принимается следующим образом:

$$d = \mathbf{1}_{(0,+\infty)}(\Lambda_M). \tag{10}$$

Обозначим: $P_k = (p_1^{(k)}, p_2^{(k)}, ..., p_M^{(k)})^T, Q_k = \{q_{ij}^{(k)}\}_{M\times M}$, $\tilde{Q}_k$ – обобщенная обратная матрица[14] для Q_k, $a_i^{(k)} = \mathbf{P}_k\{\Lambda_i \in (C_-, C_+)\}$, $a_{ij}^{(k)} = \mathbf{P}_k\{\Lambda_i, \Lambda_j \in (C_-, C_+)\}$, $b_i^{(k)} = \mathbf{P}_k\{\Lambda_i, \Lambda_{i+1}, \Lambda_M \in (C_-, C_+)\}$, $K_k^+ = \mathbf{P}_k\{N > M, \Lambda_M \in (0, C_+)\}$, $K_k^- = \mathbf{P}_k\{N > M, \Lambda_M \in (C_-, 0]\}$, $p_i^{(k)} = 1 - a_i^{(k)}$,

$$q_{ij}^{(k)} = \begin{cases} 1 - a_i^{(k)} - a_j^{(k)} + a_{ij}^{(k)}, & i \neq j, \\ 1 - a_i^{(k)}, & i = j, \end{cases} \quad (i, j = \overline{1, M}, k \in \{0, 1\}).$$

Теорема 3.5.[3] *Для рассмотренной модели наблюдений справедливы следующие неравенства:*

$$P_k^T \tilde{Q}_k P_k \leq \mathbf{P}_k\{N \leq M\} \leq M - 2 - (M-3)q_{MM}^{(k)} + \\ + \sum_{i=2}^{M-2} (q_{iM}^{(k)} - q_{ii}^{(k)} - b_i^{(k)}) - b_1^{(k)}, k \in \{0, 1\}.$$

Пусть α_M, β_M – вероятности ошибок первого и второго рода усеченного последовательного теста (3), (10) соответственно.

Теорема 3.6.[3] *Для рассмотренной модели наблюдений вероятности ошибок теста* (3), (10) *удовлетворяют неравенствам:*

$$\alpha_M \leq e^{-C_+}(1 - \beta_M) + K_0^+ - e^{-C_+} K_1^+,$$
$$\beta_M \leq e^{C_-}(1 - \alpha_M) + K_1^- - e^{C_-} K_0^-.$$

Замечание 3.3.[3] *Если $M = +\infty$, то $K_0^\pm = K_1^\pm = 0$ при условии, что $\mathbf{P}_k\{N < +\infty\} = 1, k \in \{0, 1\}$. Тогда неравенства в теореме 3.6 превращаются в известные неравенства Вальда для порогов ПКОВ: $A \geq \beta_\infty/(1 - \alpha_\infty), B \leq (1 - \beta_\infty)/\alpha_\infty$, где $A = e^{C_-}, B = e^{C_+}$.*

Пусть $p_+ = K_0^+ - e^{-C_+} K_1^+, p_- = K_1^- - e^{C_-} K_0^-$. Если при оценивании характеристик эффективности пренебречь приращением статистики Λ_n в момент первого выхода за границы интервала (C_-, C_+), то неравенства в теореме 3.6 обратятся в равенства. Решение полученной системы уравнений:

$$\begin{cases} \alpha_M &= (p_+ + e^{-C_+} - e^{C_- - C_+} - p_- e^{-C_+})/(1 - e^{C_- - C_+}), \\ \beta_M &= (p_- + e^{C_-} - e^{C_- - C_+} - p_+ e^{C_-})/(1 - e^{C_- - C_+}) \end{cases}$$

можно использовать для аппроксимации фактических значений α_M, β_M, однако при этом сложно аналитически вычислить значения $K_i^\pm, i \in \{0, 1\}$. При условиях $\mathbf{P}_i\{N > M\} \approx 1, i \in \{0, 1\}$, целесообразно использовать следующие аппроксимации:

$$K_i^+ \approx \mathbf{P}_i\{\Lambda_M \in (0, C_+)\}, K_i^- \approx \mathbf{P}_i\{\Lambda_M \in (C_-, 0]\}, i \in \{0, 1\}.$$

[14] Rao, C. R. Linear Statistical Inference and Its Applications / C. R. Rao. — New York: Wiley, 1965. — 522 p.

Рассмотрим теперь усеченный последовательный тест для модели временных рядов с трендом (1). Отметим, что $\mathbf{P}_0\{\Lambda_M \leq 0\} = \mathbf{P}_1\{\Lambda_M > 0\}$. Рассмотрим семейство $\wp(M)$, состоящее из всех линейных комбинаций $x_1, ..., x_M$, имеющих вид $b^T X_M + c$, $b = (b_1, ..., b_M)^T \in \mathbb{R}^M$, $\sum\limits_{i=1}^{M} b_i^2 \neq 0$, $X_M = (x_1, ..., x_M)^T$, $c \in \mathbb{R}$ и удовлетворяющих условию $\mathbf{P}_0\{b^T X_M + c \leq 0\} = \mathbf{P}_1\{b^T X_M + c > 0\}$.

Теорема 3.7.[3] *Для всех $L_M \in \wp(M)$ справедливо неравенство*

$$\mathbf{P}_0\{\Lambda_M \leq 0\} \geq \mathbf{P}_0\{L_M \leq 0\}.$$

Пусть $rank(A) = m$, где $A = [\psi(1), ..., \psi(M)]^T$ – матрица со строками $\psi^T(i), i = \overline{1, M}$. Обозначим $\gamma = \rho(\hat{\theta}, \theta^0) - \rho(\hat{\theta}, \theta^1) = 2(\theta^1 - \theta^0)^T \hat{\theta} + (\theta^0)^T \theta^0 - (\theta^1)^T \theta^1$, где $\hat{\theta}$ – ОМНК для θ, вычисленная по наблюдениям $\{x_t, t = \overline{1, M}\}$ модели (1), $\rho(x, y)$ – евклидово расстояние между векторами x и y. Вместо (10) на M-м шаге используем решающее правило:

$$\begin{cases} \text{принять } \mathcal{H}_0, & \text{если} \gamma \leq 0, \\ \text{принять } \mathcal{H}_1, & \text{если} \gamma > 0. \end{cases} \tag{11}$$

Теорема 3.8.[3] *В рамках модели наблюдений (1) для последовательного теста (3), (11) справедливы следующие неравенства:*

$$\mathbf{P}_0\{\gamma \leq 0\} \leq \mathbf{P}_0\{\Lambda_M \leq 0\}, \mathbf{P}_1\{\gamma > 0\} \leq \mathbf{P}_1\{\Lambda_M > 0\}.$$

В **разделе 3.3** приводятся результаты компьютерных экспериментов, иллюстрирующие теоретические результаты.

В **главе 4** решены задачи **P5**, **P6**. В **разделах 4.1 – 4.4** разработан метод вычисления характеристик эффективности ПКОВ и УПКОВ для модели разнораспределенных независимых наблюдений. Пусть $\{x_t, t \geq 1\}$ – последовательность независимых случайных величин с п.р.в. $\{p_t(x, \theta), t \geq 1\}$, зависящими от неизвестного параметра θ. Рассматриваются гипотезы (2) для теста (3). Обозначим $S_1^{(k)}(x) = \mathbf{P}_k\{\Lambda_1 < x\}$; для $n > 1$

$$S_n^{(k)}(x) = \mathbf{P}_k\{\Lambda_n < x, \Lambda_i \in (C_-, C_+), i = \overline{1, n-1}\}, k \in \{0, 1\}.$$

Функция $S_n^{(k)}(x)$ удовлетворяет рекуррентному соотношению:

$$S_n^{(k)}(x) = \int\limits_{C_-}^{C_+} F_n^{(k)}(x - y) dS_{n-1}^{(k)}(y), n > 1, k \in \{0, 1\},$$

где $F_n^{(k)}(x)$ – функция распределения вероятностей (ФР) случайной величины λ_n при справедливой $\mathcal{H}_k$, и $S_1^{(k)}(x) = F_1^{(k)}(x)$. Пусть $G^{(k)}(x) = \sum_{n=1}^{+\infty} S_n^{(k)}(x), k \in \{0, 1\}$. Тогда характеристики эффективности теста (3), (4) представимы следующим образом:

$$\alpha = 1 - G^{(0)}(C_-), \beta = G^{(1)}(C_-),$$

$$\mathbf{E}^{(k)}(N) = 1 + G^{(k)}(C_+) - G^{(k)}(C_-), k \in \{0, 1\}.$$

Пусть $H \in \mathbb{N}$ – фиксированное число, и $\{t_i, i = \overline{1, H}\}$ – разбиение отрезка $[C_-, C_+]$, где $t_i = C_- + (i-1)h, i = \overline{1, H}, h = \frac{C_+ - C_-}{H-1}$. Обозначим $p_n^{(k)} = \sum_{j=n}^{\infty} \mathbf{P}_k\{\Lambda_i \in (C_-, C_+), i = \overline{1, j}\}, k \in \{0, 1\}$, $f_i^{(j)} = S_j^{(0)}(t_i), i = \overline{1, H}$, $f^{(n)} = (f_1^{(n)}, ..., f_H^{(n)})^T$, $n \geq 1$, $f^{(n)} = D^{(n)} f^{(n-1)}$, $D^{(n)} = \{d_{ij}^n\}_{H \times H}, n \geq 2$, где

$$d_{ij}^n = \begin{cases} \frac{1}{2}\left[F_n^{(0)}\left(t_i - t_{j-1}\right) - F_n^{(0)}\left(t_i - t_{j+1}\right)\right], & i = \overline{1, H}, j = \overline{2, H-1}, \\ -\frac{1}{2}\left[F_n^{(0)}\left(t_i - t_1\right) + F_n^{(0)}\left(t_i - t_2\right)\right], & i = \overline{1, H}, j = 1, \\ \frac{1}{2}\left[F_n^{(0)}\left(t_i - t_{H-1}\right) + F_n^{(0)}\left(t_i - t_H\right)\right], & i = \overline{1, H}, j = H. \end{cases}$$

Теорема 4.3.[4] *Пусть* $\mathbf{E}^{(k)}(N) < +\infty, k \in \{0, 1\}$. *Если* $F_n^{(k)}(\cdot)$, $n \geq 1, k \in \{0, 1\}$*, имеют непрерывные производные второго порядка на отрезке* $[C_- - C_+, C_+ - C_-]$*, то справедливы следующие разложения при* $h \to 0$ *и* $\varepsilon_0 \to 0$*:*

$$\alpha = 1 - \sum_{i=1}^{n_0^{(0)}} f_1^{(i)} + O(h^2) + O(\varepsilon_0), \beta = \sum_{i=1}^{n_0^{(1)}} g_1^{(i)} + O(h^2) + O(\varepsilon_0),$$

$$\mathbf{E}^{(0)}(N) = 1 + \sum_{i=1}^{n_0^{(0)}} (f_H^{(i)} - f_1^{(i)}) + O(h^2) + O(\varepsilon_0),$$

$$\mathbf{E}^{(1)}(N) = 1 + \sum_{i=1}^{n_0^{(1)}} (g_H^{(i)} - g_1^{(i)}) + O(h^2) + O(\varepsilon_0),$$

где $n_0^{(k)} = n_0^{(k)}(\varepsilon_0) = \min\{n : p_n^{(k)} \leq \varepsilon_0\}, k \in \{0, 1\}$, $g^{(i)} = (g_1^{(i)}, ..., g_H^{(i)})^T, i = \overline{1, n_0^{(1)}}$*, вычисляются аналогично значениям* $f^{(i)}$ *с использованием ФР* $F_i^{(1)}(\cdot)$ *случайной величины* λ_i *при справедливой* $\mathcal{H}_1$ *вместо* $F_i^{(0)}(\cdot)$*.*

Если $q_n^{(k)} = \mathbf{P}_k\{\Lambda_n \in (C_-, C_+)\} \to 0$ при $n \to +\infty$, то $n_0^{(k)}, k \in \{0, 1\}$, целесообразно выбирать из более слабых условий: $n_0^{(k)} = \min\{n \in \mathbb{N} : q_n^{(k)} \leq \varepsilon_0\}$. Отметим, что $\mathbf{P}_0\{\Lambda_i \in (C_-, C_+), i = \overline{1, n}\} = f_H^{(n)} - f_1^{(n)} + O(h^2)$. Поэтому, если этот способ нахождения индексов $n_0^{(k)}, k \in \{0, 1\}$ затруднителен, то они могут быть выбраны из условий:

$$n_0^{(0)} = \inf\{n \in \mathbb{N} : f_H^{(n)} - f_1^{(n)} \leq \varepsilon_0\},$$
$$n_0^{(1)} = \inf\{n \in \mathbb{N} : g_H^{(n)} - g_1^{(n)} \leq \varepsilon_0\}.$$

Пусть M – максимально допустимое число наблюдений и N_M – необходимое случайное число наблюдений для завершения УПКОВ. Для разбиения $\{t_i, i = \overline{1, H}\}$, определенного выше, положим значение t_{i_0}, абсолютная величина которого минимальна, равным нулю.

Теорема 4.4.[4] *Если функции* $F_n^{(k)}(\cdot), n \geq 1, k \in \{0, 1\}$*, имеют непрерывные производные второго порядка на отрезке* $[C_- - C_+, C_+ - C_-]$*, то при* $h \to 0$

справедливы следующие разложения:

$$\alpha_M = 1 - \sum_{i=1}^{M-1} f_1^{(i)} - f_{i_0}^{(M)} + O(h^2), \beta_M = \sum_{i=1}^{M-1} g_1^{(i)} + g_{i_0}^{(M)} + O(h^2);$$

$$\mathbf{E}^{(0)}(N_M) = 1 + \sum_{i=1}^{M-1} (f_H^{(i)} - f_1^{(i)}) + O(h^2),$$

$$\mathbf{E}^{(1)}(N_M) = 1 + \sum_{i=1}^{M-1} (g_H^{(i)} - g_1^{(i)}) + O(h^2).$$

Формулы в теоремах 4.3, 4.4 (согласно **замечанию 4.3**) применимы также для непрерывных функций $F_n^{(k)}(x), k \in \{0,1\}$, без какого-либо заключения о порядке точности. Кроме того, для УПКОВ в силу ограниченного числа слагаемых в правых частях этих формул можно увеличить значение параметра дискретизации H для повышения точности.

В задаче **P6** предполагается, что значения случайных величин λ_n порождены искаженной моделью с ФР:

$$\bar{F}_n(x) = (1-\delta) F_n(x) + \delta \tilde{F}_n(x), n \geq 1,$$

где $\tilde{F}_n(x)$ – засоряющая ФР, $\delta \in [0, 1/2)$ – уровень засорения. Обозначим $\bar{p}_n^{(k)}, \bar{f}^{(n)}, \bar{D}^{(n)}, \bar{\alpha}, \bar{\alpha}_M$ – элементы, вычисленные аналогично $p_n^{(k)}$, $f^{(n)}$, $D^{(n)}$, α, α_M с заменой $F_n^{(0)}(x)$ на $\bar{F}_n^{(0)}(x), n \geq 1$, $\bar{N}$ и $\bar{N}_M$ – случайные моменты остановки при наличии искажений для ПКОВ и УПКОВ соответственно; $\hat{D}^{(n)}$ – элементы, вычисленные аналогично $D^{(n)}$ с заменой $F_n^{(0)}(x)$ на $\tilde{F}_n^{(0)}(x) - F_n^{(0)}(x), n \geq 1$; $Q^{(1)} = \hat{f}^{(1)}$ – элементы, вычисленные аналогично $f^{(1)}$ с заменой $F_1^{(0)}(x)$ на $\tilde{F}_1^{(0)}(x) - F_1^{(0)}(x)$; для $n \geq 2$:
$Q^{(n)} = \hat{D}^{(n)} D^{(n-1)} ... D^{(2)} f^{(1)} + ... + D^{(n)} D^{(n-1)} ... \hat{D}^{(2)} f^{(1)} + D^{(n)} D^{(n-1)} ... D^{(2)} \hat{f}^{(1)}$.

Теорема 4.5.[4] *Пусть* $\mathbf{E}^{(k)}(N) < +\infty, \mathbf{E}^{(k)}(\bar{N}) < +\infty, k \in \{0,1\}$. *Если функции* $F_n^{(k)}(\cdot)$ *и* $\tilde{F}_n^{(k)}(\cdot), n \geq 1, k \in \{0,1\}$, *имеют непрерывные производные второго порядка на отрезке* $[C_- - C_+, C_+ - C_-]$, *то при* $h \to 0, \delta \to 0, \varepsilon_0 \to 0$ *справедливы следующие разложения:*

$$\bar{\alpha} - \alpha = -\delta \sum_{i=1}^{n_0^{(0)}} Q_1^{(i)} + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

$$\bar{\beta} - \beta = \delta \sum_{i=1}^{n_0^{(1)}} R_1^{(i)} + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

$$\mathbf{E}^{(0)}(\bar{N}) - \mathbf{E}^{(0)}(N) = \delta \sum_{i=1}^{n_0^{(0)}} (Q_H^{(i)} - Q_1^{(i)}) + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

$$\mathbf{E}^{(1)}(\bar{N})-\mathbf{E}^{(1)}(N)=\delta\sum_{i=1}^{n_0^{(1)}}(R_H^{(i)}-R_1^{(i)})+O(h^2)+O(\delta^2)+O(\varepsilon_0),$$

где $n_0^{(k)}=\inf\{n\geq 1: p_n^{(k)}\leq\varepsilon_0, \bar{p}_n^{(k)}\leq\varepsilon_0\}, k\in\{0,1\}$, $R^{(n)}, n\geq 1$ – *элементы, вычисленные аналогично* $Q^{(n)}$ *с заменой* $F_n^{(0)}(x), \tilde{F}_n^{(0)}(x)$ *на* $F_n^{(1)}(x), \tilde{F}_n^{(1)}(x)$.

Теорема 4.6.[4] *Если функции* $F_n^{(k)}(\cdot)$ *и* $\tilde{F}_n^{(k)}(\cdot), n=\overline{1,M}, k\in\{0,1\}$, *имеют непрерывные производные второго порядка на отрезке* $[C_- - C_+, C_+ - C_-]$, *то при* $h\to 0, \delta\to 0$ *имеют место разложения:*

$$\bar{\alpha}_M-\alpha_M=-\delta(\sum_{i=1}^{M-1}Q_1^{(i)}+Q_{i_0}^{(M)})+O(h^2)+O(\delta^2),$$

$$\bar{\beta}_M-\beta_M=\delta(\sum_{i=1}^{M-1}R_1^{(i)}+R_{i_0}^{(M)})+O(h^2)+O(\delta^2),$$

$$\mathbf{E}^{(0)}(\bar{N}_M)-\mathbf{E}^{(0)}(N_M)=\delta\sum_{i=1}^{M-1}(Q_H^{(i)}-Q_1^{(i)})+O(h^2)+O(\delta^2),$$

$$\mathbf{E}^{(1)}(\bar{N}_M)-\mathbf{E}^{(1)}(N_M)=\delta\sum_{i=1}^{M-1}(R_H^{(i)}-R_1^{(i)})+O(h^2)+O(\delta^2).$$

В **разделе 4.5** предложен алгоритм построения робастного усеченного последовательного теста.

В **разделе 4.6** полученные результаты применены к модели временных рядов с трендом в задаче анализа робастности УПКОВ. Без ограничения общности пусть верна гипотеза $\mathcal{H}_0$. Рассмотрим следующие модели искажений:

$$\bar{x}_t=\theta^T\psi(t)+\bar{\xi}_t,\ \bar{\xi}_t=(1-\tau_t)\xi_t+\tau_t\tilde{\xi}_t, t\geq 1, \tag{12}$$

$$\bar{x}_t=\theta\tilde{\psi}(t)+\xi_t, t\geq 1, \parallel\tilde{\psi}(t)-\psi(t)\parallel=\max_{1\leq i\leq m}\sup_{t\geq 1}|\tilde{\psi}_i(t)-\psi_i(t)|\leq\delta, \tag{13}$$

где $\mathbf{P}\{\tau_t=0\}=1-\delta, \mathbf{P}\{\tau_t=1\}=\delta$; $\tau_t, \xi_t, \tilde{\xi}_t$ независимы, $\delta\in[0,1/2)$ – уровень засорения. Для моделей (12) и (13) в **теоремах 4.7** и **4.8** доказано, что при $h\to 0, \delta\to 0$:

$$\bar{\alpha}_M-\alpha_M=O(h^2)+O(\delta), \mathbf{E}^{(0)}(\bar{N}_M)-\mathbf{E}^{(0)}(N_M)=O(h^2)+O(\delta).$$

Для модели засорения $\bar{x}_t=\theta^T\tilde{\psi}(t)+(1-\tau_t)\xi_t+\tau_t\tilde{\xi}_t, t\geq 1$, где $\mathbf{P}\{\tau_t=0\}=1-\delta_1, \mathbf{P}\{\tau_t=1\}=\delta_1$; $\tau_t, \xi_t, \tilde{\xi}_t$ независимы, и $\parallel\tilde{\psi}(t)-\psi(t)\parallel\leq\delta_2$, $\delta_1\in[0,1], \delta_2>0$, согласно **теореме 4.9** при $h\to 0, \delta_1\to 0, \delta_2\to 0$:

$$\bar{\alpha}_M-\alpha_M=O(h^2)+O(\delta_1)+O(\delta_2),$$

$$\mathbf{E}^{(0)}(\bar{N}_M)-\mathbf{E}^{(0)}(N_M)=O(h^2)+O(\delta_1)+O(\delta_2).$$

В **разделе 4.7** приводятся результаты компьютерных экспериментов, иллюстрирующие теорию.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Разработан новый вероятностный метод аппроксимации характеристик эффективности (вероятностей ошибок I и II рода и условных математических ожиданий случайного числа необходимых наблюдений) для последовательного теста проверки двух простых гипотез о параметрах временных рядов с трендом [2, 5, 6, 8].

2. Установлены условия завершения последовательного теста за конечное число наблюдений, существования и конечности моментов случайного числа необходимых наблюдений для последовательных тестов проверки двух простых гипотез о параметрах временных рядов с трендом [2, 6].

3. Построены новые последовательные тесты проверки двух простых гипотез о параметрах временных рядов с трендом при пропусках наблюдений и получены неравенства, позволяющие оценить уклонения характеристик эффективности для таких тестов [1, 7, 9].

4. Получены новые неравенства для вероятностей ошибочных решений усеченного последовательного теста в модели неоднородных независимых наблюдений и доказана оптимальность правила принятия решения этим тестом в момент усечения в рамках модели временных рядов с трендом [3, 11].

5. Разработан новый численный метод для вычисления характеристик эффективности последовательных и усеченных последовательных тестов в модели неоднородных независимых наблюдений и построены асимптотические разложения, позволяющие анализировать робастность этих процедур [4, 10, 11].

Рекомендации по практическому использованию результатов

Результаты диссертации носят теоретический характер. Основные результаты в диссертационной работе могут быть применены для развития теории статистических выводов на основе последовательных тестов проверки гипотез. Результаты, полученные в диссертации, также могут быть использованы в учебном процессе при преподавании математической и прикладной статистики (имеется акт о внедрении результатов диссертации в учебный процесс БГУ).

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в рецензируемых научных журналах

1. *Харин, А. Ю.* Последовательная статистическая проверка гипотез о параметрах временных рядов с трендом при пропусках наблюдений / А. Ю. Харин, Тон Тхат Ту // Весцi НАН Беларусi. Сер. фiз. - мат. навук. — 2016. — № 3. — С. 38–46.

2. *Kharin, A.* Performance and robustness analysis of sequential hypotheses testing for time series with trend / A. Kharin, Ton That Tu // Austrian Journal of Statistics. — 2017. — Vol. 46, № 3&4. — P. 23–36.

3. *Харин, А. Ю.* О вычислении вероятностей ошибок усечённого последовательного критерия отношения вероятностей / А. Ю. Харин, Тон Тхат Ту // Журн. Белорус. гос. ун-та. Математика. Информатика. — 2018. — № 1. — С. 68–76.

4. *Харин, А. Ю.* Анализ и исследование робастности последовательного критерия отношения вероятностей для модели независимых неодинаково распределенных наблюдений / А. Ю. Харин, Тон Тхат Ту // Весцi НАН Беларусi. Сер. фiз. - мат. навук. — 2018. — № 2. — С. 179–192.

Статьи в сборниках материалов научных конференций

5. *Тон Тхат Ту.* Вычисление вероятностей ошибок последовательного теста для временных рядов с трендом / Тон Тхат Ту // Сборник работ 73-й научной конференции студентов и аспирантов Белорус. гос. ун-та, Минск, 16–25 мая 2016 г. В 3 ч. / БГУ, Гл. управление науки; отв. за выпуск С. Г. Берлинская. — Минск, 2016. — Ч. 1. — С. 99–103.

6. *Kharin, A.* Evaluation of sequential test characteristics for time series with a trend / A. Kharin, Ton That Tu // Computer Data Analysis and Modeling: Theoretical and Applied Stochastics: Proc. of the 11th Intern. Conf., Minsk, Sept. 6–10, 2016. / BSU; eds: S. Aivazian, P. Filzmoser, Y. Kharin. — Minsk: Publishing center of BSU, 2016. — P. 96–99.

7. *Харин, А. Ю.* Последовательные процедуры статистической проверки гипотез для временных рядов с трендом при наличии пропусков / А. Ю. Харин, Тон Тхат Ту // «Международный конгресс по информатике: информационные системы и технологии»: материалы Междунар. науч. конгресса, Минск, 24–27 окт. 2016 г. / БГУ; редкол.: С. В. Абламейко (гл. ред.), В. В. Казаченок (зам. гл. ред.) [и др.]. — Минск: БГУ, 2016. — С. 487–491.

Тезисы докладов на конференциях

8. *Kharin, A.* Performance analysis and robustification of sequential statistical decision rules / A. Kharin, Ton That Tu // Int. Conf. on Robust Statistics: Book of Abstracts, Geneva, 4–8 July 2016. / University of Geneva; chair: Maria-Pia Victoria-Feser. — Geneva, 2016. — P. 73–74.

9. *Ton That Tu.* Some results on sequential test for time series with a trend in the case of missing values / Ton That Tu // XII Белорусская математическая конференция: материалы Междунар. науч. конф, Минск, 5–10 сентября 2016 г. В 5 ч. / БГУ; ред.: С. Г. Красовский. — Минск: ИМ НАН Беларуси, 2016. — Ч. 4. — С. 20–21.

10. *Kharin, A.* Performance analysis and robustness for sequential testing of parametric hypotheses under deviations in data models / A. Kharin, Ton That Tu // International Conference on Robust Statistics: Book of Abstracts, Wollongong, Australia, 3–7 July 2017. / University of Wollongong, chair: Luke Prendergast. — Wollongong, 2017. — P. 54.

11. *Kharin, A. Y.* Performance and robustness of truncated sequential probability ratio test for time series with trend / A. Y. Kharin, Ton That Tu // XI международная конференция «Применение многомерного статистического анализа в экономике и оценке качества»: Труды, Москва, 21–23 августа 2018 г. / Национальный исследовательский университет «Высшая школа экономики»; ред: С. А. Айвазян, В. С. Мхитарян. — Москва, 2018. — С. 54–55.

РЕЗЮМЕ

Тон Тхат Ту

Эффективность и робастность последовательных статистических тестов проверки гипотез о параметрах неоднородных наблюдений

Ключевые слова: последовательный статистический тест, временные ряды с трендом, неоднородные наблюдения, искажения, пропуски наблюдений, характеристики эффективности, робастность.

Целью диссертационной работы является разработка методов вычисления и анализа характеристик эффективности, а также исследования робастности последовательных статистических тестов проверки простых гипотез относительно параметров модели неоднородных независимых наблюдений. При исследовании использовались методы теории вероятностей, математической статистики, теории матриц, численного анализа и компьютерного моделирования.

В работе получены следующие новые научные результаты. Построен новый метод аппроксимации характеристик эффективности для последовательного теста проверки двух простых гипотез о параметрах временных рядов с трендом. Для задачи проверки многих простых гипотез о параметрах этой модели установлены условия конечности теста и существования конечных моментов случайного момента остановки. При пропусках наблюдений построены новые последовательные тесты и неравенства для уклонений характеристик эффективности этих тестов. Получены новые неравенства для вероятностей ошибочных решений усеченного последовательного теста в модели независимых неоднородных наблюдений и доказана оптимальность правила принятия решения этого теста в момент усечения для модели временных рядов с трендом. Кроме того, в рамках модели независимых неоднородных наблюдений разработан новый численный метод для вычисления характеристик эффективности последовательных и усеченных последовательных тестов, и построены асимптотические разложения, позволяющие анализировать робастность этих процедур.

Полученные результаты могут быть применены для развития теории статистических выводов на основе последовательных тестов. Результаты диссертации использованы в учебном процессе при преподавании математической и прикладной статистики.

РЭЗЮМЭ

Тон Тхат Ту

Эфектыўнасць і рабаснасць паслядоўных статыстычных тэстаў праверкі гіпотэз пра параметры неаднародных назіранняў

Ключавыя словы: паслядоўны статыстычны тэст, часовыя шэрагі з трэндам, неаднародныя назіранні, пропускі назіранняў, скажэнні, характарыстыкі эфектыўнасці, рабаснасць.

Мэтай дысертацыйнай працы з'яўляецца распрацоўка метадаў вылічэння і аналізу характарыстык эфектыўнасці, а таксама даследавання рабаснасці паслядоўных статыстычных тэстаў праверкі простых гіпотэз адносна параметраў мадэлі неаднародных незалежных назіранняў. Пры даследаванні выкарыстоўваліся метады тэорыі імавернасцей, матэматычнай статыстыкі, тэорыі матрыц, вылічальнага аналізу і камп'ютэрнага мадэлявання.

У працы атрыманы наступныя новыя навуковыя вынікі. Пабудаваны новы метад апраксімацыі характарыстык эфектыўнасці для паслядоўнага тэсту праверкі двух простых гіпотэз пра параметры часовых шэрагаў з трэндам. Для задачы праверкі многіх простых гіпотэз пра параметры гэтай мадэлі устаноўлены ўмовы канечнасці тэсту і існавання канчатковых момантаў выпадковай колькасці назіранняў. Пры пропусках назіранняў пабудаваны новыя паслядоўныя тэсты і няроўнасці для ўхілення характарыстык эфектыўнасці гэтых тэстаў. Атрыманы новыя няроўнасці адносна імавернасцей памылковых рашэнняў усечанага паслядоўнага тэсту ў мадэлі неаднародных незалежных назіранняў і даказана аптымальнасць правіла прыняцця рашэння гэтым тэстам ў момант усячэння для мадэлі часовых шэрагаў з трэндам. Акрамя таго, у рамках мадэлі неаднародных незалежных назіранняў распрацаваны новы лікавы метад для вылічэння характарыстык эфектыўнасці паслядоўных і ўсечаных паслядоўных тэстаў, і пабудаваныя асімптатычныя раскладанні, якія дазваляюць аналізаваць рабаснасць гэтых працэдур.

Атрыманыя вынікі могуць быць ужытыя для развіцця тэорыі статыстычных высноў на аснове паслядоўных тэстаў. Вынікі дысертацыі выкарыстаны ў навучальным працэсе пры выкладанні матэматычнай і прыкладной статыстыкі.

SUMMARY

Ton That Tu

Performance and robustness of sequential statistical procedures for testing hypotheses on the parameters of inhomogeneous observations

Key words: sequential statistical test, time series with trend, inhomogeneous observations, distortions, missing observations, performance characteristics, robustness.

The goal of the dissertation is to develop methods for calculating and analyzing performance characteristics, as well as to analyze the robustness of sequential procedures for testing simple hypotheses on the parameters of the model of inhomogeneous independent observations. Methods of probability theory, mathematical statistics, matrix theory, numerical analysis and computer simulation were used in the research.

In the dissertation the following new scientific results are obtained. A new method for approximating the performance characteristics of the sequential statistical procedure for testing two simple hypotheses on the parameters of time series with a trend is constructed. For the problem of testing many simple hypotheses on the parameters of this model, the conditions for the test termination and the existence of finite moments of the random stopping time are established. For the case of missing observations, new sequential tests are constructed and inequalities for deviations of the performance characteristics of these tests are established. New inequalities are constructed for the probabilities of erroneous decisions of the truncated sequential test in the model of inhomogeneous independent observations and the optimality of the decision rule of this test at the truncation moment for the model of time series with a trend is proved. In the framework of the model of inhomogeneous independent observations, a new numerical method is developed for calculating the performance characteristics of sequential and truncated sequential tests, and asymptotic expansions are constructed to analyze the robustness of these procedures.

The obtained results can be applied for the development of the theory of statistical inference based on sequential tests. The results of the dissertation are used in the teaching process for delivering of the courses on mathematical and applied statistics.