1. Министерство образования и науки Российской Федерации
2. Санкт-Петербургский политехнический университет Петра Великого
3. —
4. Институт компьютерных наук и технологий
5. **Кафедра «Информационная безопасность компьютерных систем»**

**Лабораторная работа № 2**

1. по дисциплине «Теория вероятностей и математическая статистика»
2. Выполнила
3. студентк гр. 23508/4 Проценко М.А.

<*подпись*>

1. Проверила ассистент Лаврова Д.С.

<*подпись*>

Санкт-Петербург

1. 2016

**Формулировка задания 1**

## 1. Используя таблицу значений функции*f*(*x*), записать 100 цифр, выбирая из каждого значения функции второй знак справа(указать точность округления). C помощью критерия хи-квадрат проверить для такой выборки гипотезу о случайности цифр 0, 1, …, 9. Уровень значимости положить равным: а) 0,05; б) 0,01. Функцию *f*(*x*)брать в соответствии с вариантом.

**Решение:**

1. Была взята функция lnxв соответствии с вариантом.
2. Гипотеза H0: вторая цифра справа случайна.
3. Закон редких событий – распределение Пуассона: C:\Users\GP60\Desktop\Screenshot_13.png
4. Рассчитали количество встречаемости каждой цифры (0, 1,...,9)
5. Вычислили параметр а, который должен получиться равен среднему значению.
6. Вычислим Pi по формуле: C:\Users\GP60\Desktop\Screenshot_14.png
7. Затем вычислили n\*Pi
8. Далее посчитали (Ni - N\*pi)^2
9. Получившиеся значения разделим на N\*pi и сложим. Это и будет критерий хи-квадрат. У нас он получился равным 92.5 (наблюдаемое значение). Это больше чем 15.5 (хи-квадрат критическое при а=0.01) и больше чем 20.1 (хи-квадрат наблюдаемое при а=0.05)
10. Таким образом наша гипотеза не подтвердилась, отсюда делаем вывод, что вторая цифра в числе не случайна.

**Формулировка задания 2**

Взять три текста из разных областей знаний, объемом 2000 знаков (включая пробелы) каждый. На основе анализа частот встречаемости букв проверить гипотезу об однородности этих текстов.

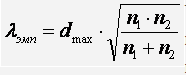
**Решение:**

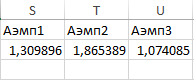
Были взяты три текста из разны областей знаний (текст из курса философии, статья про теннис из википедии, технический текст).

Для оценки однородности текстов был использован критерий Колмогорова-Смирнова.

Однако этим критерием можно сравнить только 2 выборки. Поэтому, так как у нас 3 текста, а следовательно и три выборки, нам придется применить этот критерий 3 раза (т.к. три пары выборок).

Вычисления:

1. Вычисляем относительные частоты f, равные частному от деления частот на объём выборки, для двух имеющихся выборок.
2. Далее определяем модуль разности соответствующих относительных частот для контрольной и экспериментальной выборок.
3. Среди полученных модулей разностей относительных частот выбираем наибольший модуль, который обозначается dmax.
4. Эмпирическое значение критерия λэмп определяется с помощью формулы: 
5. Чтобы сделать вывод о схожести по рассматриваемому критерию между двумя текстами, сравним экспериментальное значение критерия с его критическим значением, определяемым по специальной таблице, исходя из уровня значимости . В качестве нулевой гипотезы примем утверждение о том, что сравниваемые тексты незначительно отличаются друг от друга. При этом нулевую гипотезу следует принять в том случае, если наблюдаемое значение критерия не превосходит его критического значения.
6. По таблице определяем критическое значение критерия: λкр(0,05)=1,36.
7. Таким образом, λэмп<1,36= λкр только в первом и третьем случае. Следовательно, нулевая гипотеза принимается для текстов (1,2) и (2,3), и группы по рассмотренному признаку отличаются не существенно. Текста 1 и 3 отвергают гипотезу о том что тексты однородны и принимают конкурирующую гипотезу о том, что текста неоднородны.



**Формулировка задания № 3**

Построить критерий отношения правдоподобия для проверки двух простых гипотез о неизвестном параметре распределения из расчетных заданий № 1.3 и 1.4.

**3.1)** Для показательного распределения (задание 1.3) сформулировать гипотезы:

• для значения a = 0,134;

• для значения a = 3,13.

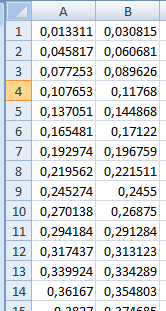
**3.2)** Для нормального распределения (задание 1.4) сформулировать гипотезы:

• для математического ожидания (M = 9);

• для дисперсии (σ 2 = 3,13).

**Выполнение задания № 3.1**

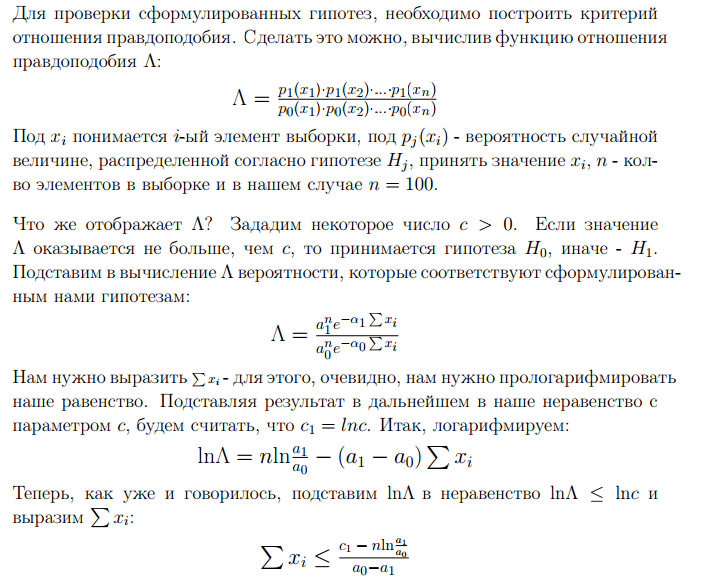
Для выполнения задания 3.1 из результатов пункта 3 первой лабораторной работы были извлечены данные о вариационных рядах выборок:

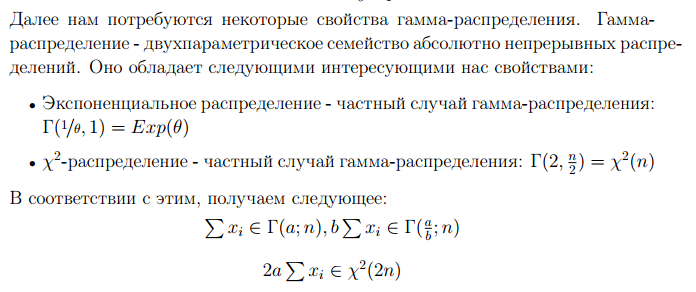
 И т.д. Всего по 100 значений в выборках.

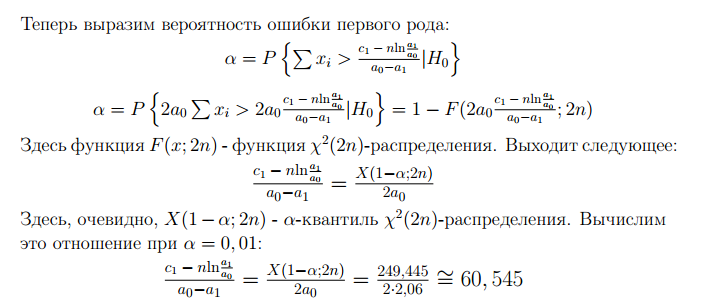
Были сформулированы две простые гипотезы:

• Гипотеза H0: элементы выборки имеют показательное распределение с параметром a0 = 3, 13 (т.е. xi ∈ exp(a0), a0 = 3, 13)

• Гипотеза H1: элементы выборки имеют показательное распределение с параметром a1 = 0, 134 (т.е. xi ∈ exp(a1), a1 = 0, 134)







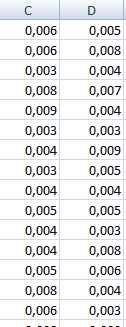
Посчитаем суммы первой и второй выборок. Для выборки 1 сумма равна 71.101, для выборки 2 сумма равна 69.923.

**Выводы по заданию № 3.1**

Т.к. сумма элементов выборок меньше, чем вычисленное нами отношение (60,545), принимается гипотеза H0. Результаты выполнения данного задания согласуются с результатами, полученными в ходе первой лабораторной работы.

**Выполнение задания № 3.2**

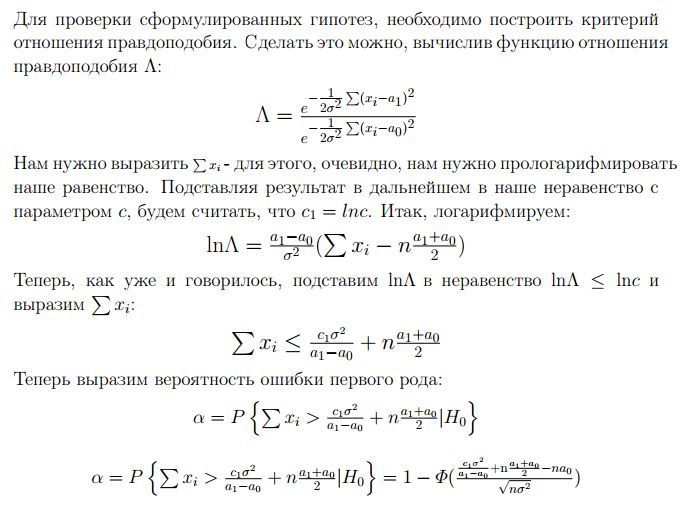
Для выполнения задания 3.2 из результатов пункта 4 первой лабораторной работы были извлечены данные о вариационном ряде следующей выборки:

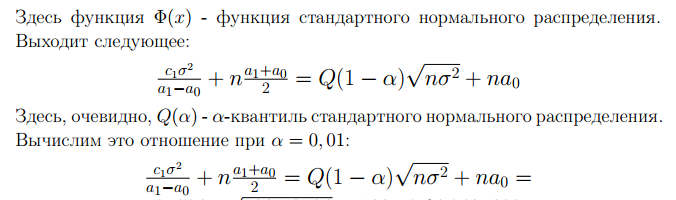
 И т.д.

Были сформулированы две простые гипотезы о мат.ожидании:

• Гипотеза H0: элементы выборки имеют нормальное распределение мат. ожиданием a0 = 6 и дисперсией σ 2 = 4, 136

• Гипотеза H1: элементы выборки имеют нормальное распределение мат. ожиданием a0 = 25 и дисперсией σ 2 = 4.136.





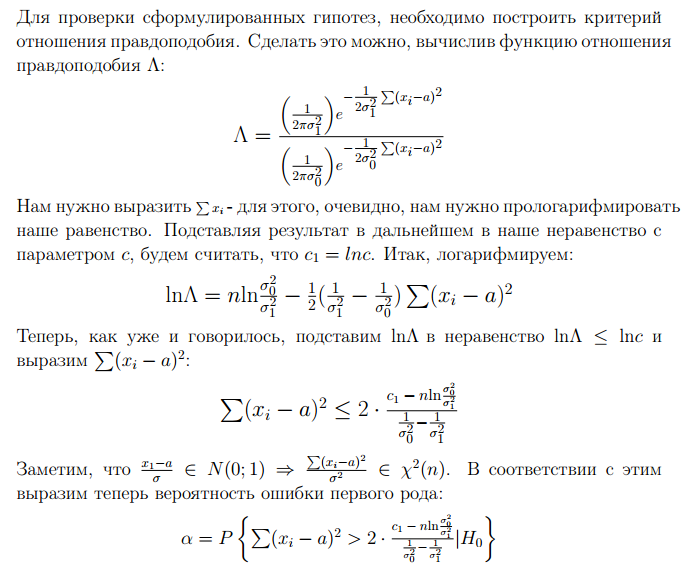
= 647.304

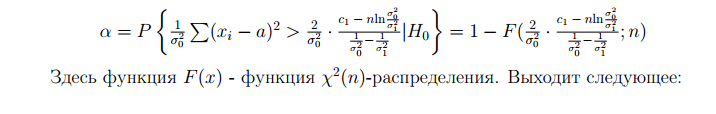
Сумма представленной выборки оказалась равно 0.561.

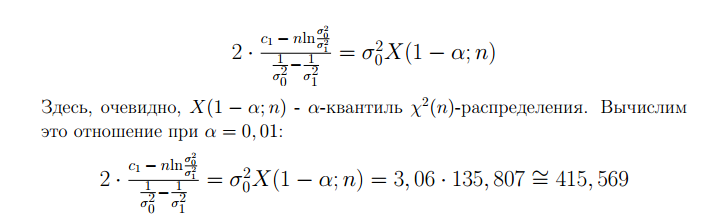
Используя ту же выборку были сформулированы две простые гипотезы о дисперсии:

• Гипотеза H0: элементы выборки имеют нормальное распределение мат. ожиданием a0 = 6 и дисперсией σ 2 = 4. 136.

• Гипотеза H1: элементы выборки имеют нормальное распределение мат. ожиданием a0 = 6 и дисперсией σ 2 = 25







Сумма квадратов (xi − a) равна 0.03.

**Выводы по заданию № 3.2**

Для выборки 1, имеем следующее: в случае с проверкой гипотез для мат. ожидания, сумма элементов выборки: Pxi = 0.561, а вычисленное ранее отношение, включающее квантиль стандартного нормального распределения: Q(1 − α) √ nσ2 + na0 = 2, 326 · √ 100 · 3, 06 + 100 · 9 ∼= 940, 688. Следовательно, принимаем гипотезу о мат. ожидании H0 при α = 0, 01. В случае проверки гипотез для дисперсии, P(xi − a) = 0.03, в то время как отношение, содержащее α-квантиль χ 2 (n)-распределения: σ 2 0X(1 − α; n) = 3, 06 · 135, 807 ∼= 415, 569. Следовательно, принимаем гипотезу о дисперсии H0 с заданным уровнем ошибки первого рода.