УДК 004.451.7 **DOI 10.52167/1609-1817-2024-131-2-376-384**

Ж.Б. Ибраева✉, Т.Д. Иманбекова, Р. Камал
Международный университет информационных технологий, Алматы, Казахстан
E-mail: zh.ibraeva@iitu.edu.kz

ОПРЕДЕЛЕНИЕ СЛУЧАЙНОГО ПРОЦЕССА В ОДНОМЕРНОМ ВРЕМЕННОМ РЯДЕ

Аннотация. На сегодняшний день наблюдается астрономический рост объема данных, связанных с информационными потоками от различной деятельности человека. Все увеличивающийся объём пропускаемой информации создаёт для опорной сети передачи данных определенную сложность при её обработке. С другой стороны, современное общество требует высоких скоростей передачи обработанной информации.

Исследование реально измеренных данных показывают, что они обладают неравномерной интенсивностью поступления пакетов на обслуживающие сетевые устройства. Модели, созданные на основе данных, описывающих изменения одного объекта в разные моменты времени (временные ряды), обычно проявляют не стационарность, что свидетельствует о многокомпонентной структуре этих данных. Выявить и количественно выразить компоненты сложной структуры – наличие/отсутствие тенденции, периодичности, случайной компоненты является основной задачей анализа временного ряда. В ходе анализа временных рядов крайне важно выявить, является ли ряд стационарным или нестационарным, поскольку они обладают различными статистическими характеристиками и, следовательно, требуют различных методов оценки [1-3].

В данной работе проведена проверка на случайный процесс временного ряда, основанного на фактических данных, полученных с магистрали в городе Алматы, (Казахстан). Для исследования временного ряда использованы методы автокорреляции, дифференциации, применены статистические тесты для определения случайного процесса, который является одним из факторов не стационарности ряда.

Ключевые слова. Одномерный временной ряд, сетевой трафик, анализ данных, автокорреляционная функция, критерии проверки случайности.

Введение.

Распространенной моделью анализа процессов современной мультисервисной сети является модель на основе временных рядов. Временные ряды представляют временную эволюцию динамического процесса. Линейный порядок временных рядов дает им особое место в анализе данных с помощью специального набора методов. Задачи, которые возникают при изучении временного ряда:

- анализ изменений в интенсивности развития явления от периода к периоду и от даты к дате;
- определение средних значений показателей временного ряда за определенный временной отрезок;
- выявление основных закономерностей динамики изучаемого явления на различных этапах и в течение всего рассматриваемого периода;
- выявление факторов, влияющих на изменение изучаемого объекта в течение времени;
- прогнозирование будущего развития изучаемого явления.

В общем случае временной ряд представляет собой совокупность случайных величин, закон распределения и числовые характеристики которых могут зависеть от времени t. Числовые значения показателя называются уровнями временного ряда.

Общая статистическая модель временного ряда:

$$y_t = f(x_t, a) + \varepsilon_t, \quad (1)$$

где y_t - ряд наблюдений, который рассматривается как сумма компоненты $f(x_t, a)$ a - параметр.

Компонента $f(x_t, a)$ ряда может быть линейной и декомпозированной на составляющие тренд, периодичности, сезонности, которые могут зависеть от времени и на авторегрессионную составляющую при $x_t = y_{t-p}$, которая описывает зависимость между текущим значением уровня ряда и прошлым значением, сдвинутым на лаг р. При этом если модель выражена как $y_t = f(t, a)$, то она относится к классу детерминированных временных рядов. Временные ряды, модель которых описывает поведение стационарных и нестационарных процессов, выглядит как $y_t = f(y_{t-p}, a) + \varepsilon_t$ и ее относят к классу стохастических [4,5].

Анализ таких рядов связан с выявлением следующих компонентов временного ряда:

- трендовой компоненты – это плавное долгосрочное изменение среднего уровня ряда;
- сезонной компоненты – это циклические изменения уровня ряда с постоянным периодом;
- циклической компоненты – это изменения уровня ряда с переменным периодом;
- случайная компонента.

В данной работе использованы критерии и тесты в среде Matlab для определения случайного процесса в одномерном временном ряде.

Случайная составляющая представляет собой результат воздействия краткосрочных и непредсказуемых факторов, проявляющих стохастическую природу изучаемого признака. Она отражает влияние на формирование значения ряда и также включает в себя погрешность измерения.

Нестационарный случайный процесс (процесс случайного блуждания) задается следующим образом:

$$X_t = X_{t-1} + \varepsilon_t. \quad (2)$$

Нестационарный процесс, который в каком-то смысле меняется во времени, представлен следующими процессами:

- процесс с детерминистическим трендом;
- процесс случайного блуждания (random walk), например,

$$\begin{cases} y_0 = 0 \\ y_t = y_{t-1} + 2 + \varepsilon_t \end{cases}, \quad (3)$$

дисперсия процесса

$$Var(y_t) = t\sigma^2. \quad (4)$$

Процесс случайного блуждания задается как:

$$Y_t = Y_{t-1} + \varepsilon_t, \varepsilon_t \sim iid(0, \sigma^2), t = 1, \ldots, n. \quad (5)$$

Случайное блуждание – нестационарный процесс, но если к нему применить оператор взятия последовательной разницы, то можно получить стационарный ряд: «белый шум» - ряд последовательных разниц первого порядка, где под порядком понимается количество последовательных применений операторов.

Материалы и методы.

В данной работе проведен анализ измеренного одномерного ряда (интенсивность пакетов протокола MPEG). Измеренные данные представляют собой последовательность пакетов, переданных по магистральной сети за каждую секунду. Общее количество точек в ряде составляет 18 000. На рисунке 1 представлен график этих измеренных данных, где по вертикали отражено количество пакетов, поступивших за период в 5 часов, а по горизонтали - время (в секундах).

Определить тренд временного ряда можно используя линейный график, наличие автокорреляции, статистические тесты со статистической значимостью и др.

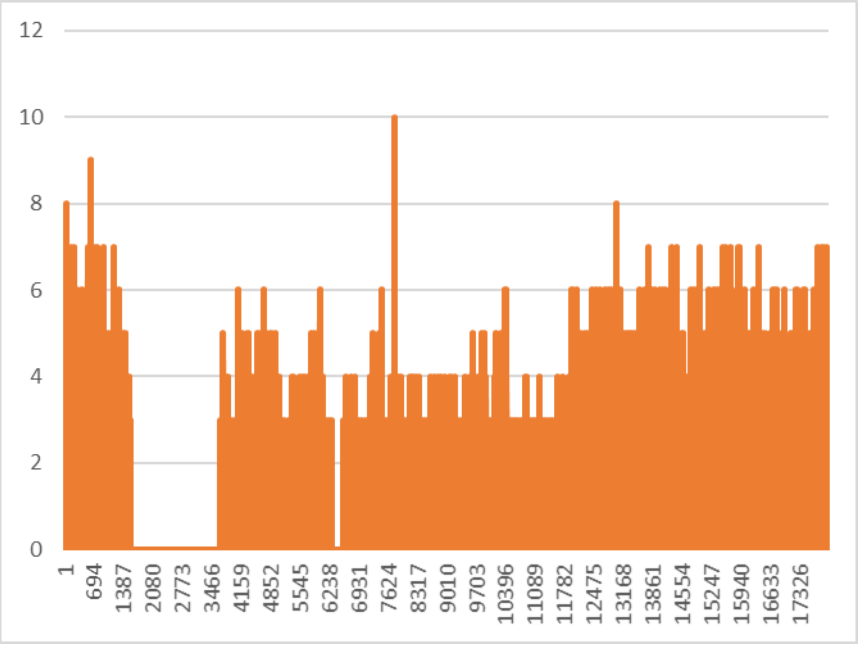

Рисунок 1 - Ряд интенсивности передачи сетевых пакетов

Результаты.

Визуальная оценка графика (рисунок 1) полученных данных временного ряда показывает, что он имеет неравномерную интенсивность, разброс уровня данных и др. Визуально можно только предположить о нестационарности ряда, дополнительно требуется проверить одномерный ряд данных несколькими другими критериями и тестами, например, корреляционную функцию на хаотический процесс и тесты проверки на случайный процесс.

Для проверки хаотичности (беспорядочности) в наборе данных используем коррелограмму.

Из рисунка 2 можно увидеть, что автокорреляционная функция убывает очень медленно, это одна из особенностей нестационарных рядов [6]. Ряд 1 – это коррелограмма, ряд 2 и ряд 3 – доверительные интервалы, соответственно нижний и верхний.

Значительная линейно затухающая автокорреляционная функция (АКФ) образца указывает на нестационарный процесс.

С помощью коррелограммы можно определить лаг наиболее высокий. В нашем случае числовое значение самого высокого уровня автокорреляции равно 0,3962896 при лаге равном 1, следовательно, исходный ряд содержит только тенденцию. Далее уровни коэффициентов автокорреляции убывают медленно и с ростом лага коррелограмма не стремится к нулю, это означает, что ряд содержит только тенденцию (тренд), что предполагает не стационарность ряда [6,7].

Характеристики коэффициента автокорреляции отражают только степень линейной связи между текущим и предыдущим рядами, и по этим данным можно сделать вывод о наличии линейной или схожей с линейной тенденции. Первым шагом при построении моделей является разность данных до тех пор, пока они не станут выглядеть стационарными [6,7].

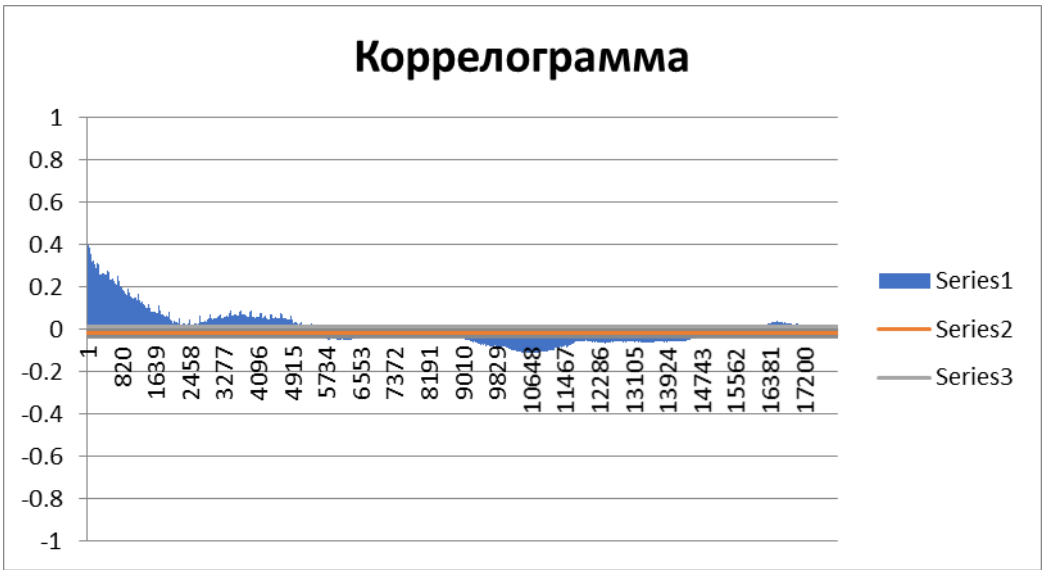

Рисунок 2 - График автокорреляции исследуемого ряда

АКФ образца разностной серии затухает быстрее, он отключается после задержки 1 (рисунок 3). Это поведение согласуется с авторегрессионной моделью первой степени (AR (1)).

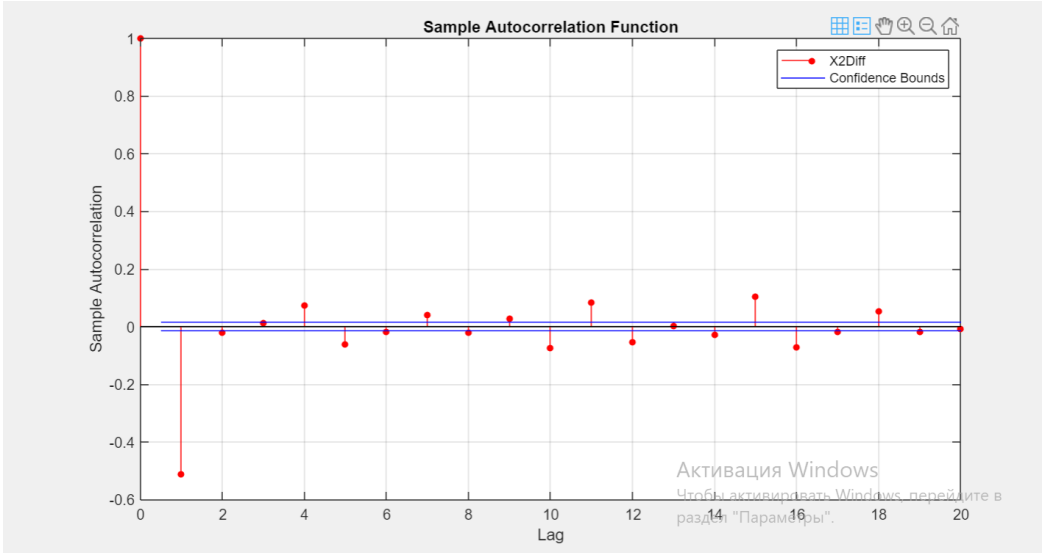

Рисунок 3 - График автокорреляции разностной серии ряда (X2Diff)

Далее ряд исследован на случайный процесс (процесс случайного блуждания). Для этого выбран *Тест отношения дисперсии для случайного блуждания* в пакете Matlab.

Тест отношения дисперсии Variance Ratio test for Random Walk оценивает нулевую гипотезу о том, что одномерный временной ряд является случайным блужданием (рисунок 4).

Variance Ratio Test for Random Walk(Y)

Null Hypothesis: Y is a random walk

Results

	Select	Null Rejected	P-Value	Test Statistic	Critical Value	Ratio	IID Innovations	Period	Significance Level
1	☐	true	0	-44.6162	1.9600	0.4887	false	2	0.0500

Рисунок 4 - Выходные данные проверки временного ряда на случайность

Уровень значимости (*Significance level*) по умолчанию должно быть 0,05, что доказывает результат теста.

Одинаково распределенные инновации (IID innovations) – скаляр или вектор логических значений, указывающих, следует ли допускать независимые IID. Отказ от нулевого случайного блуждания из-за гетероскедастичности не представляет интереса для этих случаев. Результат *IID innovations* должно быть ложным, что доказывает результат теста.

Период – это скаляр или вектор целых чисел больше единицы и меньше половины количества наблюдений в y, указывающий период, используемый для создания перекрывающихся горизонтов доходности для коэффициента дисперсии. Результат *period* равно 2, что и ожидалось.

Все результаты теста доказали, что одномерный временной ряд является случайным блужданием, что подтверждает нестационарность ряда.

В данной работе также создан объект известной современной модели ARIMA для случайного блуждания, представленного в виде (1) формулы-уравнения. Модель случайного блуждания является простейшим примером ARIMA-модели.

Смоделирован график 100 путей длиной 100 от случайного блуждания. Результаты моделирования показаны на рисунках 5 и 6. Программный код моделирования показан на рисунке 7.

```
>> rwalk

Mdl =

  arima with properties:

     Description: "ARIMA(0,1,0) Model (Gaussian Distribution)"
    Distribution: Name = "Gaussian"
               P: 1
               D: 1
               Q: 0
        Constant: 0
              AR: {}
             SAR: {}
              MA: {}
             SMA: {}
     Seasonality: 0
            Beta: [1×0]
        Variance: 1
```

Рисунок 5 - Выходные данные модели Arima для случайного блуждания

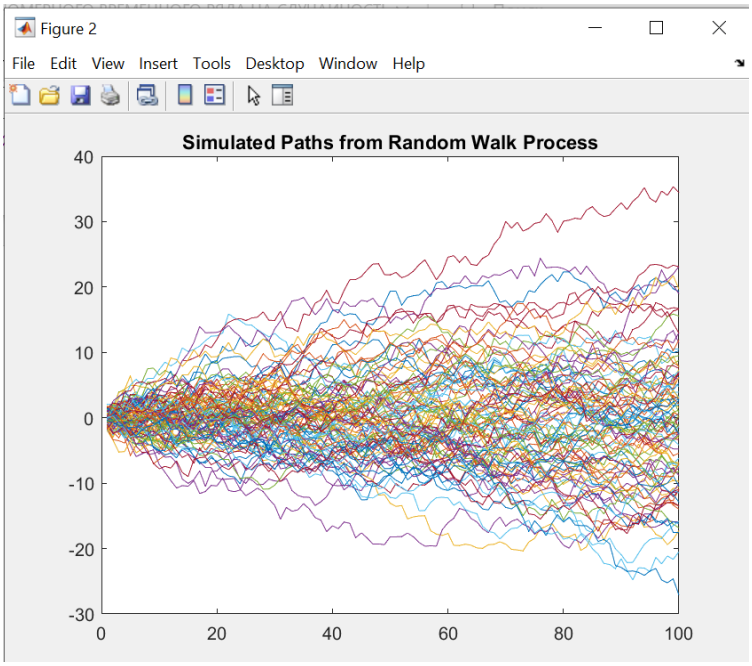

Рисунок 6 - Смоделированные пути из процесса случайного блуждания

```
2  Mdl.Constant = 0;
3  Mdl.Variance = 1;
4  Mdl
5  rng(1) % For reproducibility
6  Y = simulate(Mdl,100,'NumPaths',100);
7  plot(Y)
8  title('Simulated Paths from Random Walk Process')
```

Рисунок 7 - Фрагмент программного кода моделирования случайного блуждания

В работе также осуществлена проверка одномерного временного ряда с помощью теста Льюнг-Бокс (Q-тест). Смысл этого теста — статистический критерий, предназначенный для нахождения автокорреляции временных рядов. Вместо проведения тестирования на случайность каждого индивидуального коэффициента, данный метод осуществляет проверку на значимое отличие от нуля сразу нескольких коэффициентов автокорреляции (рисунок 8).

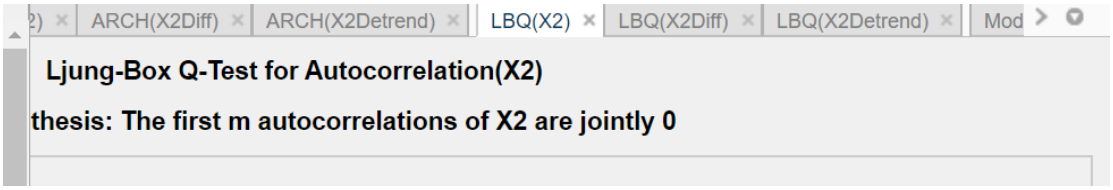

Рисунок 8 - Выходные данные анализа временного ряда с помощью Льюнг-Бокс Q-тест

Обсуждение.

В работе исследован одномерный временной ряд данных на случайность с целью дальнейшего анализа этих данных для прогнозирования сетевого трафика. На основе

полученных результатов, где исследованный ряд описывается как случайный процесс, проведение прогнозирования сетевого трафика может осуществляться с использованием методов прогнозирования временных рядов, присущих случайным процессам. В данном контексте, статистические методы, такие как модели ARIMA (авторегрессионная интегрированная скользящая средняя), GARCH (общий авторегрессионный условный дисперсионный процесс) или другие модели временных рядов, могут быть применены для анализа и прогнозирования поведения сетевого трафика.

Приведены результаты визуальной проверки, проверки нулевой гипотезы с помощью тестов о случайности распределений временного ряда и автокорреляции.

Получены такие наблюдения по тесту отношения дисперсии Variance Ratio test for Random Walk, по тесту Льюнга-Бокса, которые подтвердили, что одномерный временной ряд является случайным блужданием, по модели Arima для случайного блуждания, представленного в виде (1) формулы-уравнения. Прогнозирование сетевого трафика на основе случайных процессов может включать в себя оценку вероятностных распределений, изменчивости и корреляций в данных. Эти методы позволяют анализировать случайные флуктуации в трафике и предсказывать его будущее состояние с учетом случайных вариаций.

Заключение

Статистические характеристики стационарных и нестационарных временных рядов значительно различаются, и для их моделирования требуются различные подходы и методы. Модель стационарных временных рядов характеризуется постоянными во времени средней, дисперсией и автокорреляцией. Модель нестационарных временных рядов многокомпонентная, она содержит трендовую, сезонную и случайную составляющие.

Эта работа показала возможности методов, критериев и тестов определения случайной компоненты в ряду, наличие которой предполагает, что одномерный временной ряд данных является нестационарным. Для проверки случайности, кроме визуальной проверки и коррелограммы, использовались современные тесты в программной среде Матлаб. Все они подтвердили предположения о наличии случайной компоненты в исследуемом ряду. В дальнейшем, для прогнозирования сетевого трафика могут использоваться методы машинного обучения, такие как рекуррентные нейронные сети (RNN) или долгая краткосрочная память (LSTM), чтобы учесть сложные нелинейные зависимости в данных, которые могут быть обнаружены в случайных процессах сетевого трафика.

Важно учитывать, что выбор метода прогнозирования зависит от характера данных и конкретных особенностей сетевого трафика, и решение о применении определенного метода требует внимательного анализа полученных результатов и особенностей конкретной сети.

ЛИТЕРАТУРА

[1] Власюк А. А., Орлов Ю. Н. Точность идентификации выборочных распределений временных рядов в зависимости от типа распределения, нормы и длины выборки // Препринты ИПМ им. М.В. Келдыша. - 2015. - № 17. - С.25

[2] Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. - М.: «Статистика», 1973. –106 с.

[3] Г.У.Бектемысова, Ж.Б.Ибраева, А.Е.Кулакаева, Б.А.Кожахметова. Анализ измеренного сетевого трафика на стационарность. Вестник КазАТК, Вестник Казахской Академии Транспорта и Коммуникации им. М.Тынышпаева, ISSN 2790-5802, -№ 3 (122), -2022г. -С.: 302-308

[4] Л.В. Кирьянова. Теория случайных процессов: Курс лекций [Текст]: Л. В Кирьянова, А.Ю. Лемин, Т.А. Мацеевич. МОН РФ, Нац. исследоват. Моск. гос. строит. ун-т. – М.: Изд-во Моск. гос. строит. ун-та. - Мир, 2016. – 96 с.

[5] Дж.Бендат, Прикладной анализ случайных данных: Пер. с англ. [Текст]: Бендат, Дж. – М.: Мир, 1989. – 540 с.

[6] Sovetov, B. Y., Tatarnikova, T. M., Tsekhanovskiy, V. V. (2020) Avtoregressionnyye modeli prognozirovaniya setevogo trafika. Materialy konferentsii «Informatsionnyye tekhnologii v upravlenii». [Autoregressive models for predicting network traffic. Proceedings of the conference "Information technologies in management", 2020 St. Petersburg State University "LETI" them. IN AND. ULYANOVA (LENINA)]

[7] Beard E., Marsden J., Brown J., Tombor I., Stapleton J., Michie S., & West, R. Understanding and using time series analyses in addiction research // Society for the Study of Addiction. – 2019. – №114 (10). – P.1866-1884 DOI:10.1111/add.14643

REFERENCES*

[1] Vlasjuk A. A., Orlov Ju. N. Tochnost' identifikacii vyborochnyh raspredelenij vremennyh rjadov v zavisimosti ot tipa raspredelenija, normy i dliny vyborki // Preprinty IPM im. M.V. Keldysha. - 2015. - № 17. - S.25

[2] Kil'dishev G.S., Frenkel' A.A. Analiz vremennyh rjadov i prognozirovanie. - M.: «Statistika», 1973. –106 s.

[3] G.U.Bektemysova, Zh.B.Ibraeva, A.E.Kulakaeva, B.A.Kozhahmetova. Analiz izmerennogo setevogo trafika na stacionarnost'. Vestnik KazATK, Vestnik Kazahskoj Akademii Transporta i Kommunikacii im. M.Tynyshpaeva, ISSN 2790-5802, -№ 3 (122), -2022g. -S.: 302-308

[4] L.V. Kir'janova. Teorija sluchajnyh processov: Kurs lekcij [Tekst]: L. V Kir'janova, A.Ju. Lemin, T.A. Maceevich. MON RF, Nac. issledovat. Mosk. gos. stroit. un-t. – M.: Izd-vo Mosk. gos. stroit. un-ta. - Mir, 2016. – 96 s.

[5] Dzh.Bendat, Prikladnoj analiz sluchajnyh dannyh: Per. s angl. [Tekst]: Bendat, Dzh. – M.: Mir, 1989. – 540 s.

[6] Sovetov, B. Y., Tatarnikova, T. M., Tsekhanovskiy, V. V. (2020) Avtoregressionnyye modeli prognozirovaniya setevogo trafika. Materialy konferentsii «Informatsionnyye tekhnologii v upravlenii». [Autoregressive models for predicting network traffic. Proceedings of the conference "Information technologies in management", 2020 St. Petersburg State University "LETI" them. IN AND. ULYANOVA (LENINA)]

[7] Beard E., Marsden J., Brown J., Tombor I., Stapleton J., Michie S., & West, R. Understanding and using time series analyses in addiction research // Society for the Study of Addiction. – 2019. – №114 (10). – P.1866-1884 DOI:10.1111/add.14643

Жанар Ибраева, PhD, ассистент профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, zh.ibraeva@iitu.edu.kz

Тохтабуби Иманбекова, т.ғ.к., қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, tanya_nespayeva@mail.ru

Руслан Камал, магистр, аға оқытушы, Халықаралық акпараттық технологиялар университеті, Алматы, Қазақстан, R.kamal@iitu.edu.kz

БІР ӨЛШЕМДІ УАҚЫТТЫҚ ҚАТАРДА КЕЗДЕЙСОҚ ПРОЦЕСТІ АНЫҚТАУ

Аңдатпа. Бүгінгі күні адамның әртүрлі әрекеттері әсерінен ақпарат ағынымен байланысты деректер көлемінің астрономиялық өсуі байқалады. Жіберілетін ақпарат

көлемінің үнемі ұлғаюы оны өңдеуде деректерді негізгі желіде тасымалдау үшін белгілі бір күрделілік тудырады. Екінші жағынан, қазіргі қоғам өңделген ақпаратты тасымалдаудың жоғары жылдамдығын талап етеді.

Нақты өлшенген деректерді зерттеу олардың желілік құрылғыларға қызмет көрсететін пакеттердің біркелкі емес қарқындылығын көрсетеді. Сонымен қатар, бір объектіні уақыттың бірқатар дәйекті моменттері (уақыттық қатарлар) үшін сипаттайтын деректер негізінде құрастырылған модельдер стационарлық емес қасиетке ие, бұл олардың құрылымының көп компонентті екенін білдіреді. Күрделі құрылымның құрамдас бөліктерін анықтау және сандық анықтау – трендтің, кезеңділіктің, кездейсоқ компоненттің болуы/болмауы уақыттық қатарларды талдаудың негізгі міндеті болып табылады [1-3].

Бұл мақалада Алматы (Қазақстан) қаласындағы тас жолдан түсірілген нақты деректермен уақыт қатары талқыланады. Уақыттық қатарларды зерттеу үшін автокорреляция, дифференциация әдістері қолданылды, Matlab сандық және математикалық модельдеу ортасында қатарларды кездейсоқ процеске талдау үшін статистикалық сынақтар қолданылды.

Түйінді сөздер: бір өлшемді уақыт қатары, желілік трафик, деректерді талдау, автокорреляциялық функциялар, кездейсоқтықты тексеру критерийлері.

Zhanar Ibraeva, PhD, assistant professor, International Information Technology University, Almaty, Kazakhstan, zh.ibraeva@iitu.edu.kz

Tokhtabubi Imanbekova, c.t.s., associate professor, International University of Information Technology, Almaty, Kazakhstan, tanya_nespayeva@mail.ru

Ruslan Kamal, masters, senior lecturer, International Information Technology University, Almaty, Kazakhstan, R.kamal@iitu.edu.kz

DEFINITION OF A RANDOM PROCESS IN A ONE-DIMENSIONAL TIME SERIES

Abstract. To date, there has been an astronomical increase in the amount of data associated with information flows from various human activities. The ever-increasing volume of transmitted information creates a certain complexity for the backbone data transmission network in its processing. On the other hand, modern society requires high speeds of processed information transfer.

A study of measured data shows that they have an uneven intensity of packets arriving at serving network devices. At the same time, models built based on data characterizing one object for a number of successive moments of time (time series) have the property of non-stationarity, which means that their structure is multicomponent. The primary objective of time series analysis is to identify and quantify the components of a complex structure, including the presence or absence of a trend, periodicity, and a random component [1-3].

This article explores a time series based on real data collected from a main line in the city of Almaty (Kazakhstan). To study the time series, the methods of autocorrelation, differentiation were used, statistical tests were applied to analyze the series for a random process in the environment of numerical and mathematical modeling Matlab.

Keywords. One-dimensional time series, network traffic, data analysis, autocorrelation function, randomness testing criteria.

****