Содержание **Работа №5.** Задачи по составлению алгоритмов и структур данных с использованием двумерных массивов..

Цель работы: закрепить понятия алгоритмов, структур данных и соответствующие навыки программирования (написания псевдокода) на примере тестовых квадратов и гистограмм для числовых псевдослучайных последовательностей с заданным законом распределения.

Задачи - создать динамический документ со скриптами-псевдокодами на языке Javascript

1) показывающий тест-квадрат ;
2) формирующий и выводящую гистограмму и теоретическую кривую плотности распределения для визуального анализа последовательности;
3) со ссылкой на дополнительную страницу с ответами на контролльные вопросы.

Время выполнения: 2 часа.

1 Краткие теоретические сведения

Предполагается, что ко времени выполнения настоящей работы используемые математические понятия рассматривались по дисциплине “Прикладная математика”. Ну и, конечно, предыдущая работа №4 должна быть выполнена до выполнения настоящей работы, поскольку её результаты сейчас будут нужны.

В настоящей работе для моделирования случайных чисел используется так называемый метод обратных функций, который часто называют стандартным или универсальным методом моделирования случайных чисел. Согласно этому методу значение $\boldsymbol{x}$ случайной величины $\boldsymbol{X}$, является решением уравнения

$$F(x) = r \quad , \qquad (1)$$

где r – значение случайной величины R, равномерно распределённой на интервале [0, 1), $F(x)$ – функция распределения случайной величины X.

Легко доказывается, что решение этого уравнения можно записать в виде:

$$x = F^{-1}(r) \quad . \qquad (2)$$

Метод обратных функций можно понять и реализовать с помощью графиков, подобных изображённому на рис. 1.

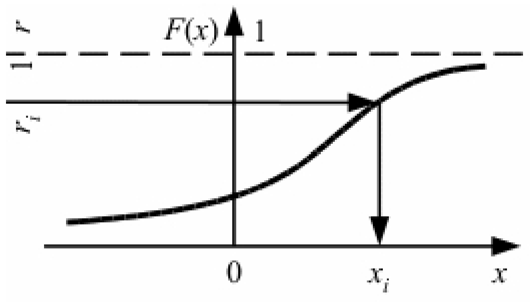


Рис. 1 Графический вариант метода обратных функций

Из рисунка видно, что, если по оси ординат взять случайное число от 0 до 1 и найти для него (показано стрелкой) число на оси абсцисс, то полученное значение будет значением случайной величины с функцией распределения *F(x)*.

На основании вышеизложенного можно предложить следующие этапы для реализации метода обратных функций:

1) получение случайной величины *R*, равномерно распределённой на интервале [0, 1);

2) вычисление по формуле (2) значений случайной величины *X*, имеющей распределение *F(x)*, удовлетворяющее уравнению (1).

Обращение функции распределения не всегда представляет собой простую задачу. Поэтому трудные для обращения функции, заменяют более легко обращаемыми аппроксимациями. Аппроксимации, в частности, часто используются для нормального закона распределения. Учитывая важность моделирования нормально распределённых случайных величин при имитационном моделировании, рассмотрим для этого распределения три таких аппроксимации.

Как известно, плотность вероятности нормально распределённой случайной величины ***x*** можно записать в виде

$$f(x) = \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\exp\left(-\frac{(x-a)^2}{2\cdot\sigma^2}\right), \quad -\infty < x < \infty \tag{3}$$

Интегральная функция распределения для стандартного нормального распределения (распределения Гаусса) имеет вид

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right) = \Phi(u) = \frac{1}{\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{x}\exp\left(-\frac{u^2}{2}\right)du, \tag{4}$$

где Ф(u) – интеграл вероятностей (гауссов интеграл) *u* – случайная величина, полученная преобразованием (говорят – стандартизацией или нормировкой) исходной случайной величины *x*

$$u = \frac{x-a}{\sigma}, \tag{5}$$

Где $\bar{x}$ - среднее арифметическое, σ – среднеквадратическое отклонение.

Учитывая свойство функции Ф(x) = 1 – Ф(-x) и формулу метода обратной функции (2), очевидно, что $Ф^{-1}(r) = -Ф^{-1}(1-r)$. Поэтому формула метода обратных функций для стандарной (с нулевым мат.ожиданием и единичной дисперсией) нормально распределённой случайной величины может иметь вид:

$$u_i = Ф^{-1}(r_i), \text{ если } 0.5 \le r'_i < 1; \tag{6}$$
$$u_i = -Ф^{-1}(r_i), \text{ если } 0 < r'_i < 0.5,$$

где r'_i – псевдослучайное число, имеющее такой же закон распределения, что и r_i. Эти числа не должны быть зависимыми. Другими словами, каждое из этих чисел должно быть получено в результате отдельного вызова генераторирующей их функции.

Обратную функцию $Ф^{-1}(r)$ заменяют аппроксимациями, например, такой:

$$Ф^{-1} = \frac{2.30753+0.27061\cdot\theta}{1+0.99229\cdot\theta+0.04481\cdot\theta^2}, \tag{7}$$

с ошибкой не превышающей $3 \cdot 10^{-3}$, или

$$Ф^{-1} = \theta - \frac{2.515517+0.802853\cdot\theta + 0.010328\cdot\theta^2}{1+1.432788\cdot\theta + 0.189269\cdot\theta^2 + 0.001308\cdot\theta^3}, \tag{8}$$

с ошибкой не превышающей $4.5 \cdot 10^{-4}$.

В формулах (7)-(8) величина θ равна $\theta(r_i)=\sqrt{-2\cdot\ln r_i}$, $0 \le r_i < 1$.

Можно, также использовать формулу, имеющую меньшую точность, но достаточную для имитационного моделирования аквакультур:

$$Ф_i^{-1} = \frac{1}{\sqrt{\frac{8}{\pi}}}\cdot\ln\frac{1+r_i}{1-r_i} \tag{9}$$

Итак, стандартную нормально распределённую величину u_i можно вычислить по формуле (6) с использованием (7), или (8), или (9).

Наконец, нужную нам случайную величину с заданными параметрами a и σ, которые в случае нормального распределения равны математическому ожиданию и среднеквадратическому отклонению, определим по формуле

$$x_i = a + \sigma \cdot u_i. \tag{10}$$

В таблице 1 приведены наиболее часто используемые функции плотностей распределения и обратные функции для них. Каждая строка этой таблицы соответствует варианту задания, равному двум последним цифрам номера зачётной книжки.

Моделируемые псевдослучайные последовательности должны быть подвергнуты статистической проверке. Для этого существует много специальных тестов, в том числе те, что использовались в предыдущей работе: тест-квадраты и автокорреляционные функции. Мы не будем их здесь рассматривать, поскольку соответствующие функции можно взять из предыдущей работы.

Одна из более тщательных проверок связана с расчётом интервального вариационного ряда, построением гистограммы и проверки соответствия гистограммы заданному закону распределения (критерий хи-квадрат Пирсона и другие).

В этой работе будет использован упрощённый визуальный метод проверки, при котором на гистограмму эмпирических плотностей вероятности будет наложена кривая соответствующей теоретической плотности распределения вероятности. При большом количестве значений последовательности (длине выборки) можно достаточно хорошо определить, подходит ли теоретический закон распределения для описания чисел генерируемой последовательности, визуально сопоставив график теоретического закона с гистограммой. Для визуального восприятия при большой длине выборки возможно удобнее анализировать ломанную линию, описываемую полигоном.

Если при неоднократных испытаниях (генерациях последовательности) огибающая гистограммы (ломанная полигона) колеблются вокруг теоретической кривой плотности, значит проверяемый генератор можно использовать для генерации псевдослучайных чисел с заданным законом распределения.

С точки зрения программирования в этой работе мы познакомимся со способами следующих задач программирования:

1) интервального вариационного ряда (или ряда частот), относительных частот и эмпирических плотностей распределения (с выдачей в виде удобной для восприятия таблицы);

2) эмпирической плотности распределения в виде гистограммы и полигона, совмещённых с кривой теоретического распределения (удобная для визуального восприятия графическая форма, умение создавать совмещённые графики с анимацией и масштабированием).

Предполагается, что теоретический материал по вычислению интервального вариационного ряда, и других интервальных характеристик, а также представление о гистограмме и полигоне усвоены к этому времени на занятиях по дисциплине “Прикладная математика”. Здесь же будет обращено внимание лишь на основные моменты:

1) количество значений генерируемой последовательности ***n*** должно быть большим

(рекомендуется $n \geq 2000$);

2) полученные значения последовательности следует округлить (обрезать) до нужного числа знаков после запятой (указано в примечании к табл.1);

3) для определения количества интервалов k, на которые разбивается весь диапазон значений последовательности, рекомендуется использовать формулу Старджеса (или Стёрджеса, Sturges, 1926)

$$k = \lfloor 1 + 3.322 \cdot lg(n) \rfloor \ , \tag{11}$$

где lg(n) – десятичный логарифм от числа значений последовательности,

или более простую формулу

$$k = \lfloor \sqrt{n} \rfloor \tag{11'}$$

где скобки – математический символ, означающий, что k необходимо округлять (обрезая) до целого числа.

Следует учитавать, что формулы (11) - (11') и другие (формул для вычисления k существует много), дают лишь приблизительное значение, на основе котрого исследователь должен принять решение. При небольшом (n < 100) числе наблюдений k, вычисляемое по всем формулам, различаются незначительно. При очень большом числе наблюдений (например, 2000 – это очень много) формулы выдают значения k различающиеся в разы. В этой работе в качестве k рекомендуется выбирать промежуточное значение между значениями, получаемыми по формулам (11) и (11') ;

4) шаг (или интервал группировки) вычисляется по формуле

$$h = (x_{max} - x_{min}) / k \tag{12}$$

где x_{max} , x_{min} – минимальное и максимальное значение среди чисел последовательности.

При этом, важно правильно округлить значение шага. При округлении шага можно использовать простое правило: после запятой надо оставлять на один знак больше, чем их содержится в числах последовательности;

5) в качестве левой границы первого интервала возьмём значение $a_0 = x_{min} - h / 2$. Тогда каждая последующая граница вычисляется прибавлением шага к предыдущей:

$$a_j = a_{j-1} + h, \quad j = 1, 2, \ldots, k+1, \tag{13}$$

при этом, средина первого интервала (первый класс) должна находиться в точке минимума исходной последовательности ($c_1 = (a_0 + a_1) / 2 = x_{min}$), а средина последнего – в точке максимума ($c_{k+1} = (a_k + a_{k+1}) / 2 = x_{max}$). Следует обратить внимание, что при таком способе расчёта интервалов (классов) их число будет на единицу больше первоначально задаваемого k;

6) подсчёт частот $\boldsymbol{m_j}$ попадания значений последовательности в интервалы выполняется в соответствии с формулой:

$$m_j = \sum_{i=1}^{n} c_i\ ; \quad c_i = 1, \text{при} \quad a_{j-1} \le x_i < a_j; \quad c_i = 0, \text{при} \quad a_{j-1} > x_i \ge a_j \tag{14}$$

где j = 1, 2, …, k+1; при этом не лишней будет проверка правильности подсчёта частот: сумма всех частот должна быть равна количеству значений в последовательности;

7) затем вычисляются относительные частоты или частости

$$w_j = m_j/n \tag{15}$$

и плотности частот или эмпирические плотности распределения вероятностей

$$v_j = w_j/h\ , \tag{16}$$

при этом следует проверить правильность выполнения расчётов: сумма всех частостей при правильном округлении должна быть равна единице. Очевидно, что этому же условию должна удовлетворять сумма всех величин ($v_j \cdot$ h), то есть площадь под гистограммой, (претендующей на эмпирический закон распределения вероятностей) должна быть равна единице.

8) поверх гистограммы следует поместить график теоретической функции, описывающей плотность распределения, чтобы можно было визуально оценить степень соответствия эмпирической и теоретической плотностей распределения. При этом, в качестве параметров теоретической функции следует брать статистические оценки (статистики), вычисляемые по значениям последовательности. Например, для вариантов с нормальным распределением параметр $\boldsymbol{a}$ приравнивается среднему арифметическому

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j\ , \tag{17}$$

а второй параметр $\boldsymbol{\sigma}$ – среднеквадратическому отклонению:

$$\sigma_x = \sqrt{D_x} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^2}\ , \tag{18}$$

где D_x — дисперсия.

При большом количестве значений последовательности вычисляемые значения статистик должны быть примерно равны задаваемым параметрам. Например, в случае рассматриваемого варианта с нормальным распределением: $\bar{x} \approx a$, $\sigma_x \approx \sigma$. Этот факт можно использовать для контроля правильности вычислений.

Далее будут приведены некоторые разъяснения по алгоритму построения интервального вариационного ряда:

1) упорядочиваем по возрастанию (ранжируем) исходную последовательность (исходный одномерный массив), допустим массив ***y***. Для этого удобно использовать метод *sort()* объекта *Array* с анонимной (не имеющей имени) функцией в качестве параметра, например,

```
y.sort(function(a,b){return a-b;})
```

В этом примере сортировка двух очередных рядом стоящих значений массива будет осуществляться по правилу, предусмотренному для метода `sort()` (см. http://javascript.ru/Array/sort). Значения отсортированного массива будут присвоены переменным *a* и *b* и функция должна выдавать: отрицательное значение, если первый аргумент меньше второго аргумента; нуль, если аргументы равны; положительное значение, если первый аргумент больше второго аргумента;

2) вычисляем по формулам (3.9.11) или (3.9.11') (или задаём промежуточное) целое количество интервалов *k*;

3) вычисляем шаг (длину интервалов), используя первое (минимум) и последнее (максимум) значение отсортированного массива:

```
var h = Math.round(((z[n-1]-z[0])/k)*l)/l;
```

где l = 1000, и тогда h округляется до тысячных в соответствии с условием задания, оставляя на один знак больше после запятой, чем в исходной последовательности;

4) создаём одномерный массив границ интервалов, например, так:

```
a[0]=Math.round((z[0]-h/2)*l)/l;
for (var j=0;j<=k;j++){
    a[j+1]=Math.round((a[j]+h)*l)/l;
}
```

5) создаём одномерный массив классов (средин интервалов):

```
   cl[0]=z[0];
for (var j=0;j<k;j++){
    cl[j+1]=Math.round((cl[j]+h)*l)/l;
}
```

6) создаём одномерный массив частот:

```
var i_0 = 0;
for (j= 1; j<= k+1; j++) {
    m [j-1] = 0;
    for (var i= i_0; i<= (n-1); i++) {
        if (z[i] < a[j]) {m [j-1] = m [j-1] + 1; i_0 = i+1;}
    }
}
```

7) создаём одномерные массивы относительных частот и эмпирических вероятностей :

```
for (var j= 0; j<= k; j++) {
    w [j] =  Math.round(m [j]/n * l) / l;
    p[j] =  Math.round(w [j]/h * l) / l;
}
```

8) если все массивы, содержащие выше перечисленные значения интервальных рядов, вычислялись в функции, то результат, возвращаемый из этой функции, можно передать в виде объекта, например, так:

```
var o_int_var_ser= {"a_int":a,"cl_int":cl,"m_int":m,"w_int":w,"p_int":p};
return o_int_var_ser;
```

2 Задания

Задание в целом состоит в том, что каждому студенту необходимо написать программу в виде динамического документа со скриптами на языке Javascript, на которой поэтапно, по событию, создаваемому пользователем (например, по клику мышкой), решались бы следующие задачи:

1) последовательность случайных чисел с законом распределения в соответствии с вариантом задания (выбирается из таблицы 1), полученная в предыдущей работе;

2) построение тест-квадрата;

3) построение эмпирических плотностей вероятности в графическом виде: в виде совмещённых графиков гистограммы, полигона и кривой теоретического распределения;

4) закрепление полученных знаний и навыков путём ответов на вопросы для самоконтроля с оформлением воросов-ответов в виде отдельной страницы.

Таблица 1 Варианты задания *)

№ варианта **)	Заданный закон плотности распределения	Формула для вычисления последовательности, имеющей заданный закон распределения ***)
00	Экспоненциальный (показательный): $f(x) = \lambda \cdot \exp(-\lambda \cdot x)$; $x \geq 0, \lambda > 0$	$x_i = -\frac{1}{\lambda} \cdot \ln(r_i)$
01	Лапласа (двусторонний экспоненциальный): $f(x) = \frac{1}{2 \cdot b} \cdot \exp(-\|x-a\|/b)$, $-\infty<x<\infty$, $b>0$	$x_i = a + b \cdot \ln(2 \cdot r_i)$, если r_i - (0, 0.5]; $x_i = a - b \cdot \ln(2 \cdot (1 - r_i))$, если r_i - (0.5,1)
02	Рэлея: $f(x) = \frac{x}{\sigma^2} \cdot \exp(-x^2/(2 \cdot \sigma^2))$; $x \geq 0$, $\sigma > 0$	$x_i = \sigma \cdot \sqrt{-2 \cdot \ln(r_i)}$
03	Вейбулла: $f(x) = \frac{c \cdot x^{c-1}}{b^c} \cdot \exp(-(\frac{x}{b})^c)$; $0 \leq x < \infty$; $c>0$; $b>0$; b, c - целые	$x_i = b \cdot (-ln(r_i))^{1/c}$
04	Коши: $f(x) = \frac{1}{\pi \cdot b \cdot (1+(y-a)^2/b^2)}$, $b > 0$	$x_i = b \cdot \tan(\pi \cdot (r - \frac{1}{2})) + a$
05	Степенной: $f(x) = a \cdot x^{a-1}$, $x \in (0, 1)$, $a > 0$	$x_i = r_i^{\frac{1}{a}}$
06	Логистический: $f(x) = \frac{1}{2 \cdot b} \cdot \exp(-\frac{x-a}{b}) \cdot (1+\exp(-\frac{x-a}{b}))^{-2}$	$x_i = a + b \cdot \ln(\frac{r_i}{1-r_i})$
07	Логарифмически нормальный (логнормальный): $f(x) = \frac{1}{x \cdot \sigma \cdot \sqrt{(2 \cdot \pi)}} \cdot \exp(-\frac{(\ln x - \ln a)^2}{2 \cdot \sigma^2})$ для $x > 0$; $\sigma > 0$	$x_i = \exp(z_i)$, где z_i - гауссовская случайная величина с параметрами (a, σ^2) (см. вариант 09)
08	Логистический (близкий к стандартному нормальному): $f(x) = \frac{\pi}{\sqrt{3}} \cdot \exp(-\frac{\pi \cdot x}{\sqrt{3}}) \cdot (1+\exp(-\frac{\pi \cdot x}{\sqrt{3}}))^{-2}$	$x_i = \sqrt{\frac{3}{\pi}} \cdot \ln(\frac{r_i}{1-r_i})$
09	Нормальный (Гаусса): $f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp(-\frac{(x-a)^2}{2 \cdot \sigma^2})$	По формуле (3.9.6) с учётом аппроксимации (3.9.7)

Продолжение табл. 1

№ варианта **)	Заданный закон плотности распределения	Формула для вычисления последовательности, имеющей заданный закон распределения ***)
10	Нормальный (Гаусса): $f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot \sigma^2}\right)$	По формуле (3.9.6) с учётом аппроксимации (3.9.9)
11	Степенной: $f(x) = a \cdot (1 - x)^{a-1}$, $x \in (0, 1)$, $a > 0$	$x_i = 1-(1-r)_i^{\frac{1}{a}}$
12	Экспоненциальный (показательный): $f(x) = \frac{1}{a} \cdot \exp\left(-\frac{x}{a}\right)$; $x \geq 0, a > 0$	$x_i = -a \cdot \ln(r_i)$
13	Лапласа (двусторонний экспоненциальный): $f(x) = \frac{1}{2 \cdot a} \cdot \exp\left(-\frac{\lvert x-m \rvert}{a}\right)$, $a > 0$, $-\infty < m < \infty$	$x_i = m + a \cdot \ln(2 \cdot r_i)$, если r_i (0, ½] $x_i = m - a \cdot \ln(2 \cdot (1 - r_i))$, если r_i (½,1)
14	Логистический: $f(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}$	$x_i = \ln\left(\frac{r_i}{1-r_i}\right)$
15	Вейбулла: $f(x) = \frac{k}{\upsilon-\varepsilon} \cdot \left(\frac{x-\varepsilon}{\upsilon-\varepsilon}\right)^{k-1} \cdot \exp\left(-\left(\frac{x-\varepsilon}{\upsilon-\varepsilon}\right)^k\right)$, для $x \geq \varepsilon$, $f(x) = 0$, для $x < \varepsilon$.	$x_i = \varepsilon + (-\ln(r_i))^{\frac{1}{k}}$
16	Логарифмически нормальный (логнормальный): $f(x)=\frac{1}{(x-a) \cdot \sigma \cdot \sqrt{(2 \cdot \pi)}} \cdot \exp\left(-\frac{(\ln(x-a)+A)^2}{2 \cdot \sigma^2}\right)$, для $x > a$; $\sigma > 0$	$x_i = a + \exp(\sigma \cdot z_i - A)$, где z_i - гауссовская случайная величина с параметрами (a, σ^2) (см. вариант 09)
17	Логистический: $f(x) = \exp\left(-\frac{x-a}{b}\right) \cdot \left(1+\exp\left(-\frac{x-a}{b}\right)\right)^{-2}$	$x = a + b \cdot \ln\left(\frac{r}{1-r}\right)$
18	Логистический (близкий к нормальному): $f(x) = \frac{\pi \cdot \exp\left(-\frac{\pi}{\sqrt{3}} \cdot \left(\frac{x-m}{\sigma}\right)\right)}{\sigma \cdot \sqrt{3} \cdot \left(1+\exp\left(-\frac{\pi \cdot (x-m)}{\sigma \cdot \sqrt{3}}\right)\right)^2}$	$x_i = m+b \cdot \sqrt{\frac{3}{\pi}} \cdot \ln\left(\frac{r_i}{1-r_i}\right)$
19	Нормальный (Гаусса): $f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot \sigma^2}\right)$	По формуле (3.9.6) с учётом аппроксимации (3.9.8)

*) числа последовательности округлить до сотых (два знака после десятичной запятой);

**) вариант должен совпадать с последними двумя цифрами зачётной книжки;

***) псевдослучайные числа r_i берутся из предыдущей работы (по своему варианту).

3 Методика выполнения

Из методически нового, что нужно сделать в этой работе, того, что не делалось ранее, - это построение гистограммы. Программу для построения гистограммы не сложно составить самостоятельно. Но зачем делать то, что неоднократно делалось другими. Можно, например, использовать программу, помещённую в качестве демонстрации использования библиотеки Protovis: http://mbostock.github.io/protovis/ex/histogram.html . Заодно полезно будет научиться пользоваться этой библиотекой. У неё есть, одно преимущество перед той же Highcharts — проще использовать, не нужна дополнительная библиотека jQuery.

Если же падёт выбор на библиотеку Highcharts, то можно поучиться программировать построение гистограммы с её помощью . О библиотеке Highcharts уже многое должно быть известно из предыдущей работы и лекций. Здесь, на примере гистограммы, будет продемонстрировано основное преимущество этой библиотеки перед другими графическими библиотеками — возможности анимации и зуммирования (масштабирования) графиков. К преимуществам можно также отнести автоматизацию построения диаграмм (графиков, гистограмм и др.), когда элементы диаграмм (оси, легенды, подписи, анимации) строятся по умолчанию - на основании анализа исходных данных без необходимости явного указания характеристик этих элементов.

Ну и, конечно, нужно будет вспомнить основные понятия о билиотеке jQuery, те, которые вы должны были усвоить из предыдущей лабораторной работы, поскольку в библиотеке Highcharts используется библиотека jQuery. Необходимо усвоить хотя бы минимум знаний о библиотеке, например, "Введение в jQuery", здесь: http://jquery.page2page.ru/ , или подраздел 4.5 в конспекте лекций на pvn.ho.ua .

Рассмотрим расположенный ниже листинг с кодом функции histogram(), предназначенной для построения гистограммы.

Листинг 1

```
/*****************************************************************************************
histogram(o_interv, st, l) -
                                                          .
    o_interv -                      -                              :
        a_int -                    ; cl_int -          (                  );
        m_int -                      ;
        w_int -          (                    );
        p_int -                      (                );
    st -                                                                   :
        st[0] -    .          , st[1] -                    ;
    l -                                           (                     )
*****************************************************************************************/
function histogram(o_interv, st, l) {
var l = Number(l)+1; l = Math.pow(10, l);
var y = new Array();
var x = new Array();
var p = new Array();
//
//                                        .
var k = 1/(st[1] * Math.sqrt(2*Math.PI));
for (var i = 0; i < o_interv.cl_int.length; i++) {
    p[i] = k * Math.exp(-Math.pow((o_interv.cl_int[i]-st[0]),2)/(2*Math.pow(st[1],2)));
    p[i] = Math.round(p[i] * l) / l;
}
//
//
for (var i = 0; i < o_interv.p_int.length; i++) {
    y[i] = Math.round(o_interv.p_int[i] * l) / l;
    x[i] = Math.round(o_interv.cl_int[i] * l) / l;
}
//                      ,
//           Highcharts:
```

Продолжение листинга 1

```
$(function () {
        $('#histogram').highcharts({
            chart: {
                margin: [ 100, 50, 100, 80],
                zoomType: 'xy' //
            },
            title: {
                text: '                                          '
            },
            subtitle: {
            text: '(                                                   )'
            },
            xAxis: [{
                categories: x,
            }],
            yAxis: {
                min: 0,
                title: {
                    text: '          (      /          )'
                }
            },
            legend: {
                enabled: true
            },
            series: [{
               type: 'column',
                  name: '                                    ',
                  data: y,
               color: '#33a',
              borderColor: '#EBBA95',
              borderWidth: 3,
                   dataLabels: { //
                        enabled: true,
                        rotation: 0,
                        color: '#33a',
                        align: 'center',
                        x: 0,
                        y: -5,
                        style: {
                        fontSize: '11px',
                        fontFamily: 'Verdana, sans-serif'
                        }
                      }
                }, {
                  type: 'line',
                   name: '                                ',
                   data: y,
              color: '#3a3'
            }, {
               type: 'spline',
                     name: '                        ',
                    data: p,
                 color: 'red'
            }
        ]
        });
    });
}
```

Построение гистограммы начинается со строк кода:

```
//                    ,
```

```
// Highcharts:
$(function () {
        $('#histogram').highcharts({...})
}
```

С этих строк, собственно, и начинаются графические построения, и с них мы ниже продолжим рассмотрение примера. А до этих строк формируются массивы ***p***, ***y*** и ***x***, в которые заносятся соответственно теоретические плотности, эмпирические плотности (плотности частот) и классы (значения середин интервалов). Эти массивы содержат всю информацию, необходимую для построения гистограммы, полигона и кривой теоретической плотности вероятности.

Приведённые выше строки кода похожи на те, которые мы обычно использовали при построении графиков. С их помощью ищется тэг с идентификатором histogram и к нему добавляется метод highcharts({...}) библиотеки Highcharts, выполняющий отображение всех элементов гистограммы. Метод выполняется с помощью анонимной функции (то есть, функции, не имеющей имени). Элементы гистограммы задаются в том объекте, который мы обозначили многоточием ({**…** }). Разберём свойства-объекты, вложенные в этот объект:

свойству chart в качестве параметров задаются отступы с использованием свойства margin, с заданием ему значений в виде массива (margin: [100, 50, 100, 80]) и зуммированием по обеим направлениям (zoomType: 'xy'). Тип графика здесь не указан, как для графика генерируемой последовательности. Поскольку в этом случае на диаграмме должно быть нанесено (совмещено) несколько графиков различных типов, то тип будет указан при задании данных ниже (см. ниже задание свойств series);

title и subtitle задаются также, как и при формировании графика генерируемой последовательности;

свойству xAxis задаются параметры, формирующие ось асцисс (ось x). Для формирования шкалы этой оси указываются “категории”, в качестве которых задаются классы (см. выше - в массив x заносится массив классов);

свойству yAxis задаются параметры, формирующие ось ординат (ось y). Для этой оси через запятую указывается два свойства: начальный отсчёт (min: 0) и надпись (title: {text: ' (/)'}). Поскольку не указаны значения для расчёта шкалы этой оси, то расчёт шкалы выполнится по умолчанию. Шкала будет одной для всех трёх типов графиков (см. ниже задание свойств series);

legend: {enabled: true} **–** задаёт отображение легенды (false – легенда не будет отображаться). В данном случае текст и обозначения не указаны, поэтому к качестве текста и обозначения легенды будет использованы соответствующие значения из описания данных (series, см. ниже);

свойству series задаются параметры для вычерчивания трёх графиков: гистограммы, полигона и теоретической плотности. Параметры задаются в виде трёх элементов массива, а каждый элемент массива является объектом. Свойства каждого объекта задают элементы каждого графика;

для первого графика в свойстве type указано значение column, определяющее в графика (type: 'column'). Это означает, что значения по оси y на этом графике будут отображаться в виде колонок, как обычно выглядит гистограмма. Свойство name: ' ' будет использовано при отображении легенды. Свойством data: y задаётся массив эмпирических плотностей, которые будут отображены заданным типом графика. О назначении остальных свойств не трудно разобраться самостоятельно;

второй график будет отображён в виде отрезков прямых, соединяющих заданные значения (type: 'line'). В качестве значений также, как и для первого графика задан массив эмпирических плотностей (data: y), но отображаться он будет не в виде гистограммы, а в виде полигона;

третий график будет отображён в виде отрезков полиномов третьей степени (сплайнов), соединяющих заданные значения (type: 'spline'). В качестве значений задан массив теоретических плотностей (data: p). Отображаться он будет не в виде гладкой кривой соединяющей заданные значения плотности. При большом количестве задаваемых значений плотности (при большом количестве классов) третий график будет аппроксимировать теоретическую плотность с точностью, достаточной для визуального восприятия (трудно отличимой от кривой Гаусса).

4 Рекомендации по обработке и оформлению полученных результатов

Оформление документа с лабораторной работой необходимо выполнить по примеру соответствующих страниц сайтов студентов прошлых лет (https://pvn.org.ru/www_JS/students/12kmk021/www/LabRabs/LabRab9/).

При загрузке документ может иметь следующий вид:

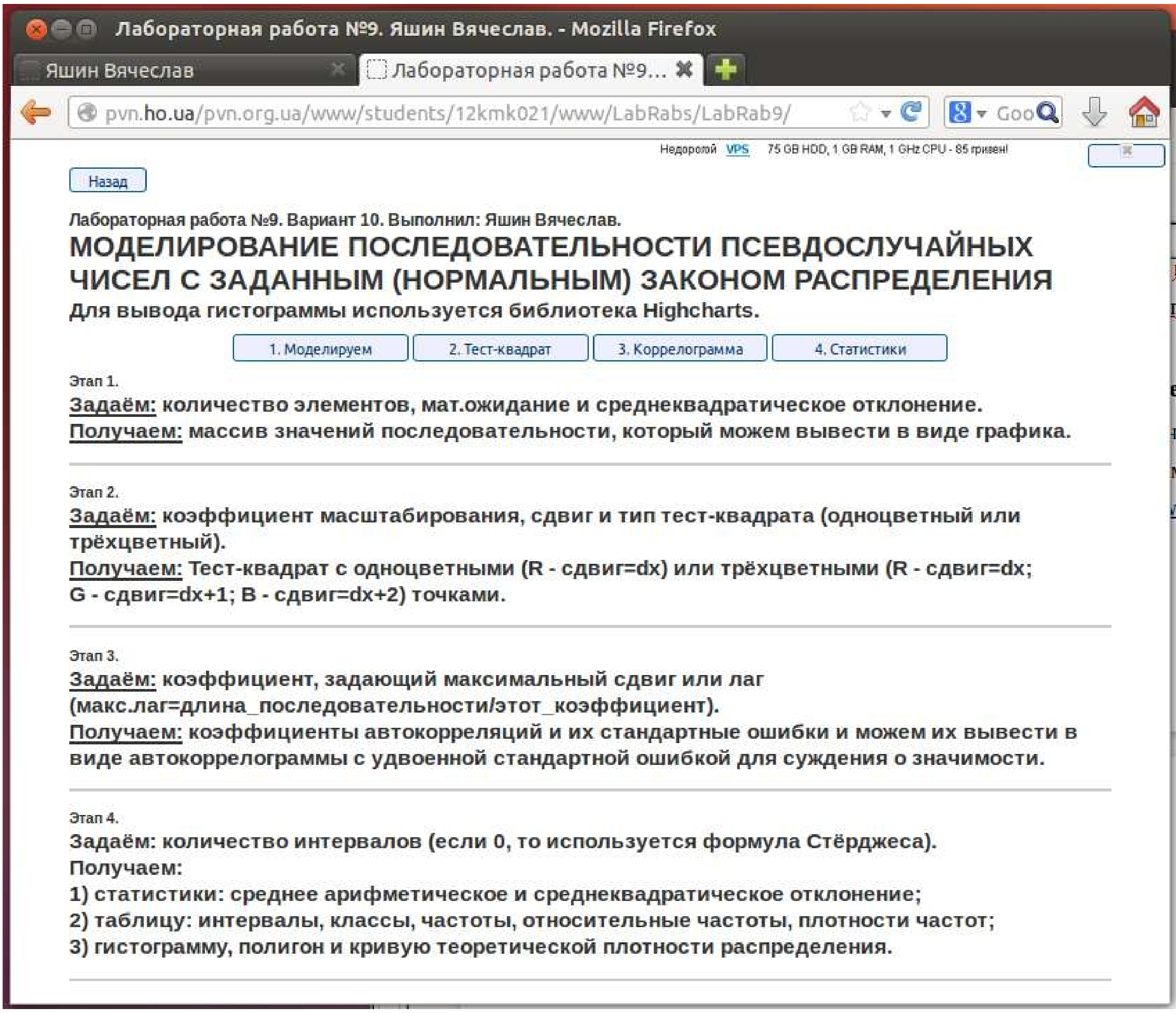


Рис. 2 Образец первоначального вида страницы

С помощью каждого из первых четырёх кнопок меню можно показать соответствующую форму, используя которые можно пройти этапы:

сгенерировать последовательность и вывести её в графическом виде;

сформировать и выести тест-квадрат;

рассчитать и вывести автокоррелограмму;

рассчитать и вывести статистики: среднее арифметическое и среднеквадратическое отклонение и вариационный ряд в виде таблицы и в виде гистограммы с полигоном и кривой теоретического распределения.

Первые три этапа должны быть понятны из предыдущей лабораторной работы. В рассматриваемом примере эти этапы сокращены: не предусмотрен вывод результатов в числовом виде. Рекомендуется просмотреть пример в действии и повторить примерно то же самое на своём варианте. Хотя, конечно, структура страницы и её дизайн могут отвечать предпочтениям каждого. Форма первого этапа появляется по кнопке меню «Моделируем» и имеет следующий вид:

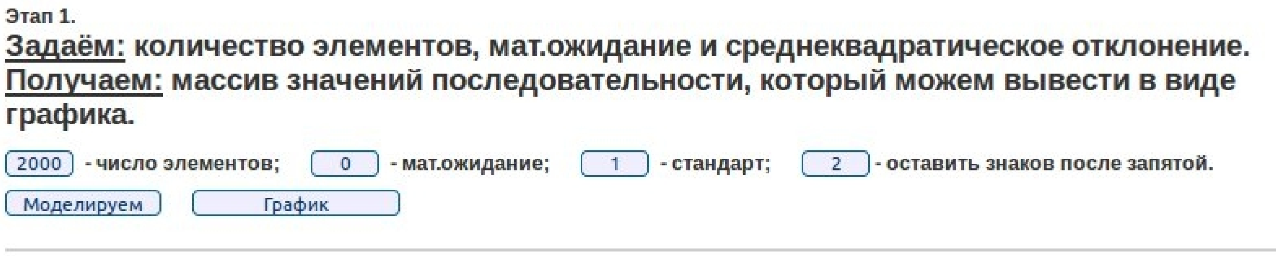


Рис. 3 Образец формы, появляющейся по меню «Моделируем»

Как видно из рисунка, в этой форме предусмотрено две кнопки: «Моделируем» и «График». По кнопке «Моделируем» в соответствии с вариантом 10 генерируется последовательность из 2-х тысяч нормально распределённых псевдослучайных чисел с нулевым математическим ожиданием и единичным среднеквадратическим отклонением. Числа округляются до двух знаков после десятичной точки. Такое количество чисел (2000) генерируется и впоследствии обрабатывается и графически отображается достаточно быстро (за 1-2 секунды) даже на самых маломощных компьютерах (и планшетах), и, в то же время, последовательность такой длины является достаточно представительной выборкой (хорошо представляет генеральную совокупность). Кроме того, такого количества чисел обычно достаточно для моделирования, которым мы будем заниматься на старших курсах.

По кнопке «График» предусмотрен вывод сгенерированной последовательности в графическом виде, показанном на рисунке ниже.

Рис. 4 Образец графика, построенного по значениям смоделированной последовательности

По виду графика можно предположить, что последовательность сгенерирована в соответствии с вариантом 10 и представляет собой выборку из генеральной совокупности чисел, имеющих стандартное нормальное распределение (почему? - попытайтесь объяснить сами). Это в какой-то мере подтверждается расположением точек на тест-квадрате, формируемом и строящемся с помощью формы на этапе 2, появляющейся по кнопке меню «Тест-квадрат».

Рис. 5 Форма и тест-квадрат сгенерированной последовательности

Но проследить, имеется ли зависимость между числами последовательности легче по коррелограмме, для расчёта и вывода графика которой предусмотрена соответствующая форма этапа3. Ниже показан результат, получаемый по меню «Коррелограмма».

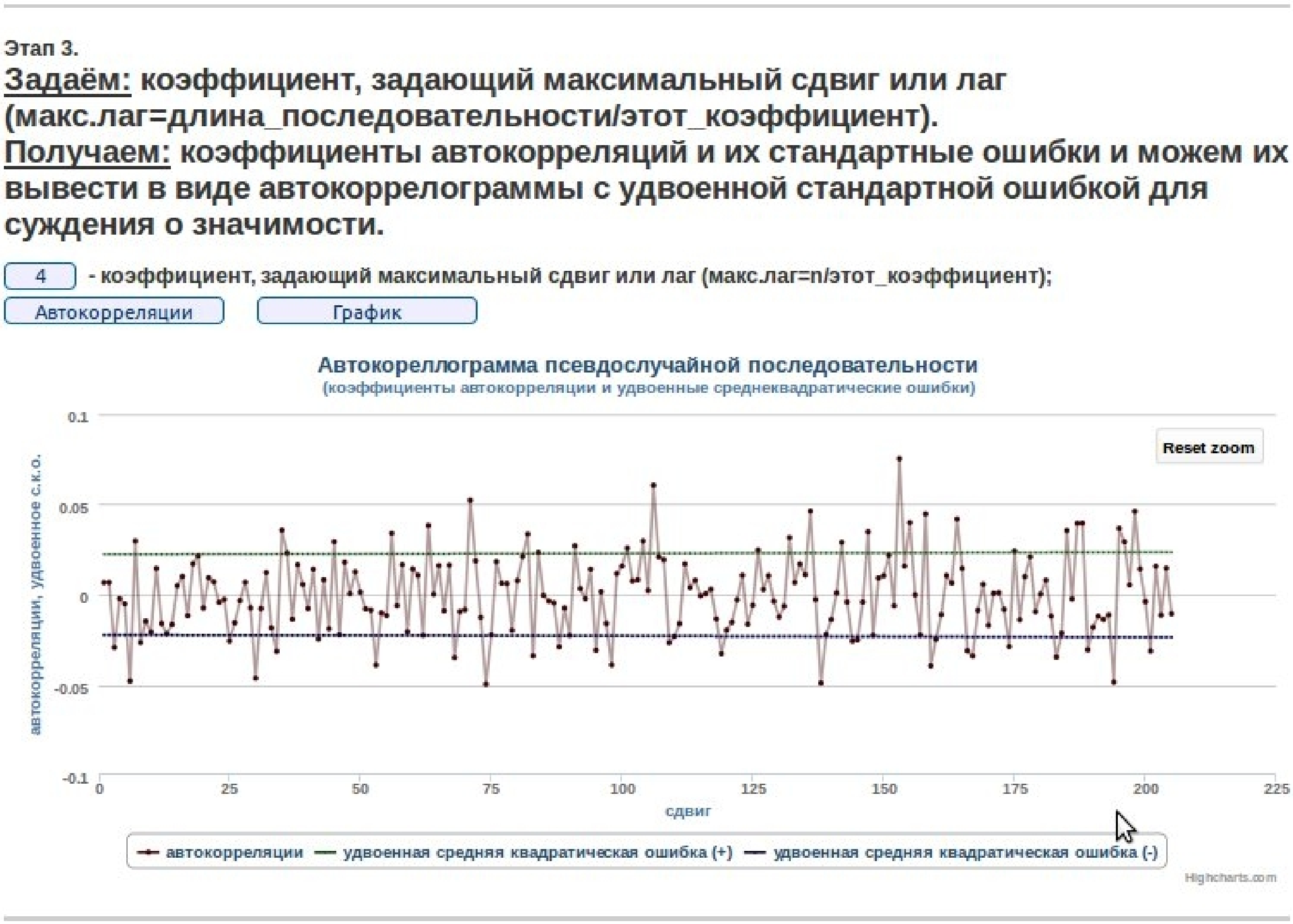


Рис. 6 Форма и фрагмент автокоррелограммы сгенерированной последовательности (фрагмент вырезан с использованием опции «зуммирование»)

Анализ гистограммы является более подходящим (хотя предыдущие этапы тоже необходимы. Почему?) для проверки соответствия вида распределения генерируемой последовательности заданному теоретическому. Рассмотрим более подробно четвёртый этап примера. По меню «Статистики» появляется следующая форма

Рис. 7 Образец формы, появляющейся по меню «Статистика»

Кнопка «Статистики» предназначена для расчёта и вывода среднего арифметического и среднеквадратического отклонения. Кнопка «Вариац.ряд» - для расчёта и вывода таблицы с интервальными значениями. Кнопка «Гистограмма» предназначена для расчёта и вывода гистограммы, полигона и кривой теоретической плотности. Результаты, полученные по первым двум кнопкам показаны на рисунке ниже.

Среднее арифметическое: -0.02
Стандартное отклонение: 1.11

Интервалы	Границы интервалов	Классы	Частоты	Относ.частоты	Эмпир.вероятности
1	-5.386 ÷ -4.873	-5.130	1	0.001	0.001
2	-4.873 ÷ -4.360	-4.617	0	0.000	0.000
3	-4.360 ÷ -3.847	-4.104	2	0.001	0.002
4	-3.847 ÷ -3.334	-3.591	4	0.002	0.004
5	-3.334 ÷ -2.821	-3.078	8	0.004	0.008
6	-2.821 ÷ -2.308	-2.565	34	0.017	0.033
7	-2.308 ÷ -1.795	-2.052	72	0.036	0.070
8	-1.795 ÷ -1.282	-1.539	108	0.054	0.105
9	-1.282 ÷ -0.769	-1.026	231	0.116	0.225
10	-0.769 ÷ -0.256	-0.513	347	0.173	0.338
11	-0.256 ÷ 0.257	-0.000	427	0.213	0.416
12	0.257 ÷ 0.770	0.513	338	0.169	0.329
13	0.770 ÷ 1.283	1.026	215	0.107	0.210
14	1.283 ÷ 1.796	1.539	120	0.060	0.117
15	1.796 ÷ 2.309	2.052	52	0.026	0.051
16	2.309 ÷ 2.822	2.565	17	0.009	0.017
17	2.822 ÷ 3.335	3.078	14	0.007	0.014
18	3.335 ÷ 3.848	3.591	5	0.003	0.005
19	3.848 ÷ 4.361	4.104	2	0.001	0.002
20	4.361 ÷ 4.874	4.617	2	0.001	0.002
21	4.874 ÷ 5.387	5.130	1	0.001	0.001

Рис. 8 Образец статистик и таблицы с интервальными значениями, появляющейся по кнопке «Статистики»

Получаемые среднее и среднеквадратическое отклонение должны быть приблизительно равны соответствующим теоретическим значениям, заданным в форме этапа 1. В рассматриваемом примере это не совсем так для среднеквадратического отклонения из-за слишком грубого метода обратных функций по варианту 10 (можете сравнить этот результат с результатами, получаемыми студентами по двум другим более точным вариантам генерации псевдонормальной последовательности).

В таблице интервальных статистик рекомендуется добавить строку сумм, свидетельствующую о правильности выполненных расчётов: сумма частот дожна быть равна 2000; сумма относительных частот — единице; сумма эмпирических вероятностей, умноженная на ширину интервала (площадь под гистограммой) — единице.

По таблице визуально трудно судить о соответствии закона распределения заданному. Вид распределения хорошо проявляется на гистограмме и полигоне с совмещённой кривой теоретического распределения. О формах кривых теоретических законов распределения к моменту выполнения настоящей работы нужно знать из дисциплины «Приклаадная математика». Графики для различных законов имеются в учебниках по теории вероятностей. Много сайтов также могут помочь в этом.

Образец результата, получаемый по кнопке «Гистограмма», показан на рисунке ниже.

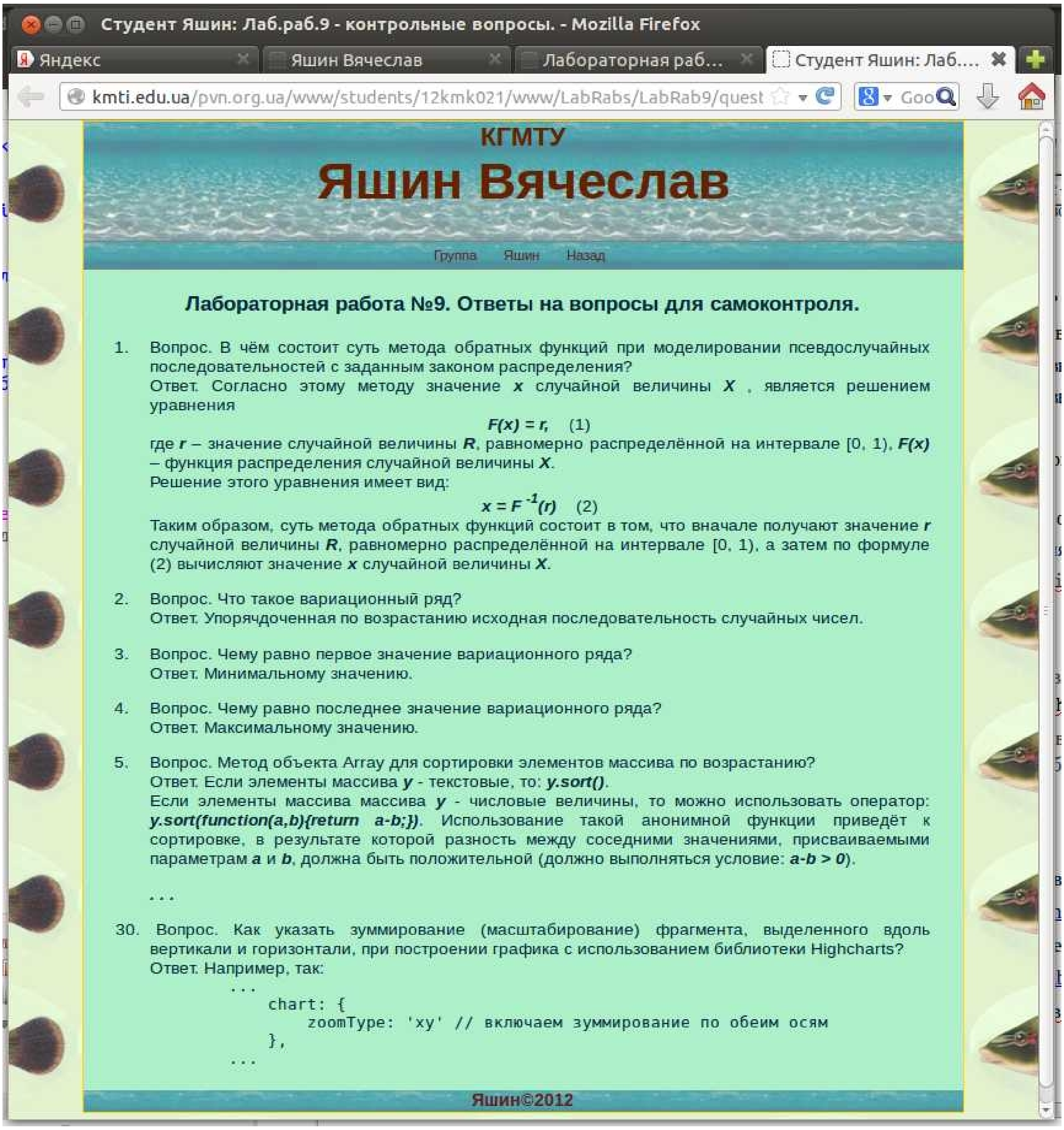
Рис. 9 Образец, получаемый по кнопке «Гистограмма»

Из рисунка видно, что гистограмма и полигон плотности эмпирического распределения хорошо соответствуют кривой плотности нормального распределения в соответствии с заданием по варианту 10. Хотя, при неоднократном повторении экспериментов, можно заметить закономерности, проявляющиеся в более высоких значениях для центральных интервалов и более низких — для интервалов, отстоящих примерно на четверть и три четверти длины всего рассматриваемого интервала (квартилей). Для более тщательного анализа гистограммы чрезвычайно полезной является опция зуммирования библиотеки Highcharts, использованная в рассматриваемом примере. На рисунке, показанном ниже, продемонстрирован фрагмент рисунка 9 после зуммирования. Показана также полезная возможность высвечивания значений *x* и *y* при подведении указателя мыши к нужной точке на графике.

Рис. 10 Фрагмент рисунка 9 после зуммирования

Кнопка «Reset zoom», предназначенная для возвращения рисунка к первоначальному виду, появляется автоматически после выполнения зуммирования.

По кнопке меню «Самоконтроль» в новом окне должна отображаться дополнительная страница с ответами на контрольные вопросы из пункта 5 настоящих методических указаний. Нумерация вопросов должна быть сохранена в точности такой, как в пункте .5. Вначале должен стоять вопрос и после него с новой строки — ответ. Образец оформления страницы с ответами приведён ниже.



Лабораторная работа №9. Ответы на вопросы для самоконтроля.

1. Вопрос. В чём состоит суть метода обратных функций при моделировании псевдослучайных последовательностей с заданным законом распределения?
Ответ. Согласно этому методу значение x случайной величины X, является решением уравнения
$$F(x) = r, \quad (1)$$
где r – значение случайной величины R, равномерно распределённой на интервале [0, 1), $F(x)$ – функция распределения случайной величины X.
Решение этого уравнения имеет вид:
$$x = F^{-1}(r) \quad (2)$$
Таким образом, суть метода обратных функций состоит в том, что вначале получают значение r случайной величины R, равномерно распределённой на интервале [0, 1), а затем по формуле (2) вычисляют значение x случайной величины X.

2. Вопрос. Что такое вариационный ряд?
Ответ. Упорядоченная по возрастанию исходная последовательность случайных чисел.

3. Вопрос. Чему равно первое значение вариационного ряда?
Ответ. Минимальному значению.

4. Вопрос. Чему равно последнее значение вариационного ряда?
Ответ. Максимальному значению.

5. Вопрос. Метод объекта Array для сортировки элементов массива по возрастанию?
Ответ. Если элементы массива ***y*** - текстовые, то: ***y.sort()***.
Если элементы массива ***y*** - числовые величины, то можно использовать оператор: ***y.sort(function(a,b){return a-b;})***. Использование такой анонимной функции приведёт к сортировке, в результате которой разность между соседними значениями, присваиваемыми параметрам ***a*** и ***b***, должна быть положительной (должно выполняться условие: ***a-b > 0***).

. . .

30. Вопрос. Как указать зуммирование (масштабирование) фрагмента, выделенного вдоль вертикали и горизонтали, при построении графика с использованием библиотеки Highcharts?
Ответ. Например, так:

```
...
    chart: {
        zoomType: 'xy' // включаем зуммирование по обеим осям
    },
...
```


Рис. 11 Образец оформления страницы с вопросами-ответами

Следует обратить внимание на то, как в рассматриваемом примере сделана ссылка на страницу с ответами на контрольные вопросы. По условию необходимо, чтобы по ссылке она открывалась в новом окне браузера. Ранее это делалось указанием значения `blank_` для атрибута `target` в тэге `<a... >`. Но в HTML5, который необходимо использовать в этой лабораторной работе, атрибут `target` отменён. Поэтому рекомендуется использовать либо стили CSS3, либо `Javascript`. В рассматриваемом примере выбран последний вариант. При установке режима для браузера «Блокировать всплывающие окна» страница с вопросами будет открываться в том же самом окне, что и исходная страница с лабораторной работой, заменяя её. Кнопка со ссылкой выглядит следующим образом:

```
<button onclick="return !window.open('questions.html')">
    5.
</button>
```

Для такого обработчика события Click функция window.open() возвратит значение true, если браузеру разрешено открывать всплывающие окна, и, наоборот, false, если открытие окон запрещено. Другими словами, если страница открылась в новом окне, то открытия этой же страницы в родном окне не происходит, так как будет return !true=false. Если же окно открыть не удалось, то будет return true, а это позволит ссылке отработать в штатном режиме, предусмотренном при запрещении браузеру открывать всплывающие окна (страница будет открыта в родном окне). Более подробно об этом можно прочитать здесь: http://www.xiper.net/collect/html-and-css-tricks/content/target-blank.html .

На странице с вопросами предусмотрена ссылка меню «Назад», построенная следующим образом: . Атрибуту href указано значение без наименования имени файла с лабораторной работой (index.html), так как в настройках web-сервера Apache, который мы используем, обычно указывается, что, если нет имени файла, то - это index.html. Если ссылка не срабатывает, то добавьте имя файла к значению атрибута href, или, лучше, научитесь настраивать web-сервер с помощью служебного файла .htaccess, размещаемого в том же директории, где находится файл index.html. Этот файл разрешается размещать на web-сервере самостоятельно владельцу сайта. О настройке web-сервера с помощью файла .htaccess можно прочитать, например, здесь: http://www.internet-technologies.ru/articles/article_1009.html .

3.9.5 Вопросы для самоконтроля

1. В чём состоит суть метода обратных функций при моделировании псевдослучайных последовательностей с заданным законом распределения?
2. Что такое вариационный ряд?
3. Чему равно первое значение вариационного ряда?
4. Чему равно последнее значение вариационного ряда?
5. Метод объекта Array для сортировки элементов массива по возрастанию?
6. Основные этапы построения интервального вариационного ряда?
7. Что вычисляется по формуле Стёрджеса?
8. Как проверить правильность построения последнего интервала?
9. Как вычисляется частота попадания случайной величины в интервал?
10. Как проверить правильность вычисления частот?
11. Что такое относительная частота?
12. Как проверить правильность вычисления относительных частот?
13. Как вычисляется плотность частот (эмпирическая плотность вероятности)?
14. Метод для поиска HTML-элемента с заданным идентификатором с помощью

библиотеки jQuery?

15. Что такое анонимная функция?
16. Можно ли пользоваться билиотекой Highcharts без подключения библиотеки jQuery?
17. Какие типы графиков билиотеки Highcharts вы использовали при выполнении заданий настоящей лабораторной работы?
18. Как создаётся объект при использовании фигурных скобок?
19. Как создаётся массив при использовании квадратных скобок?
20. Можно ли в качестве значения свойства объекта задать массив?
21. Можно ли в качестве значений свойств объекта задать массивы разных размеров?
22. Можно ли в качестве значений свойств объекта задать массивы разных мерностей?
23. Чему должна быть равна площадь под гистограммой?
24. Чему должна быть равна площадь под теоретической кривой плотности распределения вероятностей?
25. Может ли плотность распределения вероятностей принимать отрицательные значения?
26. Как строится полигон эмпирической плотности распределения вероятностей?
27. Для чего предназначена Javascript-библиотека Protovis?
28. Нужно ли для использования библиотеки Protovis дополнительно подсоединять библиотеку jQuery?
29. Как указать зуммирование (масштабирование) фрагмента, выделенного вдоль оси ординат, при построении графика с использованием библиотеки Highcharts?
30. Как указать зуммирование (масштабирование) фрагмента, выделенного вдоль вертикали и горизонтали, при построении графика с использованием библиотеки Highcharts?

6 Рекомендуемая литература:

Тараскин А.Ф. Статистическое моделирование и метод Монте-Карло: учебное пособие. Самар.гос.аэрокосм.ун-т. Самара, 1997. 62 с. (свободный доступ: http://bib.convdocs.org/

Алгоритмы моделирования случайных чисел с заданным законом распределения: http://ifour.spb.ru/www/lib/mst.pdf .

Ионнидис Я. История гистограмм (перевод Сергея Кузнецова: Yannis Ioannidis. The History of Histograms, Proceedings of 29th International Conference on Very Large Data Bases, September 9-12, 2003, Berlin, Germany (свободный доступ: http://citforum.ru/database/articles/histograms/)