

УДК 004.716 : 303.446.33

СТАТИСТИЧЕСКАЯ ОЦЕНКА ОСНОВНЫХ ТЕНДЕНЦИЙ ИЗМЕРЕННОГО ВРЕМЕННОГО РЯДА

С.У. Исакова, Ш.А. Мирзакулова

Применение непараметрических критериев на основе статистических процедур позволяют формировать успешные выводы при исследовании временного ряда. Эти методы не требуют предположения о том, что исследуемый ряд распределен по нормальному закону распределения.

Ключевые слова: временной ряд; тренд; непараметрические критерии; критерий Уоллиса – Мура; знаковый критерий Кокса – Стюарта; кореллограмма.

ЧЕНЕЛГЕН УБАКЫТ КАТАРЛАРЫНЫН НЕГИЗГИ ТЕНДЕНЦИЯЛАРЫН СТАТИСТИКАЛЫК БААЛОО

Убакыт катарларын изилдөөдө статистикалык процедуралардын негизинде параметрдик эмес критерийлерди колдонуу ийгиликтүү жыйынтык чыгарууга мүмкүндүк берет. Бул методдор изилденүүчү катар кадимки бөлүштүрүү мыйзамы боюнча бөлүштүрүлгөндүгү тууралуу божомолду талап кылбайт.

Түйүндүү сөздөр: убакыт катары; тренд; параметрдик эмес критерийлер; Уоллис–Мурдун критерийи; Кокс–Стюарттын белги критерийи; кореллограмма.

STATISTICAL ESTIMATION OF THE MAIN TRENDS OF THE MEASURED TIME SERIES»

S.U. Isakova, S.A. Mirzakulova

The use of non-parametric criteria based on statistical procedures, allows us to form successful conclusions in the study of the time series. These methods do not require the assumption that the series under study is distributed according to the normal distribution law.

Keywords: time series; trend; non-parametric criteria; Wallis – Moore criterion; Cox – Stewart sign criterion; corellogramma

Применение различных статистических тестов или критериев при исследовании временных рядов позволяет с некоторой точностью отнести их к тому или иному классу для описания математической модели. В данной работе исследован ряд на наличие определенного тренда с целью дальнейшего использования его данных для прогнозирования сетевого трафика.

Практика показывает, что большинство реальных процессов не удовлетворяет свойствам стационарности. Нестационарные ряды характеризуются трендом, систематическими изменениями дисперсии, наличием периодических флуктуаций и наличием изменяющихся взаимозависимостей между уровнями временного ряда. Для распознавания не

стационарности временного ряда могут быть использованы различные методы: визуальный анализ графического представления временного ряда на наличие тенденции и периодической составляющей, метод средних, анализ временного ряда на наличие автокорреляции и др. В статье осуществлены упрощенные проверки измеренного временного ряда на нестационарность. Используются различные критерии проверки на нормальность распределения. Применены непараметрические методы: фазо-частотный критерий Уоллиса – Мура, который отличает отклонения ряда измерений от чисто случайной последовательности и знаковый критерий Кокса – Стюарта для проверки гипотезы о наличии тренда. Построена кореллограмма, имеющая много значимых

коэффициентов автокорреляции, уровни которых превышают уровни доверительных интервалов.

Измеренный ряд динамики приведен на рисунке 1, который показывает совокупность переданных по магистральной сети пакетов за пять часов за каждые 10 с.

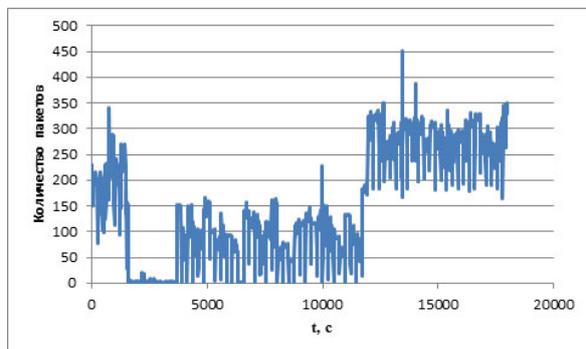


Рисунок 1 – Ряд интенсивности передачи пакетов

Для проверки гипотезы о стационарности ряда в [1] показано, что используются тесты, которые разделяются на два вида – параметрические и непараметрические. Применение параметрических тестов возможно только в случае нормального распределения данных.

Проведем исследование закона распределения временного ряда – построим его гистограмму (рисунок 2).

Полученная гистограмма не похожа на колокол, а также по данным статистических показателей видно, что они распределены не по нормальному закону, например куртозис равен значению – 1,327121 вместо значения, равного трем в случае нормального распределения данных.

Согласно критерию Колмогорова – Смирнова, расхождения между распределениями (теоретического и эмпирического) основано на максимальном значении абсолютной величины разности между этими функциями распределения. Если D статистика Колмогорова – Смирнова значима, то гипотеза о нормальном распределении отвергается. Критическое значение D статистики при объеме выборки ряда, равного значению 1800 ($n > 35$) при значении $p = 0,01$, определяется следующим выражением [2]:

$$D_{крит} = \frac{1,63}{\sqrt{1800}} = 0,038. \quad (1)$$

Так как $D_{эмп} = 0,11873$ и больше $D_{крит} = 0,038$, то гипотеза о достоверных различиях принимается.

На рисунке 3 приведены результаты обработки программой Attestat исходного временного ряда по его проверке на нормальность распределения на основе критериев: Колмогорова, Смирнова, Крамера – Мизеса, Андерсона – Дарлинга, Шапиро – Франсиа, Жарка – Бера, Гири и др.

Так как закон распределения исследуемого временного ряда отличается от нормального, для проверки гипотезы о стационарности ряда воспользуемся непараметрическими тестами.

Осуществим проверку на наличие тренда в ряду динамики двумя непараметрическими критериями, которые измеряют связи между явлениями, при этом не используют количественные значения признака, то есть не используют параметры распределения:

- фазо-частотного критерия Уоллиса – Мура;
- знакового критерия тренда Кокса – Стюарта.

Фазо-частотный критерий Уоллиса – Мура не зависит от вида распределения временных меток x_i , отличает отклонения ряда измерений ($n > 12$) от чисто случайной последовательности.

При этом исследуемый временной ряд будет случайным, если знаки разностей $(x_{i+1} - x_i)$ распределены случайно (нулевая гипотеза), а альтернативная гипотеза – значимо отличается от случайной последовательности – последовательности плюсов и минусов. Критерий Уоллиса – Мура рассматривает частоту плюсовых и минусовых фаз – одинаковых последовательностей знаков и не учитывает разности $x_{i+1} - x_i = 0$, а также начальную и конечную фазы при вычислении числа фаз L . Расчетное значение фазо-частотного критерия разностей определяется по формуле

$$\hat{z} = \frac{\left| L - \frac{2 \cdot n - 7}{3} \right| - 0,5}{\sqrt{\frac{16 \cdot n - 29}{90}}}, \quad (2)$$

где \hat{z} – фазо-частотный критерий разностей; L – число фаз; n – число уровней.

$$\hat{z} = \frac{\left| 1041 - \frac{2 \cdot 1800 - 7}{3} \right| - 0,5}{\sqrt{\frac{16 \cdot 1800 - 29}{90}}} = \frac{156,17}{17,88} = 8,73. \quad (3)$$

Так как $\hat{z} = 8,73 > z_{0,05} = 1,96 (\alpha = 5\%)$, где α – уровень значимости, нулевая гипотеза отвергается, то есть исследуемый ряд имеет возрастающий тренд. Значения $Y_{0,05}(n)$ табулированы и приведены в [3].

Метод «Знакового критерия тренда Кокса – Стюарта» требует деления исследуемого ряда,

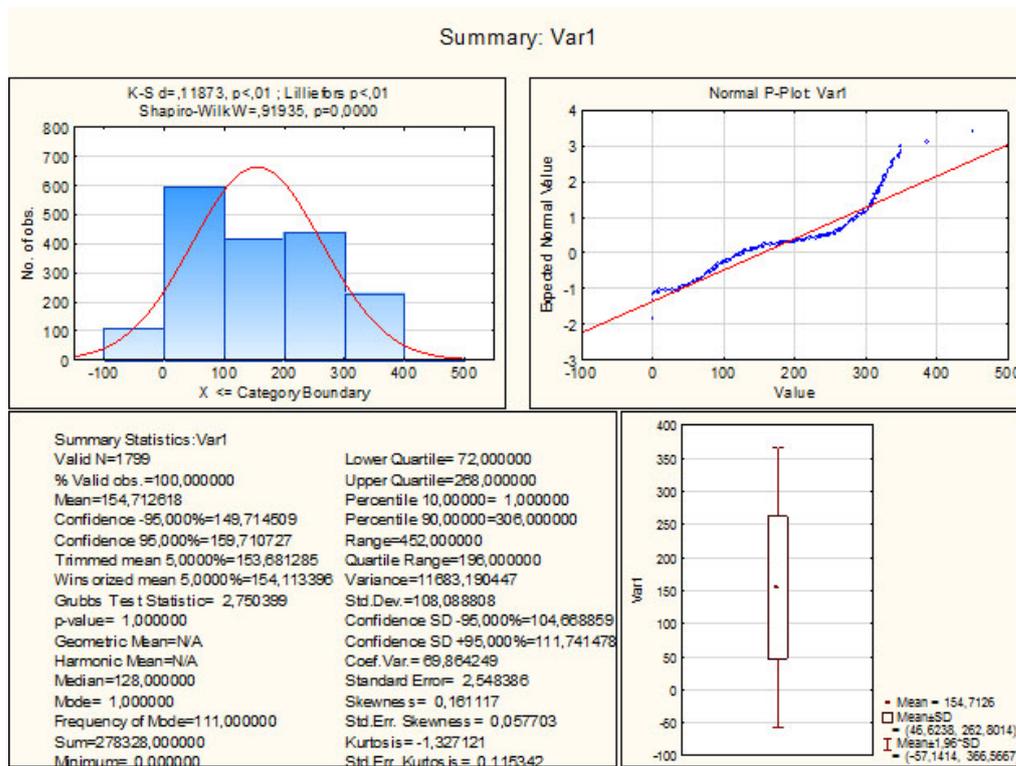


Рисунок 2 – Гистограмма распределения исходного ряда

	A	B	C	D	E	F	G	H	I	J
8	4379				Модифицированный критерий Колмогорова					
9	1920				0,168366	1,1E-05	Гипотеза о нормальности отклоняется			
10	24				Модифицированный критерий Смирнова					
11	19				0,159908	2,05E-05	Гипотеза о нормальности отклоняется			
12	27				Критерий Крамера-Мизеса					
13	64				0,729559	2,89E-07	Гипотеза о нормальности отклоняется			
14	21				Критерий Андерсона-Дарлинга					
15	13				4,253452	7,66E-07	Гипотеза о нормальности отклоняется			
16	22				Критерий Шапиро-Франка					
17	12				0,899118	1,01E-05	Гипотеза о нормальности отклоняется			
18	14				Критерий коэффициента асимметрии					
19	12				0,144743	0,274371	Гипотеза о нормальности не отклоняется			
20	12				Критерий эксцесса					
21	1570				-1,45413	0,001183	Гипотеза о нормальности отклоняется			
22	1935				Критерий Жарка-Бера					
23	1274				9,00079	0,011105	Гипотеза о нормальности отклоняется			
24	2158				Критерий Колмогорова					
25	1601				Заданное среднее значение					
26	1340				0					
27	1267				Заданное стандартное отклонение					
28	2656				1					
29	1978				1	0	Гипотеза о нормальности отклоняется			
30	1593				Критерий Гири					
31	1410				0,910526	5,61E-08	Гипотеза о нормальности отклоняется			

Рисунок 3 – Проверка нормальности распределения

состоящего из n значений, на три равные группы по 600 значений в каждом ($n/3$), как в нашем случае. При этом мы имеем первую x_i^1 , вторую x_i^2 и третью x_i^3 группы, но рассматриваем первую и третью группы (x_i^1, x_i^3) с составлением разности значений $x_i^3 - x_i^1$. В конечном счете, на основе этих данных фиксируется знак «плюс» при возрастающем тренде (s^+), знак «минус» – при убывающем тренде (s^-), нуль – при отсутствии тренда (s^0), так что $s^+ + s^- + s^0 = n/3$.

$$\hat{z} = \frac{\left| S - \frac{n}{6} \right| - 0,5}{\sqrt{\frac{n}{12}}}, \quad (4)$$

где $S = \max \{s^+, s^-\}$ приближенно распределена нормально.

Для расчета величины \hat{z} используем данные, полученные при исследовании: $s^+ = 574$, $s^- = 26$, $s^0 = 0$, $S = \max \{574, 26\} = 574$.

$$\hat{z} = \frac{\left| 574 - \frac{600}{6} \right| - 0,5}{\sqrt{\frac{600}{12}}} = \frac{473,5}{7,07} = 66,97. \quad (5)$$

Если $\hat{z} = 66,97 > z_{0,05} = 1,96$ ($\alpha = 5\%$), нуль-гипотеза отменяется, временной ряд имеет тренд. Для обнаружения связей, лежащих в основе этих зависимостей, воспользуемся статистическим методом (корреляцией), который определяет существование зависимости и насколько она сильна.

Линейную зависимость между двумя переменными описывает парная корреляция. Частная корреляция – в ней исключены влияния других переменных. На рисунке 4 показана кореллограмма ряда, при этом ряд 1 – кореллограмма, ряд 2 и ряд 3 – доверительные интервалы соответственно нижний и верхний. При этом видно, что кореллограмма имеет много значимых коэффициентов автокорреляции, уровни которых превышают уровни доверительных интервалов, наиболее высокий уровень автокорреляции при лаге, равном 1, следовательно,

ряд содержит тенденцию, и уровни коэффициентов автокорреляции убывают медленно и с ростом лага кореллограмма не стремится к нулю.



Рисунок 4 – Кореллограмма исследуемого ряда

Таким образом, статистическая оценка нормальности распределения исследуемого ряда показала расхождение между измеренной и теоретической распределениями согласно критерия Колмогорова – Смирнова, $D_{\text{эксп}}$ оказалось больше $D_{\text{крит}}$. Применение непараметрических тестов (Уоллиса – Мура и Кокса – Стюарта) к исследуемому ряду показали наличие в этом ряду тренда.

Полученная кореллограмма исследуемого ряда показала, что ряд содержит много значимых коэффициентов автокорреляции, уровни которых превышают уровни доверительных интервалов, что свидетельствует о наличии в ряде тенденции. Коэффициенты автокорреляции убывают медленно, и с ростом лага кореллограмма не стремится к нулю. Наличие тренда (тенденции) в ряде свидетельствует о нестационарности ряда.

Литература

1. Проверка исходного ряда на стационарность URL: <http://www.legalmanager.ru/lems-835-1.html>
2. Лемешко Б.Ю. Критерии проверки отклонения распределения от нормального закона: руководство по применению / Б.Ю. Лемешко. Новосибирск: НГТУ, 2014. 192 с.
3. Проверка гипотезы о существовании тенденций. URL: <http://vremennoi-ry-ru.1gb.ru/3.html>