АСТРОНОМИЯ, АСТРОФИЗИКА И КОСМОЛОГИЯ

Использование непараметрических методов математической статистики для поиска космических струн

А. В. Моргунова,[1, а] О. С. Сажина[1, 2, б]

[1] *Московский государственный университет имени М. В. Ломоносова, физический факультет, кафедра небесной механики, астрометрии и гравиметрии. Россия, 119991, Москва, Ленинские горы, д. 1, стр. 2.*

[2] *Государственный астрономический институт имени П. К. Штернберга (ГАИШ МГУ). Россия, 119991, Москва, Университетский пр-т, д. 13.*

Поступила в редакцию 06.04.2019, после доработки 24.04.2019, принята к публикации 06.05.2019.

Проводилось статистическое исследование участка небесной сферы, который по радиоданным анизотропии реликтового излучения содержит кандидата в космическую струну (поле CSc-1), с целью дополнительного обоснования этого кандидата. Космические струны — одномерные объекты космологических масштабов, которые могут проявлять себя, в частности, как гравитационные линзы, обладающие характерными особенностями. Решена задача статистического сравнения плотности распределения гравитационно-линзовых пар источников (галактик) в полях, не содержащих космических струн, с распределением аналогичных пар в поле CSc-1. Использовались непараметрические ранговые критерии сдвига, позволяющие работать с малыми выборками при неизвестном законе распределения; обсуждались пределы применимости критериев. Выявленное различие двух указанных распределений служит дополнительным аргументом в пользу наличия в поле CSc-1 струны. Обнаружение космических струн послужило бы уникальным наблюдательным тестом современных космологических теорий, в том числе многомерных моделей.

Ключевые слова: космические струны, гравитационное линзирование, статистическая обработка данных.
УДК: 524.8, 520.8, 519.254, 519.234.
PACS: 02.60.-x, 95.75.Mn, 98.70.Vc, 98.80.-k, 95.35.+d, 95.36.+x.

ВВЕДЕНИЕ

Космические струны (КС) — предсказываемые широким классом физических теорий одномерные объекты космологических масштабов с энергиями порядка шкалы Теории великого объединения (10^{16} ГэВ) [1–3]. Теоретические механизмы рождения КС разнообразны [4, 5]. Если КС образовывались в результате серии фазовых переходов при расширении Вселенной, то они принадлежат к топологическим дефектам и либо образуют замкнутые петли, либо не имеют концов. При наличии дополнительных полей имеют место гибридные модели КС (т. н. полулокальные КС) [6, 7]. КС могут оказаться результатом взаимодействия в ранней доинфляционной Вселенной многомерных структур и, дожив до сегодняшнего момента, послужить первым наблюдательным доказательством теории суперструн [8]. Обнаружение КС послужило бы уникальным наблюдательным тестом современных космологических теорий.

КС обладают рядом хорошо описанных астрофизических свойств, поскольку пространство в присутствии КС становится глобально трехмерно-коническим, где раскрыв конуса — малая величина, называемая дефицитом угла КС и определяемая энергией КС. Обладая релятивистскими скоростями, КС должны проявлять себя посредством эффекта Доплера в реликтовом излучении (РИ) [9–11]. Кроме того, КС должны порождать характерные цепочки гравитационно-линзовых (ГЛ) событий [12–14]. ГЛ-события на КС обладают локализованным вдоль линии избыточным распределением, что отличает их от обычных ГЛ-событий [15, 16]. Расстояние между компонентами ГЛ-пар пропорциональны дефициту угла КС (так, энергиям порядка 10^{16} ГэВ соответствует расстояние порядка нескольких угловых секунд). В отличие от типичных внегалактических объектов КС одномерна и обладает протяженностью космологических масштабов, больше углового масштаба галактик, вплоть до размеров видимого горизонта Вселенной. Таким образом, на всей протяженности КС ожидается формирование цепочек ГЛ-пар фоновых по отношению к струне галактик, т. н. «Млечный путь гравитационных линз», количество которых избыточно по сравнению с числом обычных ГЛ-событий [14] и плотность распределения отлична от последних.

Для выбора области поиска таких цепочек ГЛ-пар использовались данные обработки карт РИ, в которых был найден КС-кандидат (CSc-1; протяженностью от $(\alpha = 11 : 29 : 03, \delta = +15 : 23 : 37)$ до $(\alpha = 10 : 57 : 47, \delta = +25 : 03 : 51)$) [7, 17, 18]. КС создает характерный скачок анизотропии РИ, [9], который может быть выявлен путем применения ступенчатого фильтра, основанного, например, на модифицированных функциях Хаара [19].

В данной работе представлены результаты статистического сравнения распределения ГЛ-пар в полях, не содержащих КС (контрольные поля), с распределением аналогичных пар в поле CSc-1. Были выбраны непараметрические ранговые критерии сдвига,

[а] E-mail: arina.morgunova@yandex.ru
[б] E-mail: cosmologia@yandex.ru

позволяющие работать с малыми выборками, законы распределения которых неизвестны: (1) быстрый ранговый критерий, (2) критерий Ван дер Вардена, (3) критерий Манна—Уитни—Вилкоксона, (4) аппроксимация Имана. Обсуждалась правильность работы указанных критериев на тестовых синтезированных выборках. Была доказана неоднородность двух наборов данных, что говорит о статистическом различии распределений ГЛ-пар в двух указанных наборах полей, что в совокупности с аномально высоким количеством ГЛ-пар в поле CSc-1 [18] дает дополнительные аргументы в пользу существования КС в поле CSc-1.

1. ИСХОДНЫЕ ДАННЫЕ И ФОРМИРОВАНИЕ ВЫБОРОК ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА

ГЛ-пары были найдены по фотометрическому каталогу галактик DR12 SDSS. При поиске были учтены необходимые геометрические и фотометрические условия того, что галактики представляют собой ГЛ-пары (равенство фотометрических красных смещений, одинаковое отношение интенсивностей на всех доступных частотах, характерные расстояния между компонентами ГЛ-пары, от $1.3''$ до $9''$; первая величина следует из разрешения каталога, а последняя — из теоретических предсказаний максимально возможной энергии КС [18]). Отметим, что в данном исследовании речь идет только о кандидатах в ГЛ-пары, поскольку более сильным необходимым (но недостаточным) условием является идентичность спектров компонент ГЛ- пары. Кроме того, в случае линзирования галактики на КС ожидались бы характерные срезы внешних яркостных изофот и неискаженность изображений, что можно выявить либо на инструментах высокого углового разрешения, либо с помощью подходящего фильтрационного анализа. Данные оптического каталога DR12 SDSS позволяют выявлять только ГЛ-кандидаты.

Были сформированы две выборки. Каждый элемент первой выборки (далее CSc-1 выборка) — это количесто пар галактик, расположенных в площадке 1 кв. град. в поле CSc-1. Каждый элемент второй выборки (далее — контрольная выборка) — это количество пар галактик, расположенных в площадке 1 кв. град. в поле, где, по данным анизотропии РИ, ступенчатый фильтр не обнаружил КС [18]), см. табл. 1 (дробные значения количества пар происходят из-за указанной нормировки на 1 кв. град.). В табл. 2 [18] указаны координаты центров контрольных полей, которые не содержат кандидатов в КС, найденных по данным анизотропии РИ, [7] и которые расположены вдали от галактической экваториальной области, где данные анизотропии могут быть искажены вкладом излучения Галактики.

Цель работы — показать, что вне зависимости от способа группировки данных этих выборок и применяемых методов, плотности распределения данных двух выборок различны и, таким образом, CSc-1-выборка представляется аномальной по сравнению с контрольной выборкой.

Пример группировки: в поле CSc-1 число площадок, в которых количество ГЛ-пар лежит в интервале $(9, 11)$, равно 2, а для контрольного поля число ГЛ-пар в этом интервале равно 3. Эти числа можно рассматривать как случайные величины, характеризующие частоту ГЛ-пар, и их распределения использовать для доказательства неоднородности двух наборов данных.

Табл. 1. Исходные данные

N	Поле CSc-1	Контрольное поле	N	Поле CSc-1	Контрольное поле
1	52.61	20.06	17	23.23	11.01
2	32.92	16.05	18	27.53	14.44
3	37.38	23.10	19	27.64	16.87
4	26.42	30.09	20	14.91	32.72
5	26.48	27.07	21	21.37	13.98
6	24.31	32.23	22	32.19	23.35
7	13.23	31.13	23	30.11	28.89
8	25.09	20.08	24	17.38	14.23
9	30.21	15.05	25	10.87	11.74
10	10.73	10.75	26	26.21	25.88
11	30.13	19.82	27	15.37	18.19
12	36.71	23.22	28	26.50	23.61
13	38.95	30.52	29	51.08	22.85
14	30.41	12.68	30	17.05	
15	21.06	25.64	31	17.92	
16	29.49	18.85			

Табл. 2. Координаты центров полей для контрольной выборки (экваториальные координаты, к. э. 2000 г.), [18]

N	RA $[h:m:s]$	Dec $[^\circ:':'']$
1	10 : 56 : 16.000	+15 : 47 : 36.90
2	11 : 06 : 50.510	+13 : 25 : 38.70
3	11 : 43 : 50.900	+27 : 39 : 42.80
4	12 : 36 : 26.070	+07 : 26 : 50.60
5	12 : 43 : 16.160	+07 : 13 : 20.80
6	16 : 48 : 00.000	+51 : 00 : 00.00
7	14 : 12 : 00.000	+34 : 00 : 00.00
8	00 : 36 : 00.000	+39 : 00 : 00.00
9	08 : 00 : 00.000	+20 : 00 : 00.00
10	10 : 22 : 20.390	+55 : 27 : 36.82
11	10 : 46 : 52.110	+56 : 48 : 06.03
12	12 : 20 : 13.800	+62 : 45 : 32.82
13	13 : 13 : 08.520	+56 : 52 : 39.47
14	14 : 41 : 14.410	+57 : 29 : 13.74
15	14 : 55 : 23.690	+43 : 52 : 18.17
16	16 : 05 : 57.170	+40 : 55 : 21.36
17	08 : 48 : 58.430	+50 : 50 : 44.99
18	16 : 43 : 19.080	+42 : 29 : 10.73
19	16 : 33 : 32.860	+32 : 08 : 41.12
20	10 : 52 : 24.800	+54 : 37 : 02.21
21	12 : 14 : 41.820	+60 : 32 : 56.74
22	13 : 04 : 18.970	+55 : 01 : 47.58
23	14 : 43 : 03.700	+43 : 53 : 05.77
24	15 : 54 : 41.140	+42 : 02 : 56.69
25	09 : 06 : 46.280	+49 : 04 : 41.78
26	09 : 04 : 13.830	+53 : 39 : 49.67
27	09 : 48 : 07.200	+61 : 50 : 36.77
28	16 : 32 : 17.970	+43 : 57 : 46.62
29	16 : 25 : 03.380	+33 : 52 : 05.21

Табл. 3. Данные, перегруппированные по интервалам плотности числа ГЛ-пар в CSc-1 и контрольном поле

Первая перегруппировка			Вторая перегруппировка		
Интервал	Поле CSc-1 ($n_1 = 24$)	Контрольное поле ($n_2 = 13$)	Интервал	Поле CSc-1 ($m_1 = 12$)	Контрольное поле ($m_2 = 7$)
9	0	0	11	2	3
11	2	3	13	3	5
13	1	2	17	3	5
15	2	3	21	3	7
17	3	2	25	8	3
19	0	3	29	5	4
21	2	2	33	2	2
23	1	5	37	3	
25	2	2	41	0	
27	6	1	45	0	
29	1	1	49	1	
31	4	3	53	1	
33	2	2			
35	0				
37	2				
39	1				
41	0				
43	0				
45	0				
47	0				
49	0				
51	1				
53	1				
55	0				

Разбиение на интервалы велось двумя способами: CSc-1 и контрольная выборки содержали (1) 24 и 13 и (2) 12 и 7 величин соответственно (табл. 3).

2. ОБОСНОВАНИЕ ИСПОЛЬЗОВАНИЯ НЕПАРАМЕТРИЧЕСКИХ РАНГОВЫХ КРИТЕРИЕВ

Трудность анализа состоит в малости выборок и неизвестности законов распределений. Следовательно, применение стандартных параметрических критериев (Стьюдента, Фишера и др.) может привести к неудовлетворительным выводам. В данном случае целесообразно использовать непараметрические или свободные от распределения критерии однородности статистических данных, статистически проверив их применимость для случая малых выборок.

Для доказательства того, что предположение о совпадении двух распределений случайных величин неверно, следует опровергнуть хотя бы одну из гипотез о равенстве параметров масштаба или положения (сдвига).

Сдвиг определяется разностью параметров положения, которые характеризуют центры группировки случайных величин в исследуемых распределениях. Заметим, что к критериям масштаба необходимо будет обратиться только в том случае, если гипотеза о наличии сдвига будет отклонена.

Используемые критерии тестируются на искусственно сгенерированных выборках (табл. 4), распределения случайных величин в которых заведомо совпадают. Это позволяет удостовериться в адекватности применяемых методов и, следовательно, в достоверности сделанных на их основе выводов.

Ранговые критерии одни из самых эффективных методов непараметрической статистики (эффективность лучших из них составляет до 95% от мощности t-критерия Стъюдента и сопоставима с последним для случая больших выборок, [20]; далее эффективность везде указывается относительно t-критерия). Они основываются на использовании рангов, приписываемых значениям случайных величин в общей упорядоченной по возрастанию выборке (т. е. в упорядоченном ряду чисел $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ значению x_i приписывается ранг R_i). При этом одинаковым величинам присваивается усредненный ранг (табл. 1).

Ранговых критериев сдвига существует много, но есть три условия, которые существенно ограничивают нас в выборе: наличие повторяющихся значений в наборах случайных величин, неравенство объемов выборок и малый объем выборок. Учитывая вышеперечисленные особенности, можно выделить только три критерия, применение которых в рамках поставленной задачи будет обоснованным: быстрый ранговый критерий, критерий Ван дер Вардена и критерий Манна—Уитни—Вилкоксона, частным случаем которого является аппроксимация Имана.

3. БЫСТРЫЙ РАНГОВЫЙ КРИТЕРИЙ

Быстрый (грубый) ранговый критерий [20] — самый простой с точки зрения вычислительной

Табл. 4. Однородные выборки

$n_1 = 31$	$m_1 = 29$	$n_2 = 24$	$m_2 = 13$	$n_3 = 12$	$m_3 = 7$
27.97	24.34	2	5	1	6
27.71	16.71	6	1	6	6
11.37	13.55	2	2	6	5
11.64	24.4	0	4	4	5
23.29	10.46	5	6	1	1
11.56	24.96	5	3	6	2
23.27	29.43	4	2	5	
12.14	21.19	5	6	3	
27.02	10.05	0	1	6	
13.19	17.49	3	4	2	
26.14	15.55	4	3	5	
15.10	28.85	3	4		
10.44	23.89	1			
16.89	13.82	6			
15.21	26.24	2			
27.83	21.86	3			
28.9 0	29.48	2			
12.33	20.70	3			
24.95	25.91	5			
14.05	28.15	5			
15.97	9.58	4			
23.65	19.46	6			
9.53	27.38	1			
29.91	28.66				
26.87	18.33				
20.07	19.43				
28.27	25.50				
11.59	26.99				
24.16					
15.57					

Табл. 5. Присвоение рангов элементам исследуемых выборок

Первая группа данных ($\mathbf{n_1 = 24}$, $n_2 = 13$)								
N	Частота	Ранг	N	Частота	Ранг	N	Частота	Ранг
1	**0**	5.5	14	**1**	14.5	27	2	24
2	**0**	5.5	15	**1**	14.5	28	2	24
3	**0**	5.5	16	**1**	14.5	29	2	24
4	**0**	5.5	17	1	14.5	30	**3**	32
5	**0**	5.5	18	1	14.5	31	3	32
6	**0**	5.5	19	**2**	24	32	3	32
7	**0**	5.5	20	**2**	24	33	3	32
8	**0**	5.5	21	**2**	24	34	3	32
9	**0**	5.5	22	**2**	24	35	**4**	35
10	0	5.5	23	**2**	24	36	5	36
11	**1**	14.5	24	**2**	24	37	**6**	37
12	**1**	14.5	25	2	24			
13	**1**	14.5	26	2	24			
Вторая группа данных ($\mathbf{m_1 = 12}$, $m_2 = 7$)								
N	Частота	Ранг	N	Частота	Ранг	N	Частота	Ранг
1	**0**	1.5	8	**3**	10.5	15	**5**	16
2	**0**	1.5	9	**3**	10.5	16	5	16
3	**1**	3.5	10	**3**	10.5	17	5	16
4	**1**	3.5	11	**3**	10.5	18	7	18
5	**2**	6	12	3	10.5	19	**8**	19
6	**2**	6	13	3	10.5			
7	2	6	14	4	14			

сложности и обладающий эффективностью 86%. Он основан на переходе к статистике d-критерия, которая может быть аппроксимирована нормальным распределением с нулевым средним. Переход к d-статистике:

- составление общего ранжированного ряда из двух выборок объемами n_1 и n_2;
- присвоение рангов;
- рассчет суммарных усредненных рангов для каждой из групп, то есть $\overline{R_1} = \frac{1}{n_1}\sum R_i$ и $\overline{R_2} = \frac{1}{n_2}\sum R_j$.

Тогда d-критерий имеет вид $d = \overline{R_1} - \overline{R_2}$ с дисперсией:

$$s_d = \sqrt{\frac{(n_1+n_2)(n_1+n_2+1)}{12}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

Проверка гипотезы о неоднородности двух наборов статистических данных осуществляется путем сравнения полученной величины с квантилью стандартного нормального распределения $u_{\frac{1+\alpha}{2}}$. Если $|d/s_d| < 1.96$, то гипотеза сдвига отклоняется с доверительной вероятностью $\alpha = 0.95$. Для данных табл. 3 (здесь и далее с точностью до второго знака):

- $n_1 = 24$, $n_2 = 13$: $\overline{R_{n_1}} = 16.02$, $\overline{R_{n_2}} = 24.5$, $d = 8.48$, $s_d = 3.73$, $|d/s_d| = 2.27 > 1.96$;
- $m_1 = 12$, $m_2 = 7$: $\overline{R_{m_1}} = 8.25$, $\overline{R_{m_2}} = 13$, $d' = 4.75$, $s'_d = 2.25$, $|d'/s'_d| = 1.77 < 1.96$.

Гипотеза наличия сдвига, полученная быстрым ранговым критерием, подтверждается только для большой выборки, поэтому требуются дополнительные исследования другими более эффективными методами.

4. КРИТЕРИЙ ВАН ДЕР ВАРДЕНА

Критерий Ван—дер—Вардена [21] основан на алгоритме перехода к статистике X-критерия и различен для выборок средних и малых объемов. Для второй группы данных (табл. 1) суммарным объемом $m_1 + m_2 = 19$ статистика Ванн—дер—Вардена имеет вид

$$X_m = \sum_{j=1}^{m_2} u_{\gamma_j},$$

где u_{γ_j} есть γ_j-квантиль стандартного нормального распределения (суммирование можно вести относительно m_1, результат не меняется). Величину u_γ можно определить по приближенной формуле

$$u_{\gamma_j} \approx 4.91\left[\gamma_j^{0.14} - (1-\gamma_j)^{0.14}\right],$$

где $\gamma_j = R_l/(m_1 + m_2 + 1)$, $j = \overline{1, m_2}$.

Для второй группы данных: $\gamma_1 = 0.3$, $u_{\gamma_1} = -0.52$, $\gamma_2 = 0.53$, $u_{\gamma_2} = 0.06$, $\gamma_3 = 0.7$, $u_{\gamma_3} = 0.52$, $\gamma_4 = 0.8$, $u_{\gamma_4} = 0.84$, $\gamma_5 = 0.9$, $u_{\gamma_5} = 1.28$, $X_m = -0.52 + 0.06 \cdot 2 + 0.52 + 0.84 \cdot 2 + 1.28 = 3.02$, что меньше критического значения 3.62 ($\alpha = 0.95$ и $m_1 = 12$, $m_2 = 7$ [20]).

Для первой группы данных (табл. 5) $n = n_1 + n_2 = 37$ и статистику Ван—дер—Вардена можно аппроксимировать нормальным распределением со средним $m = 0$ и дисперсией

$$D(X) = \sqrt{\frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)} \sum_{p=1}^{n_1+n_2} u^2_{\gamma_p}},$$

где $\gamma_p = R_p/(n_1+n_2+1)$, $p = \overline{1, n}$.

Соответствующие расчеты дают: $X_n = 5.64$, $D(X) = 2.49$. Как и в случае быстрого рангового критерия, $|X_n/D(X)|$ сравнивается с $u_{0.95}$: $|X_n/D(X)| = 2.26 > 1.96$.

Для первой группы данных гипотеза сдвига справделива. Однако по отношению к малым выборкам требуется дополнительное исследование, поскольку метод обладает высокой эффективностью (его мощность равна мощности t-критерия Стьюдента) только для больших выборок.

5. КРИТЕРИЙ МАННА—УИТНИ—ВИЛКОКСОНА

Критерий Манна-Уитни-Вилкоксона [22], основан на U-статистике Манна—Уитни и R-статистике Вилкоксона. Его эффективность 95%. U-статистика определяется как

$$U = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} h_{ij}, \quad \text{где} \quad h_{ij} = \begin{cases} 1, & \text{если } x_i < y_j, \\ 0, & \text{если } x_i > y_j. \end{cases}$$

Здесь k_1, k_2 — объемы двух выборок в рассматриваемой группе; $i = \overline{1, k_1}$, $j = \overline{1, k_2}$. Для вычисления U необходимо подсчитать количество элементов первой выборки, не превосходящих по своему значению случайные величины из второй выборки. При этом не важно, относительно какой из выборок ведется суммирование. Мы сопоставили величины, соответствующие CSc-1 и контрольному полю, и получили следующие значения U для первой и второй перегруппировок данных: $U_n = 57$, $U_m = 15$.

В случае малых выборок гипотеза сдвига отклоняется, если найденная величина U не входит в числовой интервал, определяемый критическими значениями U_1 и U_2. Для $m_1 = 12$ и $m_2 = 7$ имеем $U_1 = 18$ и $U_2 = 63$ [20], следовательно, $U = 15 \notin [18, 63]$.

Для выборки большего объема лучшую оценку дает R-статистика:

$$R = n_1 n_2 + \frac{n_2(n_2+1)}{2} - U,$$

которая аппроксимируется W-распределением:

$$W = \frac{R - n_2(n_1+n_2+1)/2}{g}.$$

Величина g в знаменателе:

$$g = \left[\frac{n_1 n_2 (n_1+n_2+1)}{12} \times \right.$$

$$\left. \times \left(1 - \frac{\sum_{s=1}^{q} t_s(t_s^2 - 1)}{(n_1+n_2)(n_1+n_2+1)(n_1+n_2-1)} \right) \right]^{0.5}, \quad (1)$$

где s — количество групп, значения элементов в которых одинаковы, t_s — количество элементов в каждой группе, q — общее число групп. При этом элементы, численно равные друг другу, но принадлежащие разным выборкам, не учитываются. Величина g учитывает совпадающие элементы в выборках, благодаря чему полученное распределение может быть аппроксимировано нормальным. Для $n_1 = 24$ и $n_2 = 13$ получаем $R = 346$. Величина $q = 4$, следовательно, $t_1(0) = 10$, $t_2(1) = 8$, $t_3(2) = 11$, $t_4(3) = 5$. (В скобках при t_s указаны численные значения совпадающих элементов в наборе данных). Тогда $g = 30.51$ и $|W| = 3.25 > 1.96$ (для $\alpha = 0.95$).

Таким образом, гипотеза сдвига принимается как для первой, так и для второй (малой) группировки исходных данных.

6. АППРОКСИМАЦИЯ ИМАНА

Одним из наиболее точных, наряду с рассмотренными непараметрическими методами, является метод аппроксимации Имана [23]. Его эффективность 95%. Соответствующая J-статистика строится на основе W-статистики Вилкоксона:

$$J = \frac{W}{2} \left[1 + \left(\frac{n_1+n_2-2}{n_1+n_2-1-W^2} \right)^{0.5} \right].$$

Далее полученное значение сравнивается с критическим: $J_{\alpha'} = (z_{\alpha'} + t_{\alpha'})/2$. Здесь $z_{\alpha'}$ и $t_{\alpha'}$ есть α'-квантили нормального распределения и распределения Стьюдента с $r = n_1 + n_2 - 2$ степенями свободы соответственно. Для $\alpha' = 0.95$ критическая статистика есть $J_{\alpha'} = 2.00$. Соответствующее расчетное значение для второй группы данных $J = 3.53 > 2.00$, что свидетельствует о подтверждении гипотезы сдвига.

7. РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ НЕГРУППИРОВАННЫХ ИСХОДНЫХ ДАННЫХ

Приведем результаты статистической обработки исходных наблюдательных данных ($N_1 = 31$, $N_2 = 29$, табл. 1). Так как со стороны рассмотренных непараметрических критериев никаких теоретических ограничений сверху на объемы выборок нет, то мы можем проанализировать их с помощью этих критериев. Отметим, что теперь рангами рассматриваемых величин будут их порядковые номера в общем ранжированном ряду, потому что совпадающих значений среди элементов выборок нет. В табл. 6 приводятся результаты вычислений для исходных данных, а также сводка всех расчетных величин, полученных при первой и второй группировках исходных данных.

Табл. 6. Итоговые результаты для исследуемых выборок

Непараметрический критерий сдвига	$N_1 = 31$, $N_2 = 29$	$n_1 = 24$, $n_2 = 13$	Критическое значение	$m_1 = 12$, $m_2 = 7$	Критическое значение
Быстрый ранговый	2.17	2.27	1.96	1.77	1.96
Ван—дер—Вардена	2.12	2.26	1.96	3.02	3.62
Манна—Уитни—Вилкоксона	2.18	3.25	1.96	15	18 <. . . < 63
Аппроксимация Имана	2.22*	3.53	2.00 (1.98*)	—	—

Табл. 7. Результаты анализа однородных выборок

Непараметрический критерий сдвига	$n_1 = 31$, $m_1 = 29$	$n_2 = 24$, $m_2 = 13$	Критическое значение	$n_3 = 12$, $m_3 = 7$	Критическое значение
Быстрый ранговый	1.04	0.03	1.96	0.08	1.96
Ван—дер—Вардена	0.97	0.06	1.96	0.02	3.62
Манна—Уитни—Вилкоксона	1.04	0.67	1.96	31	18 <...< 63
Аппроксимация Имана	1.04*	0.67	2.00 (1.98*)	—	—

8. ТЕСТИРОВАНИЕ ИСПОЛЬЗУЕМЫХ МЕТОДОВ НА СИНТЕЗИРОВАННЫХ ВЫБОРКАХ

Чтобы продемонстрировать справедливость полученных результатов, мы протестировали все использованные ранговые критерии сдвига на заведомо однородных выборках, представленных в табл. 4. Были взяты выборки того же объема и структуры, что и рассмотренные ранее. Так, например, в однородных выборках с $n_2 = 24, m_2 = 13$ имеются повторяющиеся числа, а диапазон значений — целочисленный, от 0 до 6. Как можно видеть из табл. 7, использованные для анализа критерии однозначно подтверждают одинаковость распределения случайных величин в тестовых выборках, что указывает на устойчивость работы критериев.

ЗАКЛЮЧЕНИЕ

В данной работе ставилась задача статистического сравнении распределения ГЛ-пар источников (галактик) в полях, заведомо не содержащих КС (контрольное поле), с распределением аналогичных пар в поле CSc-1, где, по данным анизотропии РИ, расположен КС кандидат. В случае действительного присутствия КС в соответствующем поле не только должен наблюдаться статистический избыток ГЛ-пар (который был получен в [18]), но и сама плотность распределения ГЛ-пар должна отличаться от соответствующей плотности распределения классических ГЛ-событий в не содержащих струн полях.

Сравнительный анализ плотности распределения пар линзированных галактик проводился путем доказательства неоднородности двух наборов статистических данных. Выбранные для этой цели методы математической статистики, а именно непараметрические ранговые критерии сдвига позволили работать с малыми выборками, плотности распределения которых заранее не известны. Обсуждалась правильность работы используемых критериев на основе статистической обработки синтезированных выборок с заранее известными параметрами.

Полученные результаты указывают на статистическое различие распределений ГЛ-пар в CSc-1 поле по сравнению с контрольным полем. Выявленное различие служит дополнительным аргументом в пользу наличия в исследуемом поле КС.

Таким образом, существование КС в поле от ($\alpha = 11 : 29 : 03$, $\delta = +15 : 23 : 37$) до ($\alpha = 10 : 57 : 47$, $\delta = +25 : 03 : 51$)) [7, 17, 18] находит подтверждение двумя независимыми астрофизическими методами: как с помощью фильтрационной обработки данных анизотропии РИ, так и с помощью оптических методов, путем обнаружения статистического избытка с аномальным законом распределения ГЛ-пар.

Авторы выражают благодарность профессору М. В. Сажину за обсуждения и многочисленные полезные замечания. Работа выполнена при поддержке гранта Программы развития МГУ «Ведущая научная школа “Физика звезд, релятивистских объектов и галактик”».

СПИСОК ЛИТЕРАТУРЫ

1. *Copeland E. J., Kibble T. W. B.* // Proc. Roy. Soc. Lond. 2010. **A466**. P. 623.
2. *Vilenkin A.* // Phys. Rev. D. 1981. **23**. P. 852.
3. *Zeldovich J. B.* // MNRAS. 1980. **192**. P. 663.
4. *Davis A.-C., Brax P., van de Bruck C.* // Phil. Trans. Roy. Soc. Lond. 2008. **A366**. P. 2833.
5. *Durrer R., Kunz N., Melchiorri A.* // Phys. Rep. 2002. **364**. P. 1.
6. *Kibble T. W. B., Vachaspati T.* // J. Phys. G: Nucl. Part. Phys. 2015. **42**. 094002.
7. *Sazhina O. S., Scognamiglio D., Sazhin M. V.* // Eur. Phys. J. C. 2014. **74**. P. 2972.
8. *Polchinski J.* // Lectures presented at the 2004 Cargese Summer School. 2004. hep-th/0412244.
9. *Stebbins A.* // Ap. J. 1988. **327**. P. 584.
10. *Sazhina O. S.,Sazhin M. V., Sementsov V. N.* // JETP. 2008. **133**, N 5. P. 1005.
11. *Sazhina O. S., Sazhin M. V., Sementsov V. N.* et al. // 5th International Seminar on High Energy Physics QUARKS 2008. Electronic proceedings.
12. *Sazhin M. V. et al.* // MNRAS 2003. **343**, N 2. P. 353.
13. *Sazhin M. V., Capaccioli M., Longo G.* et al. // Astrophys. J. 2005. **636**. P. L5.

14. *Sazhin M. V., Khovanskaya O. S.* et al. // MNRAS. 2007. **376**. P. 1731.
15. *Keeton C. R.* // A Catalog of Mass Models for Gravitational Lensing. 2002. arXiv:astro-ph/0102341v2.
16. *Vilenkin A., Shellard E. P.S.* // Cosmic Strings and Other Topological Defects. Cambridge University Press, 2001.
17. *Sazhina O. S., Sementsov V. N., Ashimbaeva N. T.* // Astron. Rep. 2014 **58** **1**. P. 16.
18. *Sazhina O. S.* et al. // MNRAS 2019. **485**. P. 1876.
19. *Сажина О. С.* // Вестн. Моск. ун-та. Физ. Астрон. 2011. № 6. С. 93 (*Sazhina O. S., Sazhin M. V.* // Moscow Univ. Phys. Bull. 2011. **66**. P. 588.)
20. *Кобзарь А. И.* // Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006.
21. *ван дер Варден Б. Л.* // Математическая статистика. М., 1960.
22. *Mann H. B., Whitney D. R.* // The Ann. of Math. Stats. 1947. **18**, N 1. P. 50.
23. *Iman Ronald L.* // Proceedings of the Seventh Annual SAS Users Group International Conference. 1982. **7**. P. 676.

The Use of Nonparametric Methods of Mathematical Statistics to Search for Cosmic Strings

A. V. Morgunova[1,a], **O. S. Sazhina**[2,b]

[1]*Department of Physics, Lomonosov Moscow State University. Moscow 119991, Russia.*
[2]*Sternberg State Institute of Astronomy, Lomonosov Moscow State University. Moscow 119191, Russia.*
E-mail: [a]*arina.morgunova@yandex.ru*, [b]*cosmologia@yandex.ru*.

A part of the celestial sphere which has a cosmic string candidate (CSc-1 field) according to the radio data on cosmic microwave background anisotropy is statistically investigated to further substantiate this candidate. Cosmic strings are 1D cosmological-scale objects that can manifest themselves in particular as gravitational lenses with characteristic features. The problem of statistical comparison of the density distribution of gravitational-lens pairs of sources (galaxies) in the fields without cosmic strings with the distribution of similar pairs in the CSc-1 field is solved. Nonparametric rank tests are used to handle small samples under an unknown distribution law. The limits of applicability of the test are discussed. The revealed difference between the two distributions serves as an additional argument in favor of the presence of a string in the CSc-1 field. The discovery of cosmic strings would serve as a unique observational test of modern cosmological theories, including multidimensional models.

Keywords: cosmic strings, gravitational lensing, statistical data processing.
PACS: 02.60.-x, 95.75.Mn, 98.70.Vc, 98.80.-k, 95.35.+d, 95.36.+x.
Received 06 April 2019.

English version: *Moscow University Physics Bulletin*. 2019. **74**, No. 5. Pp. 529–536.

Сведения об авторах

1. Моргунова Арина Владимировна — студент; e-mail: arina.morgunova@yandex.ru.
2. Сажина Ольга Сергеевна — доктор физ.-мат. наук, вед. науч. сотрудник, доцент; тел.: (495) 939-50-06, e-mail: cosmologia@yandex.ru.