

Министерство образования и науки Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого

—
Институт компьютерных наук и технологий
Кафедра «Информационная безопасность компьютерных систем»

**ОТЧЕТ
ПО РАСЧЕТНОМУ ЗАДАНИЮ № 1**

**«Получение на практике и анализ выборок,
приближенных к теоретическим распределениям»**

по дисциплине
«Теория вероятностей и математическая статистика»

Выполнил
студент группы 23508/4

Е.Г. Проценко

Проверила
ассистент

Д.С.Лаврова

Санкт-Петербург
2016

ХОД РАБОТЫ

Задание 1

Постановка задачи

Получить реализацию двух выборок равномерно распределенной случайной величины по правилам:

- $x_{k+1} = \{K \cdot x_k\}, k = 0, 1, \dots, 100;$
- $x_{k+1} = \{11 \cdot x_k + \pi\}, k = 0, 1, \dots, 100,$

где $\{x\}$ - дробная часть числа. $x_0 = 0.089, K = 53.$

Для полученных выборок необходимо построить вариационный ряд, найти выборочное среднее, выборочную дисперсию, исправленную выборочную дисперсию, медиану, функции распределения и сравнить эти параметры с теоретическими параметрами равномерно распределенной случайной величины.

Выполненная работа

С помощью написанной программы были получены два вариационных ряда.

Вариационный ряд 1 ($x_{k+1} = \{K \cdot x_k\}, k = 0, 1, \dots, 100$):

0.001	0.002	0.011	0.233	0.254	0.259	0.577	0.579	0.580	0.777	0.782	0.808
0.016	0.021	0.046	0.291	0.341	0.342	0.583	0.593	0.594	0.814	0.820	0.824
0.053	0.057	0.061	0.349	0.353	0.366	0.598	0.606	0.616	0.842	0.848	0.869
0.073	0.094	0.101	0.380	0.398	0.417	0.617	0.626	0.638	0.887	0.898	0.934
0.106	0.109	0.113	0.420	0.422	0.429	0.647	0.660	0.672	0.938	0.940	0.943
0.118	0.126	0.140	0.434	0.438	0.446	0.678	0.686	0.694	0.979	0.980	0.982
0.141	0.143	0.153	0.460	0.462	0.472	0.701	0.709	0.714	0.989		
0.174	0.178	0.181	0.482	0.486	0.497	0.717	0.727	0.737			
0.214	0.220	0.222	0.502	0.531	0.546	0.740	0.758	0.766			

Выборочное среднее - 0.486; теоретическое значение - 0.5

Выборочная дисперсия - 0.084; теоретическое значение - 0.083

Исправленная выборочная дисперсия - 0.085; теоретическое значение - 0.083

Медиана - 0.491; теоретическое значение - 0.5

Вариационный ряд 2 ($x_{k+1} = \{11 \cdot x_k + \pi\}, k = 0, 1, \dots, 100$):

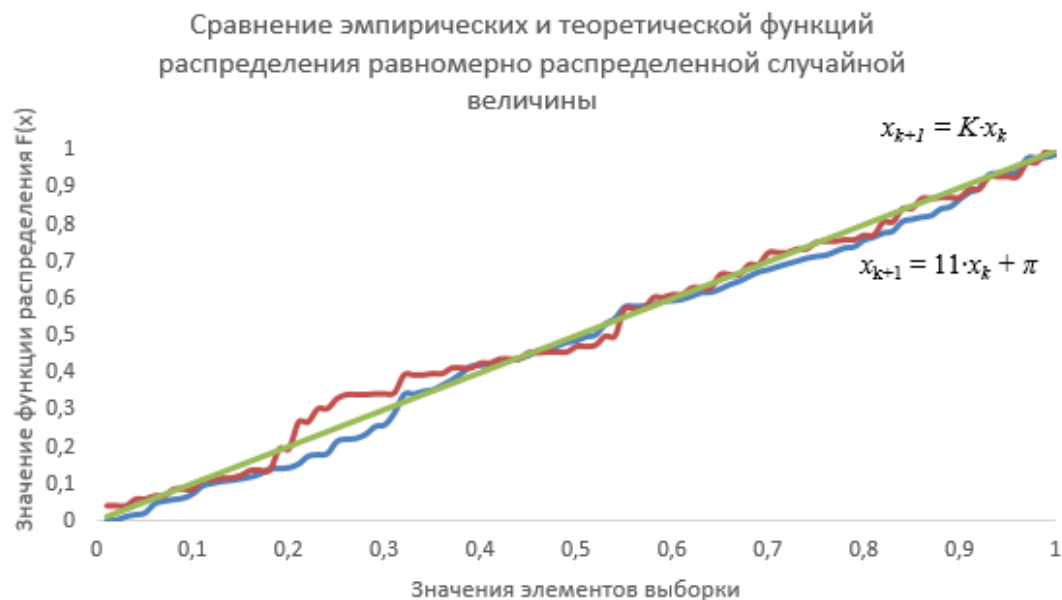
0.039	0.039	0.039	0.339	0.341	0.341	0.570	0.570	0.570	0.805	0.805	0.841
0.057	0.057	0.066	0.344	0.392	0.392	0.601	0.601	0.609	0.841	0.868	0.868
0.066	0.084	0.084	0.392	0.396	0.396	0.609	0.627	0.627	0.871	0.871	0.871
0.084	0.109	0.109	0.411	0.411	0.411	0.627	0.663	0.663	0.893	0.893	0.926
0.114	0.114	0.121	0.424	0.424	0.435	0.663	0.690	0.690	0.927	0.927	0.927
0.135	0.135	0.135	0.435	0.435	0.452	0.722	0.722	0.722	0.965	0.965	0.997
0.193	0.193	0.265	0.452	0.454	0.454	0.732	0.732	0.750	0.997		
0.265	0.301	0.301	0.454	0.469	0.469	0.753	0.753	0.757			
0.328	0.339	0.339	0.473	0.497	0.497	0.757	0.769	0.769			

Выборочное среднее - 0.509; теоретическое значение - 0.5;

Выборочная дисперсия - 0.078; теоретическое значение - 0.083;

Исправленная выборочная дисперсия - 0.079; теоретическое значение - 0.083;

Медиана - 0.469; теоретическое значение - 0.5.



Выводы

Характеристики для полученных выборок отличаются от теоретических значений на величину порядка 0.01.

На графике выше можно увидеть, что полученные выборки достаточно хорошо приближены к теоретическим значениям равномерно распределенной случайной величины.

Задание 2

Постановка задачи

На основе выборок из п.1 смоделировать последовательность испытаний Бернулли с вероятностью успеха $P = x_0$. Вычислить частоты $\frac{\mu_m}{m}$, где $\mu_m = x_1 + \dots + x_m$; $m = 10, \dots, 100$. Построить график зависимости $\frac{\mu_m}{m}$ от m .

Выполненная работа

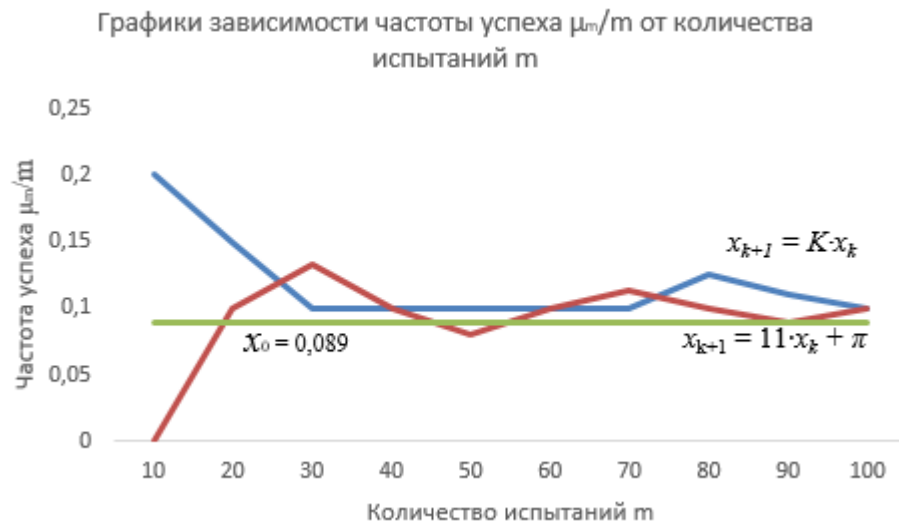
С помощью написанной программы были смоделированы последовательности испытаний Бернулли на основе выборок из п.1 и вычислены частоты успеха. Моделирование испытаний Бернулли было произведено в соответствии с данным правилом:

$$\begin{cases} b_i = 1, x_i < x_0 \\ b_i = 0, x_i \geq x_0, \end{cases}$$

где x_i - элемент выборки из п.1, b_i - смоделированное испытание Бернулли.

Зависимость частоты успехов от количества испытаний Бернулли

Испытания 1		Испытания 2	
Количество испытаний	Частота успеха	Количество испытаний	Частота успеха
10	0.200	10	0.000
20	0.150	20	0.100
30	0.100	30	0.133
40	0.100	40	0.100
50	0.100	50	0.080
60	0.100	60	0.100
70	0.100	70	0.114
80	0.125	80	0.100
90	0.111	90	0.089
100	0.100	100	0.100



Выводы

На графике выше можно увидеть, что при увеличении количества испытаний частота успеха стремится к $x_0 = 0.243$, что соответствует теории.

Задание 3

Постановка задачи

На основе выборки из п.1 получить выборки показательно распределенной случайной величины с параметрами $a_1 = 0.409$ и $a_2 = 3.09$.

Для полученных выборок необходимо построить гистограмму, найти выборочное среднее, выборочную дисперсию, исправленную выборочную дисперсию, медиану, функции распределения и сравнить эти параметры с теоретическими параметрами показательно распределенной случайной величины.

Выполненная работа

Для выполнения данного задания был использован метод обратного преобразования (Метод Н.В. Смирнова), суть которого заключается в использовании данной формулы: $X_i = \frac{\ln(1 - U_i)}{-a}$ [1], где U_i - элемент реализации выборки равномерно распределенной на $[0,1]$ случайной величины, X_i - элемент реализации выборки показательно распределенной случайной величины с параметром a .

На основе выборки 2 из п.1 были получены две выборки показательно распределенной случайной величины с различными параметрами.

Выборка 1 ($a = 0.409$)

0.097 0.097 0.097 0.143	1.020 1.020 1.031 1.217	2.063 2.246 2.246 2.296	4.496 4.951 4.951 5.007
0.143 0.167 0.167 0.215	1.217 1.217 1.233 1.233	2.296 2.411 2.411 2.411	5.007 5.007 5.464 5.464
0.215 0.215 0.282 0.282	1.294 1.294 1.294 1.349	2.659 2.659 2.659 2.864	6.366 6.399 6.399 6.399
0.296 0.296 0.315 0.355	1.349 1.396 1.396 1.396	2.864 3.130 3.130 3.130	8.197 8.197 14.203
0.355 0.355 0.524 0.524	1.471 1.471 1.480 1.480	3.219 3.219 3.389 3.419	14.203
0.753 0.753 0.876 0.876	1.480 1.548 1.548 1.566	3.419 3.459 3.459 3.583	
0.972 1.012 1.012 1.012	1.680 1.680 2.063 2.063	3.583 3.997 3.997 4.496	

Выборочное среднее - 2.473; теоретическое значение - 3.164;

Выборочная дисперсия - 6.262; теоретическое значение - 10.014;

Исправленная выборочная дисперсия - 6.325; теоретическое значение - 10.014;

Медиана - 1.548; теоретическое значение - 2.194.

Выборка 2 ($a = 3.09$)

0.013 0.013 0.013 0.019	0.135 0.135 0.136 0.161	0.273 0.297 0.297 0.304	0.595 0.655 0.655 0.663
0.019 0.022 0.022 0.028	0.161 0.161 0.163 0.163	0.304 0.319 0.319 0.319	0.663 0.663 0.723 0.723
0.028 0.028 0.037 0.037	0.171 0.171 0.171 0.179	0.352 0.352 0.352 0.379	0.843 0.847 0.847 0.847
0.039 0.039 0.042 0.047	0.179 0.185 0.185 0.185	0.379 0.414 0.414 0.414	1.085 1.085 1.880 1.880
0.047 0.047 0.069 0.069	0.195 0.195 0.196 0.196	0.426 0.426 0.449 0.453	
0.100 0.100 0.116 0.116	0.196 0.205 0.205 0.207	0.453 0.458 0.458 0.474	
0.129 0.134 0.134 0.134	0.222 0.222 0.273 0.273	0.474 0.529 0.529 0.595	

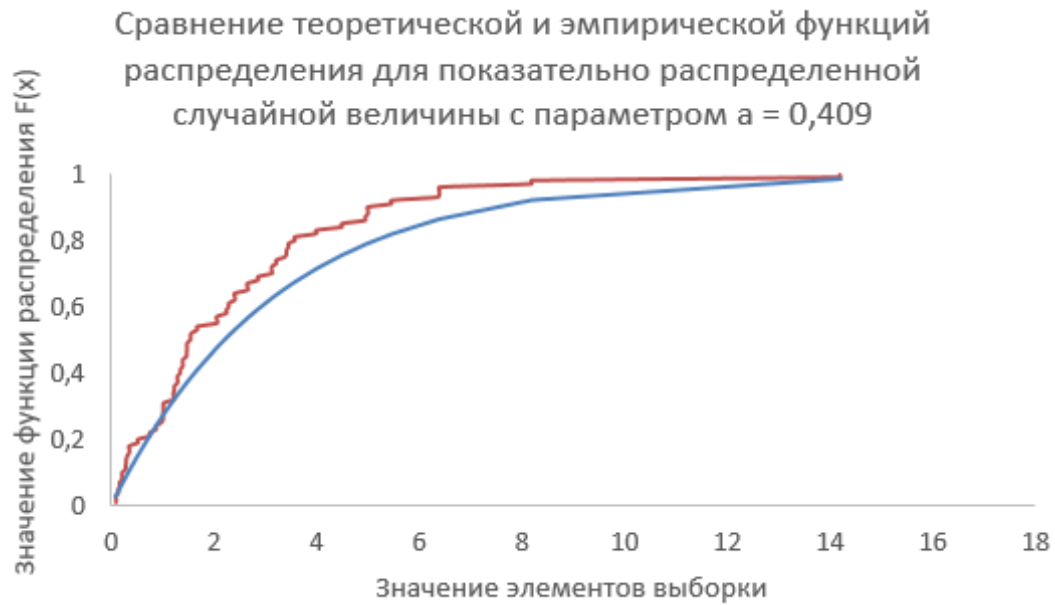
Выборочное среднее - 0.327; теоретическое значение - 0.463;

Выборочная дисперсия - 0.110; теоретическое значение - 0.214;

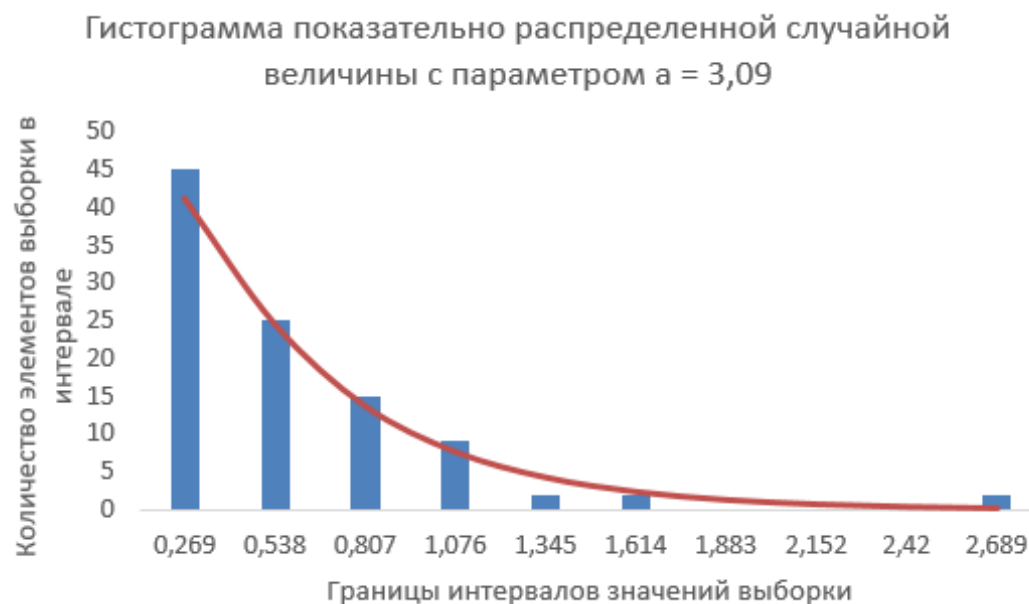
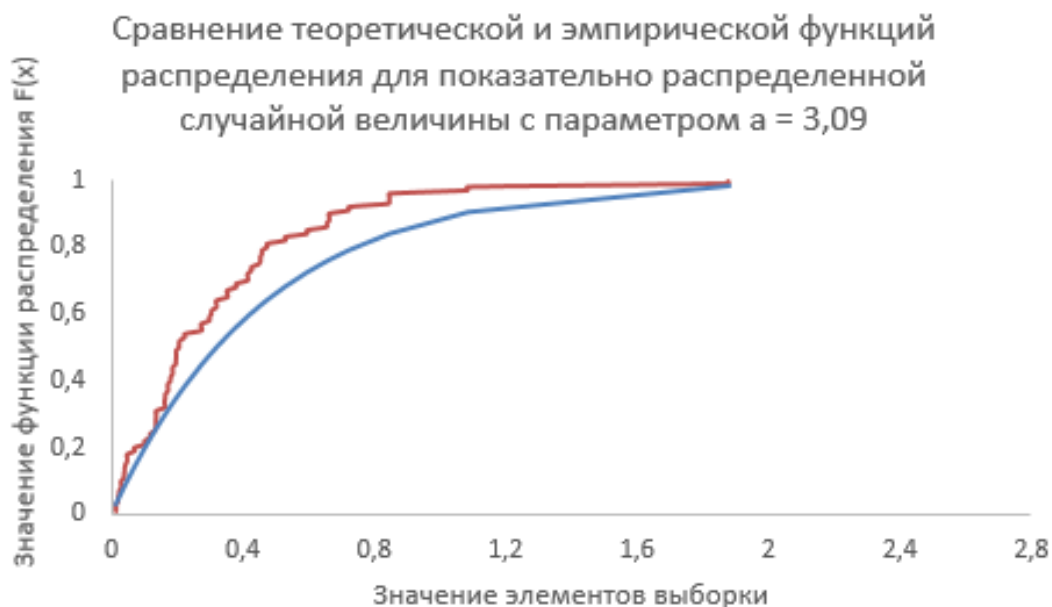
Исправленная выборочная дисперсия - 0.111; теоретическое значение - 0.214;

Медиана - 0.205; теоретическое значение - 0.321.

Графики для выборки 1



Графики для выборки 2



Выводы

На графиках выше можно увидеть, что полученные выборки достаточно хорошо приближены к теоретическим значениям экспоненциально распределенной случайной величины.

Характеристики для полученных выборок отличаются от теоретических значений на величину порядка 0.01.

Задание 4

Постановка задачи

На основе выборок из п.1 получить выборки нормально распределенной случайной величины с $M = 9$ и $D = 4.09$.

Для полученных выборок необходимо построить гистограмму, найти выборочное среднее, выборочную дисперсию, исправленную выборочную дисперсию, медиану, функции распределения и сравнить эти параметры с теоретическими параметрами нормально распределенной случайной величины.

Выполненная работа

С помощью формул $\eta_1 = \cos(2\pi\xi_1)\sqrt{-2\ln\xi_2}$ и $\eta_2 = \sin(2\pi\xi_1)\sqrt{-2\ln\xi_2}$, где η_1 и η_2 - элементы выборок нормально распределенных случайных величин, ξ_1 и ξ_2 - элементы выборок равномерно распределенных случайных величин из п.1, были получены две выборки случайной величины с нормальным стандартным распределением. Далее получившиеся величины были приведены к нормальному распределению с параметрами $M = 9$ и $D = 4.09$, используя формулу $\xi = \mu + \sigma z$, где z - случайная величина с нормальным стандартным распределением, $\mu = M$ и $\sigma = \sqrt{D}$.

Выборка 1 ($\eta_1 = \cos(2\pi\xi_1)\sqrt{-2\ln\xi_2}$)

4.010 4.178 4.978 5.568	7.638 7.712 7.790 7.913	9.266 9.268 9.283 9.295	10.815 10.822 11.009
5.797 5.941 6.026 6.082	7.924 7.929 7.968 7.968	9.409 9.662 9.667 9.671	11.220 11.240 11.475
6.202 6.351 6.550 6.564	8.112 8.147 8.271 8.279	9.833 9.833 9.836 9.853	11.501 11.522 11.524
6.842 6.890 6.905 6.936	8.279 8.279 8.378 8.419	9.938 9.954 9.954	11.699 11.729 11.778
7.035 7.127 7.241 7.276	8.520 8.608 8.852 8.875	10.140 10.147 10.178	11.792 11.897 11.974
7.281 7.393 7.436 7.463	8.887 8.907 8.946 9.075	10.356 10.359 10.473	12.118 12.764 12.948
7.469 7.518 7.565 7.631	9.141 9.212 9.217 9.223	10.475 10.506 10.719	13.317 13.767

Выборочное среднее - 8.957; теоретическое значение - 8;

Выборочная дисперсия - 4.039; теоретическое значение - 4.09;

Исправленная выборочная дисперсия - 4.079; теоретическое значение - 4.09;

Медиана - 8.927; теоретическое значение - 8.

Выборка 2 ($\eta_2 = \sin(2\pi\xi_1)\sqrt{-2\ln\xi_2}$)

4.175 4.710 4.828 4.857	7.646 7.695 7.718 7.847	9.315 9.317 9.352 9.423	10.279 10.282 10.421
4.924 4.932 4.950 5.880	7.905 7.923 8.221 8.222	9.435 9.455 9.548 9.657	10.450 10.467 10.500
6.302 6.364 6.662 6.742	8.230 8.259 8.266 8.407	9.681 9.744 9.778 9.832	10.502 10.813 10.858
6.910 6.920 6.972 6.975	8.423 8.442 8.503 8.640	9.883 9.890 9.919 9.966	11.029 11.033 11.078
6.990 7.047 7.103 7.156	8.683 8.795 8.853 8.897	9.986 10.040 10.079	11.347 11.605 11.899
7.232 7.283 7.347 7.442	8.986 9.016 9.037 9.049	10.120 10.162 10.169	12.667 12.842 12.867
7.513 7.526 7.536 7.584	9.069 9.127 9.139 9.299	10.169 10.238 10.239	12.913

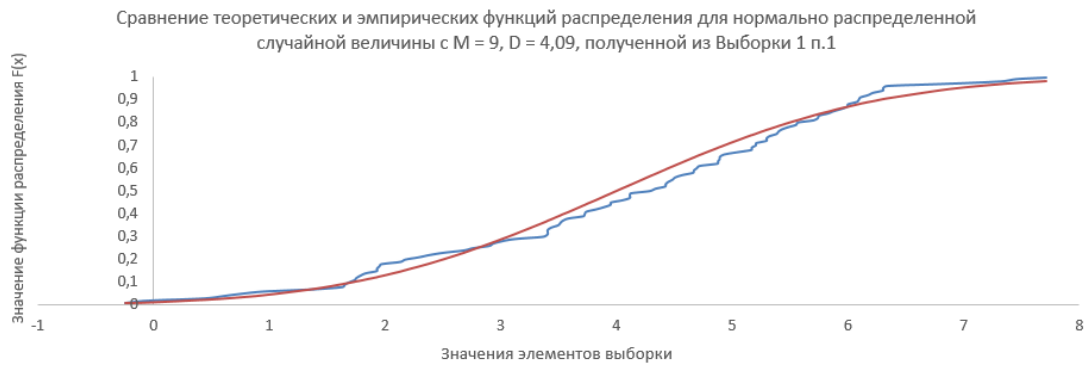
Выборочное среднее - 8.784; теоретическое значение - 8;

Выборочная дисперсия - 3.521; теоретическое значение - 4.09;

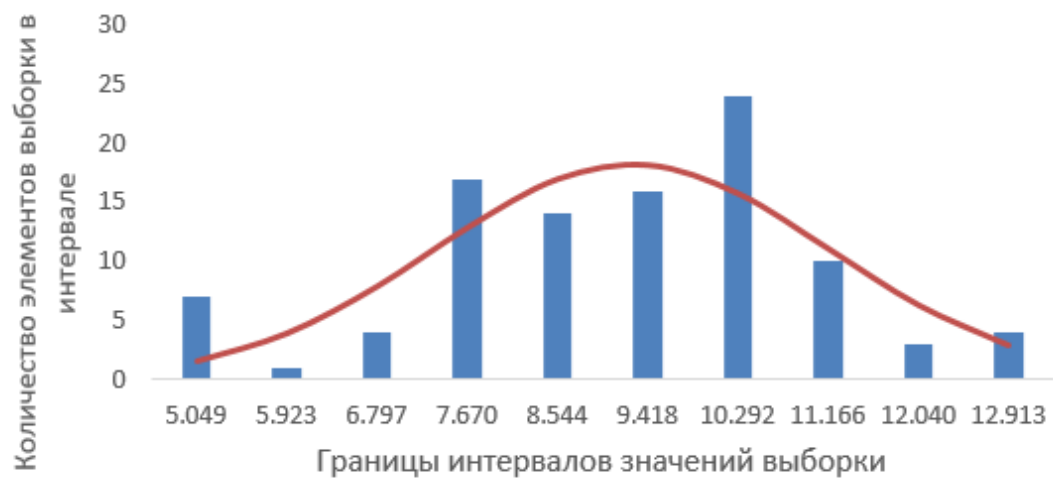
Исправленная выборочная дисперсия - 3.557; теоретическое значение - 4.09;

Медиана - 9.026; теоретическое значение - 8.

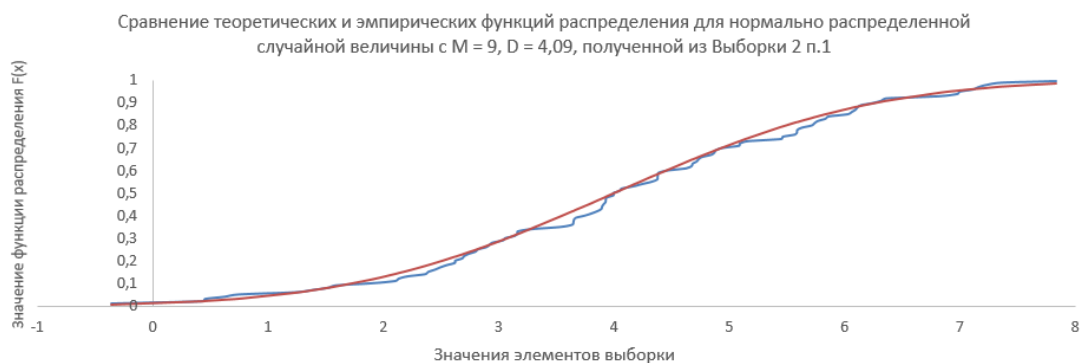
Графики для выборки 1



Гистограмма для Выборки 2 нормально распределенных случайных величин



Графики для выборки 2



Гистограмма для Выборки 1 нормально распределенных случайных величин



Выводы

На графиках выше можно увидеть, что полученные выборки достаточно хорошо приближены к теоретическим значениям нормально распределенной случайной величины.

Характеристики для полученных выборок отличаются от теоретических значений на величину порядка 0.1.

АЛГОРИТМ РАБОТЫ ПРОГРАММЫ

В моей программе было написано 4 функции, вычисляющие основные оценки выборок по следующим формулам:

1. Выборочное среднее: $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$

2. Выборочная дисперсия: $s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$

3. Исправленная выборочная дисперсия: $\tilde{s}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$

4. Медиана:

$$m = \begin{cases} x_{(k)}, & n = 2k \\ \frac{x_{(k)} + x_{(k+1)}}{2}, & n = 2k + 1, \end{cases}$$

где n - объем выборки x

При выполнении первого задания вычисляются две реализации выборок равномерно распределенных случайных величин по формулам, представленным в пункте Постановка задачи для Задания 1. Далее они копируются в два новых массива и сортируются в порядке возрастания. После с помощью отдельных функций вычисляются основные оценки этих выборок.

При выполнении второго задания моделируются испытания Бернулли по формулам, представленным в пункте Выполненная работа для Задания 2, используя массивы неотсортированных реализаций выборок из п. 1. Далее вычисляются частоты успехов по формуле, представленной в пункт Постановка задачи для Задания 2.

При выполнении третьего задания вычисляются выборки показательно распределенных случайных величин с различными параметрами по формуле, представленной в пункте Выполненная работа для Задания 3, используя массивы отсортированных реализаций выборок из п. 1. Далее области значений выборок разбиваются на 10 равных интервалов, и вычисляется количество элементов выборок, попавших в каждый из этих интервалов. После с помощью отдельных функций вычисляются основные оценки этих выборок.

При выполнении четвертого задания вычисляются выборки нормально распределенных случайных величин с параметрами $M = 4$ и $D = 3.16$ по формулам, представленным в пункте Постановка задачи для Задания 4, используя массивы неотсортированных реализаций выборок из п. 1. После полученные выборки сортируются в порядке возрастания. Далее области значений выборок разбиваются на 10 равных интервалов, и вычисляется количество элементов выборок, попавших в каждый из этих интервалов. После с помощью отдельных функций вычисляются основные оценки этих выборок.

В конце программы происходит вывод результатов в файл с названием result.txt.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Вадзинский Р.Н. Справочник по вероятностным распределениям. - СПб.: Наука, 2001, 295 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика (5-е изд.). М.: Высшая школа, 1977, 479 с.

ПРИЛОЖЕНИЕ

Листинг программы, написанной в ходе выполнения расчетного задания.

```
double calc_mo(vector<double> x) //Вычисление выборочного среднего
{
double mo = 0;
for (unsigned int i = 0; i < x.size(); i++) mo += x[i];
return mo /= x.size();
}

double calc_dispers(vector<double> x, double mo)//Вычисление выборочной
дисперсии
{
double dispers = 0;
for (unsigned int i = 0; i < x.size(); i++) dispers += (x[i] - mo)*(x[i] - mo);
return dispers /= x.size();
}

double calc_dispers_v(vector<double> x, double disp)//Вычисление исправленной
выборочной дисперсии
{
return disp = disp * x.size() / (x.size() - 1);
}

double calc_mediana(vector<double> x) //Вычисление медианы
{
if (!(x.size() % 2))
return((x[(int)(x.size() / 2)] + x[(int)(x.size() / 2) - 1]) / 2);
else return(x[(int)(x.size() / 2)]);
}

void main()
{
vector<double> x1, x2, x1_orig, x2_orig;
double x1_mo, x2_mo, x1_disp, x2_disp;
double x1_disp_v, x2_disp_v, x1_mediana, x2_mediana;
const int K = 87;
const double set = 0.243;
unsigned int i;

//////////ЗАДАНИЕ 1
x1.push_back((double)((int)(K * set * 1000) % 1000) / 1000);
x2.push_back((double)((int)((11 * set + pi) * 1000) % 1000) / 1000);
for (i = 1; i < 100; i++)
{
x1.push_back((double)((int)(K * x1[i - 1] * 1000) % 1000) / 1000);
x2.push_back((double)((int)((11 * x2[i - 1] + pi) * 1000) % 1000) / 1000);
}
```

```

x1_orig = x1;
x2_orig = x2;

sort(x1.begin(), x1.end());
sort(x2.begin(), x2.end());

x1_mo = calc_mo(x1);
x2_mo = calc_mo(x2);

x1_disp = calc_dispers(x1, x1_mo);
x2_disp = calc_dispers(x2, x2_mo);

x1_disp_v = calc_dispers_v(x1, x1_disp);
x2_disp_v = calc_dispers_v(x2, x2_disp);

x1_mediana = calc_mediana(x1);
x2_mediana = calc_mediana(x2);

//////////ЗАДАНИЕ 2
double bern1[100], bern2[100];
double freq1[10], freq2[10];
double mu1, mu2;
int m, j;
for (i = 0; i < 100; i++)
{
if (x1_orig[i] < set) bern1[i] = 1;
else bern1[i] = 0;
if (x2_orig[i] < set) bern2[i] = 1;
else bern2[i] = 0;
}
mu1 = 0.0;
mu2 = 0.0;
for (m = 10; m <= 100; m += 10)
{
for (j = m - 10; j < m; j++)
{
mu1 += bern1[j];
mu2 += bern2[j];
}
freq1[m / 10 - 1] = mu1 / m;
freq2[m / 10 - 1] = mu2 / m;
}

//////////ЗАДАНИЕ 3
const double a1 = 0.316, a2 = 2.16;
vector <double> expo1, expo2;

```

```

double ravn[100];
double exp1_mo, exp2_mo, exp1_disp, exp2_disp;
double exp1_disp_v, exp2_disp_v, exp1_mediana, exp2_mediana;
for (i = 0; i < 99; i++)
    ravn[i] = 0.01*(i + 1);
ravn[99] = 0.997;
for (i = 0; i < 100; i++)
{
    expo1.push_back(-1.0 / a1*log(1 - x2[i]));
    expo2.push_back(-1.0 / a2*log(1 - x2[i]));
}
double d1_exp, d2_exp, max_exp1, max_exp2;
vector<int> count_exp1, count_exp2;
d1_exp = expo1[99] / 10.0;
d2_exp = expo2[99] / 10.0;
i = 0;
int k = 0, p = 0, q = 0;
for (j = 1; j <= 10; j++)
{
    max_exp1 = d1_exp * j;
    max_exp2 = d2_exp * j;
    count_exp1.push_back(0);
    count_exp2.push_back(0);
    while (i < 100 && expo1[i] <= max_exp1 && expo1[i] > max_exp1 - d1_exp)
    {
        count_exp1[j - 1]++;
        i++;
    }
    while (k < 100 && expo2[k] <= max_exp2 && expo2[k] > max_exp2 - d2_exp)
    {
        count_exp2[j - 1]++;
        k++;
    }
}

exp1_mo = calc_mo(expo1);
exp2_mo = calc_mo(expo2);

exp1_disp = calc_dispers(expo1, exp1_mo);
exp2_disp = calc_dispers(expo2, exp2_mo);

exp1_disp_v = calc_dispers_v(expo1, exp1_disp);
exp2_disp_v = calc_dispers_v(expo2, exp2_disp);

exp1_mediana = calc_mediana(expo1);
exp2_mediana = calc_mediana(expo2);
print_file();
}

```