



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«МИРЭА – Российский технологический университет»**

ИНСТИТУТ КИБЕРНЕТИКИ

КАФЕДРА ВЫСШЕЙ МАТЕМАТИКИ

## **Лабораторная работа 4**

по курсу «Теория вероятностей и математическая статистика, часть 2»

Тема: **Проверка статистических гипотез с помощью критерия Колмогорова**

Выполнил:  
Студент 3-го курса  
Жолковскиц Д.А.

Группа: КМБО-01-16

МОСКВА 2019

**Лабораторная работа по Математической статистике № 4**  
**«Проверка статистических гипотез с помощью критерия**  
**Колмогорова»**

**Задание I. Проверка гипотезы о равномерном распределении.**

Из файла UD-2 в соответствии с номером варианта взять выборку  $\{x_1, \dots, x_N\}$  и значения  $a$  и  $b$ .

**Построить:**

на одном рисунке график эмпирической функции распределения этой выборки и график функции распределения равномерного закона на отрезке  $[a, b]$ .

**Проверить** в соответствии с Указаниями гипотезу о соответствии выборки равномерному распределению на отрезке  $[a, b]$  при уровне значимости 0,05 двумя способами – непосредственным расчетом и с помощью функции Python (`scipy.stats.kstest(sorted(array), 'uniform', (a, b))`).

**II. Проверка гипотезы об одинаковом распределении случайных величин.**

В соответствии с номером варианта взять выборку  $\{x_1, \dots, x_N\}$  из файла UD-2 и выборку  $\{y_1, \dots, y_M\}$  из файла UD-3.

**Построить:**

на одном рисунке графики эмпирических функций распределения этих выборок.

**Проверить** в соответствии с Указаниями гипотезу об одинаковом распределении двух случайных выборок при уровне значимости 0,05 двумя способами – непосредственным расчетом и с помощью функции Python (`scipy.stats.ks_2samp(arr_1, arr_2)`).

Графики должны быть отчётливыми: размер по горизонтали – по ширине страницы, по вертикали – не менее половины страницы; деления по горизонтали – через 1, по вертикали – через 0,1.

Вычисления проводить с точностью до 0,00001.

## Краткие теоретические сведения

### Равномерное распределение на отрезке $[a, b]$

Характеристика	Значение
Плотность	$f(x) = \begin{cases} 0, x \notin (a, b) \\ \frac{1}{b-a}, x \in (a, b) \end{cases}$
Функция распределения	$F(x) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, x \in (a, b) \\ 1, x \geq b \end{cases}$
Математическое ожидание	$\frac{a+b}{2}$
Дисперсия	$\frac{(b-a)^2}{12}$
Среднее квадратичное отклонение	$\frac{b-a}{2\sqrt{3}}$

Пусть имеется случайная выборка  $\{X_1, \dots, X_N\}$  из непрерывного распределения с функцией распределения  $F(x)$ .

Эмпирическая функция распределения выборки

$$F_N^{\exists}(x; x_1, x_2, \dots, x_N) = \sum_{x_i \leq x} I_{(-\infty; x]}(x_j)$$

Статистика Колмогорова

$$D_N = \sup_{-\infty < x < +\infty} |F_N^{\exists}(x) - F(x)|$$

Теорема Колмогорова

Пусть  $F(x)$  – непрерывна. Тогда

$$P(D_N \sqrt{N} \leq z) \rightarrow_{N \rightarrow \infty} K(z) = \begin{cases} 0, z \leq 0 \\ \sum_{j=-\infty}^{+\infty} (-1)^j I^{-2j^2 z^2}, z > 0 \end{cases}$$

Общая схема проверки с помощью критерия Колмогорова статистических гипотез о соответствии выборки равномерному распределению

Функция распределения равномерного закона на отрезке  $[a, b]$

$$F(x) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, x \in (a, b) \\ 1, x \geq b \end{cases}$$

Эмпирическая функция распределения выборки:

$$F_N^{\exists}(x; x_1, x_2, \dots, x_N) = \sum_{x_i \leq x} I_{(-\infty; x]}(x_j)$$

$$D_N = \max_{1 \leq j \leq N} (\max(|F_N(x_{(j)}) - F(x_{(j)})|, |F_N(x_{(j)} - 0) - F(x_{(j)})|))$$

Если  $D_N \sqrt{N} < k_{\alpha}$ , то гипотеза принимается.

Если  $D_N \sqrt{N} \geq k_{\alpha}$ , то гипотеза отвергается.

Критерий Колмогорова проверки гипотезы об одинаковом распределении наблюдаемых величин

$$D_{N,M} = \max_{j,k} \{ |F_N(x_{(j)}) - F_M(x_{(j)})|, |F_N(x_{(j)} - 0) - F_M(x_{(j)})|, \\ |F_N(y_{(k)}) - F_M(y_{(k)})|, |F_N(y_{(k)} - 0) - F_M(y_{(k)} - 0)| \};$$

$$K_{N,M} = D_{N,M} \sqrt{\frac{NM}{N+M}};$$

## **Средства языка программирования Python которые использованы в программе расчета**

В программе расчёта используются следующие средства языка:

Функции:

- `numpy.sort()` – сортирует вектор в порядке возрастания.
- `sorted(arr)` – возвращает отсортированную в порядке возрастания копию массива `arr`.
- `max(arr)` – возвращает максимальный элемент из массива `arr`.
- `min(arr)` – возвращает минимальный элемент из массива `arr`.
- `scipy.stats.kstest(array, 'uniform', (a, b)).pvalue` – проверка гипотезы о принадлежности выборки `data` распределению при помощи критерия Колмогорова-Смирнова.
- `scipy.stats.ks_2samp(arr_1, arr_2).pvalue` – проверка являются ли выборки из одного распределения.

Построение графиков и гистограмм было осуществлено с помощью библиотеки `matplotlib`. А построение таблиц было реализовано средствами библиотеки `docx`.

## Результаты расчетов

Задание 1)

$a = 2.0$

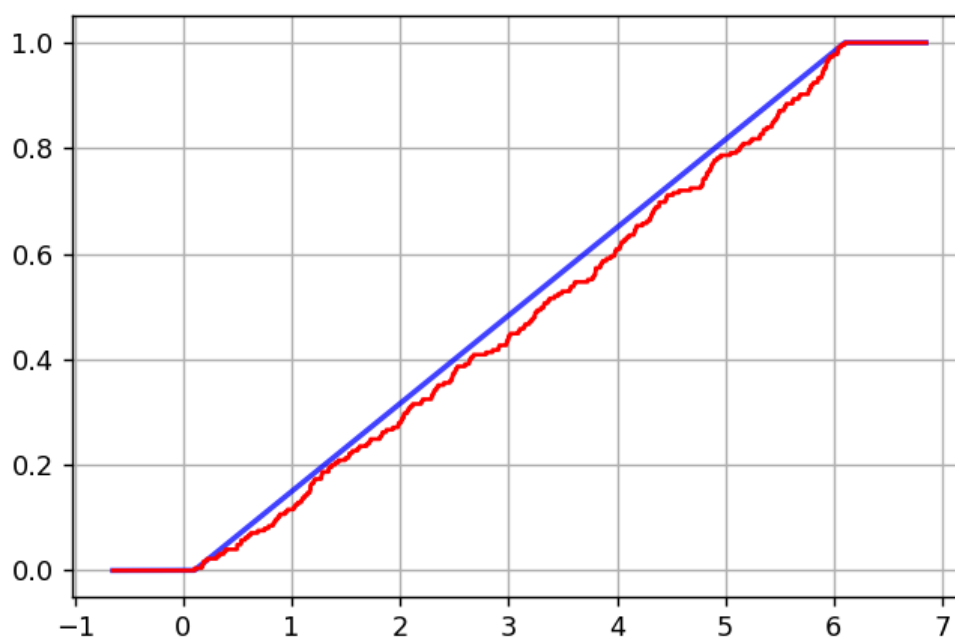
$b = 7.5$

Полученная выборка:

5.48848	3.01468	1.52398	4.17952	3.77314	4.15594	5.93128	0.78484	4.35748	5.76826
2.50318	0.49498	1.08742	5.12908	4.03246	2.48062	5.49532	4.83544	0.38986	5.56102
5.76352	4.31086	4.30144	5.81002	3.32386	2.99866	1.11922	1.98514	3.90286	3.85360
4.31812	3.99022	1.36768	5.90710	2.97988	0.49612	0.87382	3.18976	5.61922	4.17046
4.38838	5.46724	5.04448	3.97648	5.62390	1.01638	3.32800	4.87192	4.27534	5.93236
4.08688	2.29270	4.79428	4.57180	1.14088	5.92648	3.72838	5.78992	2.91190	5.95954
2.31490	5.22262	5.90020	4.33204	0.59668	5.44888	4.11088	3.56350	2.99590	4.16296
2.63944	5.99356	3.98080	2.48494	2.63464	2.33446	3.20950	4.81978	1.05436	5.42752
0.52954	0.82522	2.52532	2.90710	5.89408	2.09272	4.86634	2.06416	2.51224	0.93814
2.07454	1.33936	5.78236	2.03500	1.39936	0.10372	0.95908	3.61018	1.17760	5.48110
1.21030	3.78706	2.30752	4.46176	0.18784	2.67502	0.18010	0.30586	0.85420	2.79004
5.43328	4.33996	6.03304	1.42936	2.03332	4.88668	2.40160	4.79224	1.27930	5.68678
3.24352	3.22804	5.16154	5.94868	2.20612	2.28754	1.27138	1.93174	5.88634	1.68946
0.74812	1.72516	2.11474	3.94108	5.67454	4.82668	4.03888	5.54794	5.85418	4.06408
1.87168	0.62050	1.71376	3.24424	3.26290	4.23484	0.53296	4.45528	4.51744	4.44412
3.80044	0.16912	6.09958	4.38868	3.80344	1.19848	1.17004	1.17172	0.37240	1.81198
2.00848	3.13882	4.95424	4.85422	5.14078	2.59702	0.88810	1.83604	0.21754	2.34778
3.15160	0.68506	2.98336	1.62460	0.83896	2.45788	1.34152	1.55722	4.67602	2.47720
6.06682	2.01586	1.03162	3.49174	3.29734	5.84986	5.38360	5.34892	4.02508	3.57148
3.10180	4.89724	1.98040	2.19358	4.78906	6.03904	1.09852	1.49374	5.31676	3.60034
4.92538	3.80140	2.85154	3.36118	3.42166	3.86650	3.37708	6.03988	3.09022	5.24374
3.84316	0.57010	1.16368	4.77148	1.60462	1.53160	5.10202	1.27390	3.45700	5.35168
1.82686	0.32992	5.32156	2.65498	5.53840					

Упорядоченная выборка:

0.10372	0.16912	0.18010	0.18784	0.21754	0.30586	0.32992	0.37240	0.38986	0.49498
0.49612	0.52954	0.53296	0.57010	0.59668	0.62050	0.68506	0.74812	0.78484	0.82522
0.83896	0.85420	0.87382	0.88810	0.93814	0.95908	1.01638	1.03162	1.05436	1.08742
1.09852	1.11922	1.14088	1.16368	1.17004	1.17172	1.17760	1.19848	1.21030	1.27138
1.27390	1.27930	1.33936	1.34152	1.36768	1.39936	1.42936	1.49374	1.52398	1.53160
1.55722	1.60462	1.62460	1.68946	1.71376	1.72516	1.81198	1.82686	1.83604	1.87168
1.93174	1.98040	1.98514	2.00848	2.01586	2.03332	2.03500	2.06416	2.07454	2.09272
2.11474	2.19358	2.20612	2.28754	2.29270	2.30752	2.31490	2.33446	2.34778	2.40160
2.45788	2.47720	2.48062	2.48494	2.50318	2.51224	2.52532	2.59702	2.63464	2.63944
2.65498	2.67502	2.79004	2.85154	2.90710	2.91190	2.97988	2.98336	2.99590	2.99866
3.01468	3.09022	3.10180	3.13882	3.15160	3.18976	3.20950	3.22804	3.24352	3.24424
3.26290	3.29734	3.32386	3.32800	3.36118	3.37708	3.42166	3.45700	3.49174	3.56350
3.57148	3.60034	3.61018	3.72838	3.77314	3.78706	3.80044	3.80140	3.80344	3.84316
3.85360	3.86650	3.90286	3.94108	3.97648	3.98080	3.99022	4.02508	4.03246	4.03888
4.06408	4.08688	4.11088	4.15594	4.16296	4.17046	4.17952	4.23484	4.27534	4.30144
4.31086	4.31812	4.33204	4.33996	4.35748	4.38838	4.38868	4.44412	4.45528	4.46176
4.51744	4.57180	4.67602	4.77148	4.78906	4.79224	4.79428	4.81978	4.82668	4.83544
4.85422	4.86634	4.87192	4.88668	4.89724	4.92538	4.95424	5.04448	5.10202	5.12908
5.14078	5.16154	5.22262	5.24374	5.31676	5.32156	5.34892	5.35168	5.38360	5.42752
5.43328	5.44888	5.46724	5.48110	5.48848	5.49532	5.53840	5.54794	5.56102	5.61922
5.62390	5.67454	5.68678	5.76352	5.76826	5.78236	5.78992	5.81002	5.84986	5.85418
5.88634	5.89408	5.90020	5.90710	5.92648	5.93128	5.93236	5.94868	5.95954	5.99356
6.03304	6.03904	6.03988	6.06682	6.09958					



$a$	$b$	$N$	$D_N$	$D_N\sqrt{N}$	$x^*$	$F(x^*)$	$F_N(x^*)$	$F_N(x^* - 0)$
0.10000	6.10000	225	0.06108	0.91618	3.77314	0.61219	0.55556	0.55111

При втором способе проверки гипотезы о соответствии выборки равномерному распределению с помощью критерия Колмогорова значение функции:  $pval = 0.35840$

Задание 2)

Полученная выборка:

X

5.48848	3.01468	1.52398	4.17952	3.77314	4.15594	5.93128	0.78484	4.35748	5.76826
2.50318	0.49498	1.08742	5.12908	4.03246	2.48062	5.49532	4.83544	0.38986	5.56102
5.76352	4.31086	4.30144	5.81002	3.32386	2.99866	1.11922	1.98514	3.90286	3.85360
4.31812	3.99022	1.36768	5.90710	2.97988	0.49612	0.87382	3.18976	5.61922	4.17046
4.38838	5.46724	5.04448	3.97648	5.62390	1.01638	3.32800	4.87192	4.27534	5.93236
4.08688	2.29270	4.79428	4.57180	1.14088	5.92648	3.72838	5.78992	2.91190	5.95954
2.31490	5.22262	5.90020	4.33204	0.59668	5.44888	4.11088	3.56350	2.99590	4.16296
2.63944	5.99356	3.98080	2.48494	2.63464	2.33446	3.20950	4.81978	1.05436	5.42752
0.52954	0.82522	2.52532	2.90710	5.89408	2.09272	4.86634	2.06416	2.51224	0.93814
2.07454	1.33936	5.78236	2.03500	1.39936	0.10372	0.95908	3.61018	1.17760	5.48110
1.21030	3.78706	2.30752	4.46176	0.18784	2.67502	0.18010	0.30586	0.85420	2.79004
5.43328	4.33996	6.03304	1.42936	2.03332	4.88668	2.40160	4.79224	1.27930	5.68678
3.24352	3.22804	5.16154	5.94868	2.20612	2.28754	1.27138	1.93174	5.88634	1.68946
0.74812	1.72516	2.11474	3.94108	5.67454	4.82668	4.03888	5.54794	5.85418	4.06408
1.87168	0.62050	1.71376	3.24424	3.26290	4.23484	0.53296	4.45528	4.51744	4.44412
3.80044	0.16912	6.09958	4.38868	3.80344	1.19848	1.17004	1.17172	0.37240	1.81198
2.00848	3.13882	4.95424	4.85422	5.14078	2.59702	0.88810	1.83604	0.21754	2.34778
3.15160	0.68506	2.98336	1.62460	0.83896	2.45788	1.34152	1.55722	4.67602	2.47720
6.06682	2.01586	1.03162	3.49174	3.29734	5.84986	5.38360	5.34892	4.02508	3.57148
3.10180	4.89724	1.98040	2.19358	4.78906	6.03904	1.09852	1.49374	5.31676	3.60034

4.92538	3.80140	2.85154	3.36118	3.42166	3.86650	3.37708	6.03988	3.09022	5.24374
3.84316	0.57010	1.16368	4.77148	1.60462	1.53160	5.10202	1.27390	3.45700	5.35168
1.82686	0.32992	5.32156	2.65498	5.53840					

Y

5.99054	8.21068	8.29752	3.25086	8.02166	7.52643	2.88738	6.81390	5.44012	3.01913
6.05931	3.95117	7.67268	3.16291	4.54228	5.19436	4.42853	4.57575	8.16635	2.94503
7.02866	1.88878	3.63975	7.39344	8.05735	6.62832	4.24256	3.00347	2.75653	6.62345
2.64876	3.01484	2.43381	7.88529	1.85368	2.56894	5.17447	6.06672	4.18569	2.89778
8.03837	4.16333	7.19616	5.21015	6.96671	3.15258	2.33072	2.53963	2.20507	8.02413
3.77027	5.02113	3.22720	3.11910	4.16352	3.35713	5.50298	5.02016	2.69017	3.56026
4.11594	2.49010	6.30202	5.17031	3.48232	3.94506	4.39993	4.69873	4.92259	4.05731
8.00171	4.72824	7.90791	6.35123	1.82424	7.09405	6.48597	6.37177	5.89441	6.92212
3.76403	7.48268	4.46311	5.80919	1.82411	3.05625	4.45095	6.28694	3.60381	8.22921
6.75046	6.97061	6.01017	3.54511	2.08567	3.13841	7.26682	5.19384	4.45258	4.06167
4.27136	3.64287	3.50423	8.03700	5.61237	6.88098	6.45848	7.19356	4.94599	3.21589
5.32780	2.59689	7.44628	5.09582	4.07688	5.57526	7.26565	5.12169	4.44640	6.62091
7.28703	4.40981	5.55387	4.96458	5.63337	5.98151	8.03401	7.11309	5.02107	5.64396
6.73817	4.09963	8.07490	2.29627	6.93727	5.13846	2.70213	5.95746	7.89465	4.43470
3.99842	6.60947	6.61363	6.62579	4.49034	2.24258	2.40430	3.67524	5.09257	5.89129
2.76654	2.95946	5.78573	8.22940	3.52932	7.99898	7.27709	4.20434	2.17127	5.19878
7.35372	5.40658	5.60919	2.38441	6.78218	3.91054	8.16752	7.34384	3.37592	5.07352
1.98427	2.67041	6.06165	6.04443	8.19775	7.42308	3.33932	7.11836	7.46494	6.93603

Упорядоченная выборка:

X

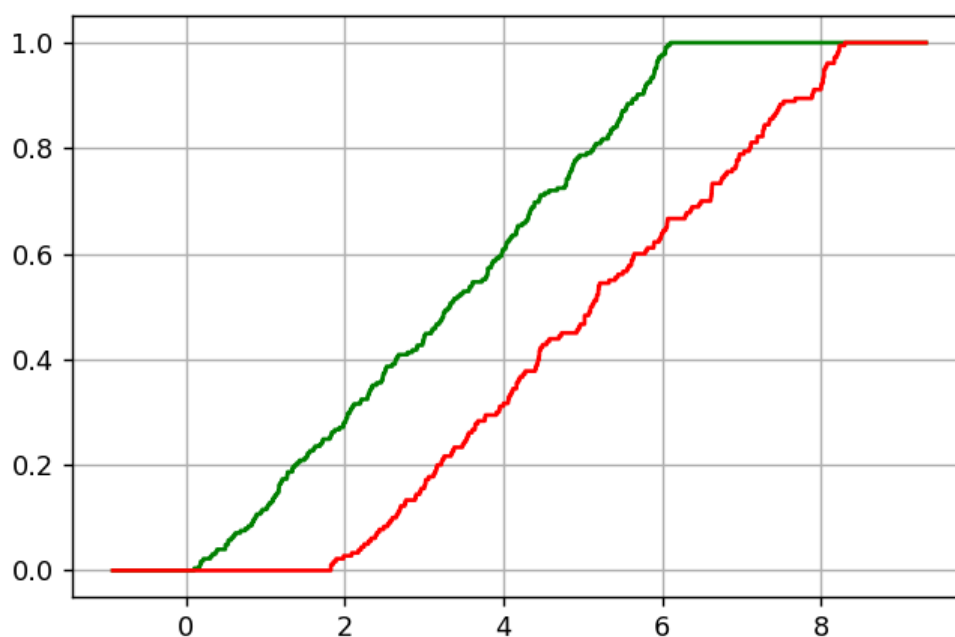
0.10372	0.16912	0.18010	0.18784	0.21754	0.30586	0.32992	0.37240	0.38986	0.49498
0.49612	0.52954	0.53296	0.57010	0.59668	0.62050	0.68506	0.74812	0.78484	0.82522
0.83896	0.85420	0.87382	0.88810	0.93814	0.95908	1.01638	1.03162	1.05436	1.08742
1.09852	1.11922	1.14088	1.16368	1.17004	1.17172	1.17760	1.19848	1.21030	1.27138
1.27390	1.27930	1.33936	1.34152	1.36768	1.39936	1.42936	1.49374	1.52398	1.53160
1.55722	1.60462	1.62460	1.68946	1.71376	1.72516	1.81198	1.82686	1.83604	1.87168
1.93174	1.98040	1.98514	2.00848	2.01586	2.03332	2.03500	2.06416	2.07454	2.09272
2.11474	2.19358	2.20612	2.28754	2.29270	2.30752	2.31490	2.33446	2.34778	2.40160
2.45788	2.47720	2.48062	2.48494	2.50318	2.51224	2.52532	2.59702	2.63464	2.63944
2.65498	2.67502	2.79004	2.85154	2.90710	2.91190	2.97988	2.98336	2.99590	2.99866
3.01468	3.09022	3.10180	3.13882	3.15160	3.18976	3.20950	3.22804	3.24352	3.24424
3.26290	3.29734	3.32386	3.32800	3.36118	3.37708	3.42166	3.45700	3.49174	3.56350
3.57148	3.60034	3.61018	3.72838	3.77314	3.78706	3.80044	3.80140	3.80344	3.84316
3.85360	3.86650	3.90286	3.94108	3.97648	3.98080	3.99022	4.02508	4.03246	4.03888
4.06408	4.08688	4.11088	4.15594	4.16296	4.17046	4.17952	4.23484	4.27534	4.30144
4.31086	4.31812	4.33204	4.33996	4.35748	4.38838	4.38868	4.44412	4.45528	4.46176
4.51744	4.57180	4.67602	4.77148	4.78906	4.79224	4.79428	4.81978	4.82668	4.83544
4.85422	4.86634	4.87192	4.88668	4.89724	4.92538	4.95424	5.04448	5.10202	5.12908
5.14078	5.16154	5.22262	5.24374	5.31676	5.32156	5.34892	5.35168	5.38360	5.42752
5.43328	5.44888	5.46724	5.48110	5.48848	5.49532	5.53840	5.54794	5.56102	5.61922
5.62390	5.67454	5.68678	5.76352	5.76826	5.78236	5.78992	5.81002	5.84986	5.85418
5.88634	5.89408	5.90020	5.90710	5.92648	5.93128	5.93236	5.94868	5.95954	5.99356
6.03304	6.03904	6.03988	6.06682	6.09958					

Y

1.82411	1.82424	1.85368	1.88878	1.98427	2.08567	2.17127	2.20507	2.24258	2.29627
2.33072	2.38441	2.40430	2.43381	2.49010	2.53963	2.56894	2.59689	2.64876	2.67041
2.69017	2.70213	2.75653	2.76654	2.88738	2.89778	2.94503	2.95946	3.00347	3.01484
3.01913	3.05625	3.11910	3.13841	3.15258	3.16291	3.21589	3.22720	3.25086	3.33932
3.35713	3.37592	3.48232	3.50423	3.52932	3.54511	3.56026	3.60381	3.63975	3.64287
3.67524	3.76403	3.77027	3.91054	3.94506	3.95117	3.99842	4.05731	4.06167	4.07688
4.09963	4.11594	4.16333	4.16352	4.18569	4.20434	4.24256	4.27136	4.39993	4.40981
4.42853	4.43470	4.44640	4.45095	4.45258	4.46311	4.49034	4.54228	4.57575	4.69873
4.72824	4.92259	4.94599	4.96458	5.02016	5.02107	5.02113	5.07352	5.09257	5.09582
5.12169	5.13846	5.17031	5.17447	5.19384	5.19436	5.19878	5.21015	5.32780	5.40658



5.44012	5.50298	5.55387	5.57526	5.60919	5.61237	5.63337	5.64396	5.78573	5.80919
5.89129	5.89441	5.95746	5.98151	5.99054	6.01017	6.04443	6.05931	6.06165	6.06672
6.28694	6.30202	6.35123	6.37177	6.45848	6.48597	6.60947	6.61363	6.62091	6.62345
6.62579	6.62832	6.73817	6.75046	6.78218	6.81390	6.88098	6.92212	6.93603	6.93727
6.96671	6.97061	7.02866	7.09405	7.11309	7.11836	7.19356	7.19616	7.26565	7.26682
7.27709	7.28703	7.34384	7.35372	7.39344	7.42308	7.44628	7.46494	7.48268	7.52643
7.67268	7.88529	7.89465	7.90791	7.99898	8.00171	8.02166	8.02413	8.03401	8.03700
8.03837	8.05735	8.07490	8.16635	8.16752	8.19775	8.21068	8.22921	8.22940	8.29752



$N$	$M$	$D_{N,M}$	$K_{N,M}$	$x^*$	$F_N(x^*)$	$F_N(x^* - 0)$	$F_M(x^*)$	$F_M(x^* - 0)$
225	180	0.34667	3.46667	5.94868	0.96889	0.96444	0.62222	0.62222

При втором способе проверки гипотезы о соответствии выборки равномерному распределению с помощью критерия Колмогорова значение функции:  $pval = 0.00000$

### Анализ результатов и выводы

Таблица критических значений:

$\alpha$	0.01	0.02	0.05	0.1	0.2
$K_\alpha$	1.63	1.57	1.36	1.22	1.07

Задание 1)

$D_N\sqrt{N}$	$\alpha$	$K_\alpha$	Верность гипотезы
0.91618	0.05	1.36	Да

$\alpha$	pval	Верность гипотезы
0.05000	0.35840	Да

Гипотеза о соответствии выборки равномерному распределению на отрезке  $[a, b]$  не противоречит экспериментальным данным (может быть принята) при уровне значимости  $\alpha = 0,05$ .

Задание 2)

$K_{N,M}$	$\alpha$	$K_\alpha$	Верность гипотезы
0.05000	0.35840	0.05000	Да

$\alpha$	pval	Верность гипотезы
0.05	0.00000	Нет

Гипотеза об одинаковом распределении двух случайных выборок противоречит экспериментальным данным (не может быть принята) при уровне значимости  $\alpha = 0,05$ .

### **Список использованной литературы**

1. Лобузов А.А. Математическая статистика [Электронный ресурс]: Методические указания по выполнению лабораторных работ / под ред. Ю. И. Худака. Москва: Московский технологический университет (МИРЭА), 2017. 36 с.
2. Чернова Н. И. Математическая статистика: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2007. 148 с

### Приложение (Листинг программы)

```

import math
import numpy

import matplotlib.mlab as mlab
import matplotlib.pyplot as plt

import scipy.stats as pystats
import scipy

from docx import Document

kolm = {
    0.01 : 1.63,
    0.02 : 1.57,
    0.05 : 1.36,
    0.1 : 1.22,
    0.2 : 1.07
}

head_1_str = [
    'a', 'b', 'N', 'D_N', 'D_N sqrt{N}', 'x^*', 'F (x^*)', 'F_N (x^*)', 'F_N
(x^* - 0)'
]
head_2_str = [
    'N', 'M', 'D_N,M', 'K_N,M', 'x^*', 'F_N (x^*)', 'F_N (x^* - 0)', 'F_M
(x^*)', 'F_M (x^* - 0)'
]

head_3_str = [
    'D_N sqrt{N}', 'alpha', 'k_alpha', 'Верность гипотезы'
]
head_4_str = [
    'alpha', 'pval', 'Верность гипотезы'
]

head_5_str = [
    'K_N,M', 'alpha', 'k_alpha', 'Верность гипотезы'
]
head_6_str = [
    'alpha', 'pval', 'Верность гипотезы'
]

```

```
hyp_text = {
    True : 'Да',
    False : 'Нет'
}
```

```
def strm(a):
    n = len(a)
    b = a[0]
    m = len(b)
    for i in range(n):
        for k in range(m):
            if type(a[i][k]) == int:
                a[i][k] = str(a[i][k])
            elif type(a[i][k]) != str:
                a[i][k] = str('%0.5f' % a[i][k])
```

```
def strm_lite(a):
    n = len(a)
    b = a[0]
    m = len(b)
    for i in range(n):
        for k in range(m):
            if type(a[i][k]) == int:
                a[i][k] = str(a[i][k])
            elif type(a[i][k]) != str:
                a[i][k] = str(a[i][k])
```

```
def line_out(l, cols = 10):
    tbl = []
    q = []
    for i in range(len(l)):
        q.append(l[i])
        if (len(q) == 10) or (i+1 == len(l)):
            while(len(q) != 10):
                q.append("")
            tbl.append(q)
            q = []
    return tbl
```

```
def onetablein(docname):
    document = Document(docname)
    table = document.tables[0]
    array = []
```

```

for i in range(0,len(table.rows)):
    row = table.rows[i]
    for cell in row.cells:
        array.append(float(cell.text.replace(',','.')))
return array

```

```

ud2 = onetablein('UD-2_10.docx')
ud2a = 2.0
ud2b = 7.5
ud3 = onetablein('UD-3_10.docx')
print(ud2, ud2a, ud2b)
print()
print(ud3)

```

```

class UniSample():
    def __init__(self, array, a, b):
        self.array = [i for i in array]
        self.n = len(array)
        self.a = a
        self.b = b

    def Fj(self,x):
        s = 0
        for i in self.array:
            if i <= x:
                s = s + 1
        return s/self.n

    def Fjo(self,x):
        s = 0
        for i in self.array:
            if i < x:
                s = s + 1
        return s/self.n

    def F(self,x):
        if x < self.a:
            return 0
        if x > self.b:
            return 1
        return (x-self.a)/(self.b-self.a)

    def Dn(self):

```

```

        a = [max([abs(self.Fj(i)-self.F(i)), abs(self.Fjo(i)-self.F(i))]) for i
in self.array]

```

```

        return(max(a))

```

```

def xx(self):
    for i in self.array:
        if self.Dn() == max([abs(self.Fj(i)-self.F(i)), abs(self.Fjo(i)-
self.F(i))]):
            return i

```

```

def out_1(self):
    x = self.xx()
    l = [
        self.a,
        self.b,
        self.n,
        self.Dn(),
        self.Dn()*(self.n**0.5),
        x,
        self.F(x),
        self.Fj(x),
        self.Fjo(x)
    ]
    return l

```

```

def Graf(self, filename = 'Uni_Graf.png', dx = 0.00001):
    xmin = self.a - (self.b - self.a)/8
    xmax = self.b + (self.b - self.a)/8
    xlist = mlab.frange (xmin, xmax, dx)
    ylist = [self.F(x) for x in xlist]
    zlist = [self.Fj(x) for x in xlist]
    plt.plot(xlist, ylist, color = 'blue', alpha = 0.75, linewidth = 2)
    plt.plot(xlist, zlist, color = 'red', linewidth = 1.5)
    plt.grid(True)
    plt.savefig(filename, dpi = 128)
    plt.clf()
    print('Done! [gr_1]')

```

```

def out_3(self, alpha = 0.05):
    dn = self.Dn()*(self.n**0.5)
    ka = kolm[alpha]
    l = [
        dn,
        str(alpha),
        str(ka),

```

```

        hyp_text[dn <= ka]
    ]
    return l

    def out_4(self, alpha = 0.05):
        g = pystats.kstest(self.array, scipy.stats.uniform(loc = self.a, scale
= self.b-self.a).cdf).pvalue
        l = [
            alpha,
            g,
            hyp_text[g >= alpha]
        ]
        return l

    def hyp(self, alpha = 0.05):
        dn = self.Dn()*(self.n**0.5)
        ka = kolm[alpha]
        return dn <= ka

us = UniSample(ud2, ud2a, ud2b)
print(us.out_1())
us.Graf(dx = 0.01)
print(us.out_3())
print(us.out_4())

class DoubleSample():
    def __init__(self, arr_1, arr_2):
        self.arr_1 = [i for i in arr_1]
        self.arr_2 = [i for i in arr_2]
        self.n = len(arr_1)
        self.m = len(arr_2)
        self.a = min(sorted(arr_1)[0], sorted(arr_2)[0])
        self.b = max(sorted(arr_1)[self.n - 1], sorted(arr_2)[self.m - 1])

    def Fn(self,x):
        s = 0
        for i in self.arr_1:
            if i <= x:
                s = s + 1
        return s/self.n

    def Fno(self,x):
        s = 0
        for i in self.arr_1:

```



```

        if i < x:
            s = s + 1
        return s/self.n

def Fm(self,x):
    s = 0
    for i in self.arr_2:
        if i <= x:
            s = s + 1
    return s/self.m

def Fmo(self,x):
    s = 0
    for i in self.arr_2:
        if i < x:
            s = s + 1
    return s/self.m

def Dnm(self):
    l = []
    l.extend([abs(self.Fn(i) - self.Fm(i)) for i in self.arr_1])
    l.extend([abs(self.Fno(i) - self.Fm(i)) for i in self.arr_1])
    l.extend([abs(self.Fn(i) - self.Fm(i)) for i in self.arr_2])
    l.extend([abs(self.Fn(i) - self.Fmo(i)) for i in self.arr_2])
    return max(l)

def Knm(self):
    return ((self.n*self.m/(self.n+self.m))**0.5)*self.Dnm()

def xx(self):
    for i in self.arr_1:
        if (self.Dnm() == abs(self.Fn(i) - self.Fm(i))):
            return i
        if (self.Dnm() == abs(self.Fno(i) - self.Fm(i))):
            return i
    for i in self.arr_2:
        if (self.Dnm() == abs(self.Fn(i) - self.Fm(i))):
            return i
        if (self.Dnm() == abs(self.Fn(i) - self.Fmo(i))):
            return i

def out_2(self):
    x = self.xx()
    l = [
        self.n,
        self.m,

```

```

        self.Dnm(),
        self.Knm(),
        x,
        self.Fn(x),
        self.Fno(x),
        self.Fm(x),
        self.Fmo(x)
    ]
    return l

def Graf(self, filename = 'Double_Graf.png', dx = 0.00001):
    xmin = self.a - (self.b - self.a)/8
    xmax = self.b + (self.b - self.a)/8
    xlist = mlab.frange (xmin, xmax, dx)
    ylist = [self.Fn(x) for x in xlist]
    zlist = [self.Fm(x) for x in xlist]
    plt.plot(xlist, ylist, color = 'Green', linewidth = 1.5)
    plt.plot(xlist, zlist, color = 'red', linewidth = 1.5)
    plt.grid(True)
    plt.savefig(filename, dpi = 128)
    plt.clf()
    print('Done! [gr_2]')

def out_5(self, alpha = 0.05):
    Knm = self.Knm()
    ka = kolm[alpha]
    l = [
        Knm,
        str(alpha),
        str(ka),
        hyp_text[Knm <= ka]
    ]
    return l

def out_6(self, alpha = 0.05):
    g = scipy.stats.ks_2samp(self.arr_1, self.arr_2).pvalue
    l = [
        str(alpha),
        g,
        hyp_text[g >= alpha]
    ]
    return l

def hyp(self, alpha = 0.05):
    Knm = self.Knm()
    ka = kolm[alpha]

```

```
return Knm <= ka
```

```
usd = DoubleSample(ud2, ud3)
print()
print(head_2_str)
print(usd.out_2())
print()
#usd.Graf(dx = 0.01)
print(head_5_str)
print(usd.out_5())
print()
print(head_6_str)
print(usd.out_6())
```

```
def tabler(document, out, head = None):
```

```
    tl = []
    if head != None:
        tl.append(head)
    for i in out:
        tl.append(i)
    strm(tl)
```

```
    table = document.add_table(rows = len(tl), cols = len(tl[0]))
    for i in range(len(tl)):
        hdr_cells = table.rows[i].cells
        for k in range(len(tl[0])):
            if type(tl[i][k]) == str:
                hdr_cells[k].text = tl[i][k]
            else:
                pass #LaTeh
```

```
text_anal_1 = {
    True : 'Гипотеза о соответствии выборки равномерному
распределению на отрезке [a, b] не противоречит экспериментальным
данным (может быть принята) при уровне значимости alpha = 0,05.',
    False : 'Гипотеза о соответствии выборки равномерному
распределению на отрезке [a, b] противоречит экспериментальным
данным (не может быть принята) при уровне значимости alpha = 0,05.'
}
```

```
text_anal_2 = {
```

True : 'Гипотеза об одинаковом распределении двух случайных выборок не противоречит экспериментальным данным (может быть принята) при уровне значимости  $\alpha = 0,05$ .'

True : 'Гипотеза об одинаковом распределении двух случайных выборок противоречит экспериментальным данным (не может быть принята) при уровне значимости  $\alpha = 0,05$ .'

}

```
def doccreator(arr_1, arr_2, a,b, filename = 'some'):
    document = Document()
    document.add_paragraph('Результаты расчетов')

    document.add_paragraph('Задание 1')
    tbl = line_out(arr_1)
    document.add_paragraph('Полученная выборка:')
    tabler(document,tbl)
    document.add_paragraph("")
    tbl = line_out(sorted(arr_1))
    document.add_paragraph('Упорядоченная выборка:')
    tabler(document, tbl)
    document.add_paragraph("")
    us = UniSample(arr_1, a, b)
    us.Graf(filename = filename + '_1.png')
    document.add_picture(filename + '_1.png')
    document.add_paragraph("")
    out_1 = us.out_1()
    tabler(document, [out_1], head_1_str)
    document.add_paragraph("")
    s = 'При втором способе проверки гипотезы о соответствии'
    s = s + ' выборки равномерному распределению с помощью'
    s = s + ' критерия Колмогорова значение функции: pval = '
    s = s + str("%.5f" % us.out_4()[1])
    document.add_paragraph(s)

    document.add_paragraph("")
    document.add_paragraph("")

    document.add_paragraph('Задание 2')
    document.add_paragraph('Полученная выборка:')
    document.add_paragraph('X')
    tbl = line_out(arr_1)
    tabler(document,tbl)
    document.add_paragraph("")
```

```

document.add_paragraph('Y')
tbl = line_out(arr_2)
tabler(document,tbl)
document.add_paragraph("")
document.add_paragraph('Упорядоченная выборка:')
document.add_paragraph('X')
tbl = line_out(sorted(arr_1))
tabler(document,tbl)
document.add_paragraph("")
document.add_paragraph('Y')
tbl = line_out(sorted(arr_2))
tabler(document,tbl)
document.add_paragraph("")
ds = DoubleSample(arr_1, arr_2)
ds.Graf(filename = filename + '_2.png')
document.add_picture(filename + '_2.png')
out_2 = ds.out_2()
tabler(document, [out_2], head_2_str)
s = 'При втором способе проверки гипотезы о соответствии'
s = s + ' выборки равномерному распределению с помощью'
критерия Колмогорова значение функции: rval = '
s = s + str("%.5f"% ds.out_6()[1])
document.add_paragraph(s)

document.add_paragraph("")
document.add_paragraph("")

document.add_paragraph('Анализ результатов и выводы')
document.add_paragraph('Таблица критических значений:')
tbl = [['alpha'], ['k_alpha']]
tbl[0].extend([str(i) for i in kolm])
tbl[1].extend([str(kolm[i]) for i in kolm])
tabler(document, tbl)
document.add_paragraph("")
document.add_paragraph('Задание 1')
tabler(document, [us.out_3()], head_3_str)
document.add_paragraph("")
tabler(document, [us.out_4()], head_4_str)
document.add_paragraph(text_anal_1[us.hyp()])
document.add_paragraph("")
document.add_paragraph('Задание 2')
tabler(document, [ds.out_5()], head_5_str)
document.add_paragraph("")
tabler(document, [ds.out_6()], head_6_str)
document.add_paragraph(text_anal_2[ds.hyp()])
document.add_paragraph("")

```

```
document.save(filename + '.docx')  
print('Done! [doc]')
```

```
doccreator(ud2, ud3, ud2a, ud2b)
```