Министерство образования и науки Российской Федерации

Санкт-Петербургский политехнический университет Петра Великого

—

**Институт** **кибербезопасности и защиты информации**

Отчёт

по расчетному заданию

**Логнормальное распределение**

по дисциплине «Теория вероятностей и математическая статистика»

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнили: | студентка группы 4851001/00002 |  | К.В. Зима |
|  |  | (подпись, дата) |  |
|  | студент группы 4851001/00002 |  | Г.М. Дзюнов |
|  |  | (подпись, дата) |  |
|  |  |  |  |
|  |  |  |  |
| Проверил: | Профессор |  | Е.Б. Александрова |
|  |  | (подпись, дата) |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

Санкт-Петербург

2022

1. **Провести исследование распределения из варианта: история его появления, его назначение, применение в научных исследованиях**

История появления распределения

Логнормальное распределение в теории вероятностей — это двухпараметрическое семейство абсолютно непрерывных распределений. Изначально во многих отраслях было достаточно применять Гауссово распределение, однако многие измерения показывают более или менее асимметричное распределение. Особенно распространены асимметричные распределения, когда средние значения малы, а дисперсии велики и значения не могут быть отрицательными, как, например, в случае численности видов, длительности латентных периодов инфекционных заболеваний, распределения полезных ископаемых в земной коре. Такие асимметричные распределения часто близко соответствуют логарифмически нормальному распределению. Логарифмически нормальные распределения обычно характеризуются с точки зрения логарифмически преобразованной переменной с использованием в качестве параметров ожидаемого значения или среднего значения ее распределения и стандартного отклонения. Эта характеристика может быть полезной, поскольку по определению логарифмически нормальные распределения снова симметричны на логарифмическом уровне [1].

Назначение:

Нормальное или логарифмически нормальное распределение обычно используется при отсутствии подробной информации. Логнормальное распределение используется для описания переменных нагрузки, тогда как нормальное распределение используется для описания переменных сопротивления. Однако переменной, которая, как известно, никогда не принимает отрицательных значений, обычно назначается логарифмически нормальное распределение, а не нормальное распределение [2].

Применение в научных исследованиях:

Логнормальное распределение используется, например, при моделировании таких переменных, как доходы, допустимое отклонение от стандарта вредных веществ в продуктах питания и т.д. Так же при анализе поведения человека, например Длина комментариев, размещаемых на дискуссионных форумах в Интернете, подчиняется логарифмически нормальному распределению. В биологии и медицине используется для меры размеров жировой ткани, в размерах лавин разрывов в цитоскелете живых клеток показаны логарифмически нормальные распределения, причем в раковых клетках размеры значительно выше, чем в здоровых и так далее [3].

1. **Понятие эмпирической функции распределения**
   1. **Определить вид функции распределения для распределения из варианта**

Логнормальное распределение случайной величины X задается плотностью вероятности, имеющей вид

,

## где >0, >0, ∈ R

## Имеет вид при :

## :

## 

## Рисунок 1 – График функции логнормального распределения

## Построить эмпирическую функцию распределения для k случайных элементов выборки (k=10, 50, 100)

## 

## Рисунок 2 - Эмпирическая функция распределения для k=10

## Эмпирическая функция распределения для k=50:

## 

## Рисунок 3 - Эмпирическая функция распределения для k=50

## Эмпирическая функция распределения для k=100:

## 

## Рисунок 4 - Эмпирическая функция распределения для k=100

* 1. **Сравнить теоретическую и эмпирическую функции распределения на графике**

Выборочная (эмпирическая) функция распределения в [математической статистике](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0) — это приближение теоретической [функции распределения](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F), построенное с помощью выборки из него.

Рассмотрев графики, можно заметить, что эмпирическая функция повторяет форму теоретической функции распределения.

1. **Понятие гистограммы**
   1. **Определить вид функции плотности распределения для распределения из варианта**

Для того что бы определить плотность распределения необходимо взять интеграл от функции распределения. По полученной плотности был построен график плотности логнормального распределения. График представлен на рисунке 5.

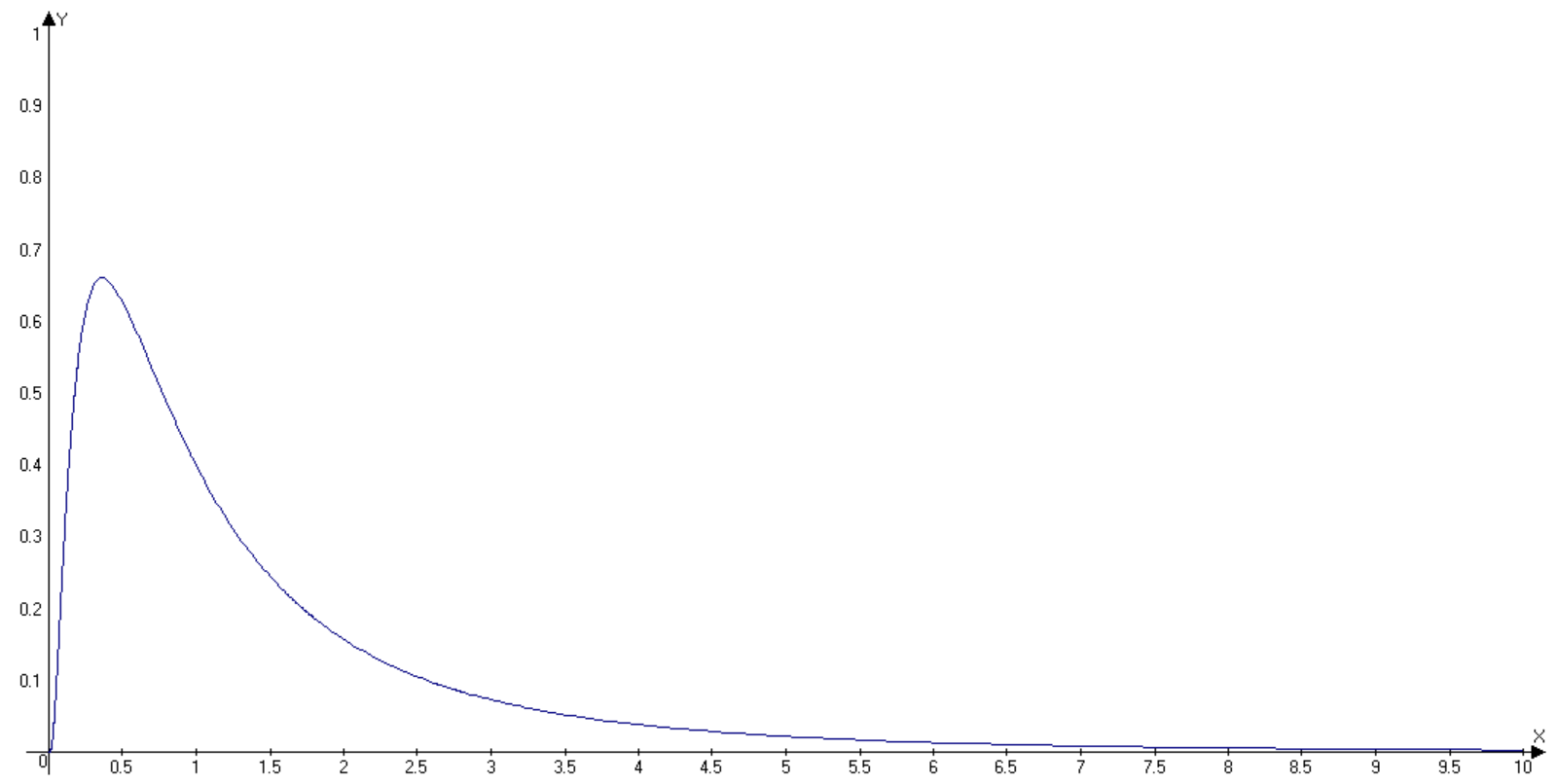


Рисунок 5 – График плотности логнормального распределения.

* 1. **Построить гистограмму (по относительным частотам!) для k случайных элементов выборки (k=10, 50, 100).**

Рисунок 6 – Гистограмма для k=10

Рисунок 7 - Гистограмма для k=50

Рисунок 8 – Гистограмма для k=100

* 1. **. Определить вид полигона относительных частот для k случайных элементов выборки (k=10, 50, 100)**

Рисунок 9 – Полигон для k=10

Рисунок 10 – Полигон для k=50

Рисунок 11 – Полигон для k=100

* 1. **Выводы**

Проанализировав график плотности логнормального распределения, можно заметить, что большая часть элементов входит в первые интервалы, а далее попадание в интервалы уменьшается, стремясь к единице. То же самое можно заметить на гистограммах для k=10, 50 и 100 элементов. Полигоны практически в точности повторяет форму гистограмм.

1. **Понятие точечных оценок**
   1. **Оценить параметры распределения выборки методом моментов**

Метод моментов — метод оценки неизвестных параметров распределений в математической статистике и эконометрике, основанный на предполагаемых свойствах моментов (Пирсон, 1894 г.). Идея метода заключается в замене истинных соотношений выборочными аналогами, т.е. выразить числовые параметры теоретического распределения через моменты распределения, оценённые по выборки. Число моментов должно соответствовать числу неизвестных параметров распределения (чаще всего используют первые два момента).Для каждого элемента была найдена относительная частота , затем с помощью относительных частот были рассчитаны математическое ожидание по формуле и дисперсия по формуле .

Далее необходимо было решить систему уравнений

; ;

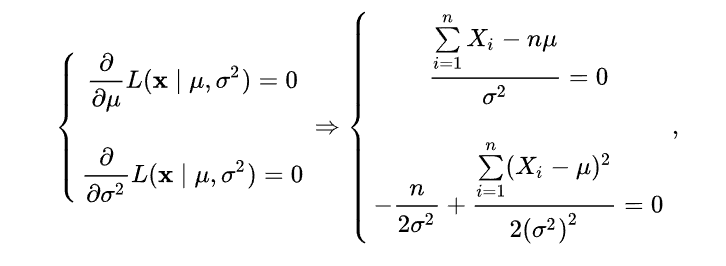
* 1. **Оценить параметры распределения выборки методом максимального правдоподобия**

**Метод максимального правдоподобия -** метод оценивания неизвестного параметра путём максимизации функции правдоподобия.

Для логнормального распределения функция правдоподобия имеет вид  
Изображение выглядит как текст

Автоматически созданное описание

Чтобы найти параметры возьмем частные производные от функции



Тогда из первого уравнения получается что /n=, следовательно = 2,52262E+21

Подставляя по вторую производную полученное значениеполучем =3,23\*1022

* 1. **Оценить параметры распределения выборки методом максимального правдоподобия. Выводы**

**Несмещенная оценка –** это точечная оценка, математическое ожидание которой равно оцениваемому параметру.

Для параметра несмещенной оценкой является среднее значение X, тогда получается, что оценка , полученная методом максимального правдоподобия является несмещенной.

Для параметра несмещенная оценка равняется = 1,05007E+45.

**Состоятельная оценка** – это точечная оценка, сходящаяся по вероятности к оцениваемому параметру.

Таким образом, по теореме Чебышева состоятельной оценкой является выборочное среднее, значит параметр , полученный методом максимального правдоподобия является состоятельной оценкой.

**Эффективная оценка** – это несмещенная оценка, имеющая наименьшую дисперсию из всех возможных несмещенных оценок данного параметра. Оценка из метода максимального правдоподобия является несмещенной. Проверим, является ли она эффективной. Дисперсия данной оценки равна . Приравняем ее к неравенству Рао-Крамера , где I()=. Получается, что оценка из метода максимального правдоподобия является эффективной.

Оценку, для которой в неравенстве Рао — Крамера достигается равенство, иногда называют **R-эффективной оценкой**. Исходя из доказательства эффективности оценки из метода максимального правдоподобия можно сказать, что она является R-эффективной.

Подводя итоги, можно сказать, что только оценка из метода максимального правдоподобия подходят все вышеперечисленные свойства.

1. **Понятие интервальных оценок**
   1. **. Оценить параметры распределения выборки с помощью интервальной оценки с уровнями доверия**

**Интервальное оценивание** — один из видов статистического оценивания, предполагающий построение интервала, в котором с некоторой вероятностью находится истинное значение оцениваемого параметра.

1. *Уровень доверия .*

Сначала оценим математическое ожидание: M(X)==2,52262E+21

Оценим дисперсию D(x)~s2==1,05007E+45

В данных условиях по теореме Фишера, величину имеет распределение хи-квадрат с n-1 степенью свободы. Для величины v1 входы таблицы r=n-1=163, =(1+)/2=0,925. Для v2 r=n-1=163, =(1-)/2=0,075. Тогда v1=189,66, v2=137,77.

Получается, что 137,77<<189,66, следовательно

9,02465E+44<<1,24237E+45

3,00411E+22<<3,52473E+22 тогда, по формуле дисперсии интервал для равен нулю.

1. *Уровень доверия*

M(x) и D(x) рассчитали в предыдущем пункте.

Для величины v1 входы таблицы r=n-1=163, =(1+)/2=0,975. Для v2 r=n-1=163, =(1-)/2=0,025. Тогда v1=200,24, v2=129,54.

Получается, что 129,54<<200,24, следовательно

8,54782E+44<< 1,3213E+45

2,92367E+22<<3,63497E+22 тогда, по формуле дисперсии интервал для равен нулю.

1. *Уровень доверия*

M(x) и D(x) рассчитали в пункте 1).

Для величины v1 входы таблицы r=n-1=163, =(1+)/2=0,995. Для v2 r=n-1=163, =(1-)/2=0,005. Тогда v1=213,25, v2=120,25.

Получается, что 120,25<<213,25, следовательно

8,02633E+44<< 1,42338E+45

2,83308E+22<<3,77277E+22 тогда, по формуле дисперсии интервал для равен нулю.

* 1. **Выводы**

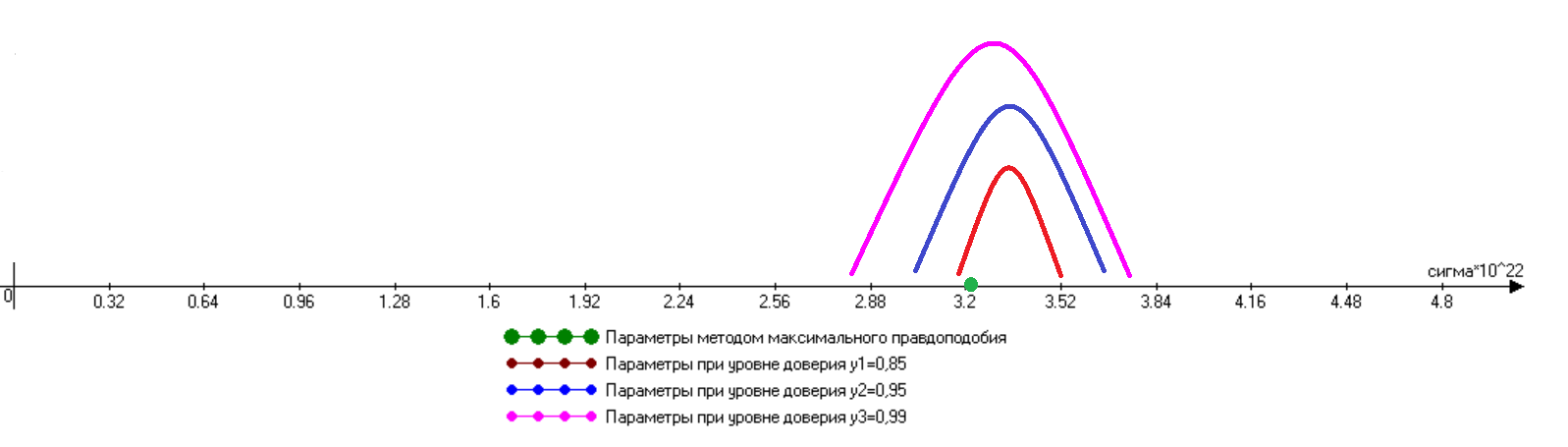


Рисунок 12 – Графическое соотношение интервальных и точечных оценок

Чем выше значение , тем шире доверительный интервал. Точечное значение находится в пределах границ доверительных интервалов, следовательно, чем выше уровень доверия тем выше процент попадания в значения исследуемой совокупности.

1. **Понятие статистических критериев**
   1. **Гипотезы о параметрах распределениях**

Метод моментов

,

Для построения наиболее мощного критерия воспользуемся леммой Неймана – Пирсона [4].

Для этого найдем совместную плотность распределения независимых случайных величин ξ1,…,ξn, каждая из которых имеет логнормальное распределение с параметрами , =1,05007E+45

=

ТогдаW0= или

Найдем константу λ из условия

Следовательно, и =+, где

Получается, что   
+

Для n=164 2,52262E+21, +2,53162E+21

Отклоняя нулевую гипотезу, с вероятностью 0,01 мы совершаем ошибку

При =4,54, тогда +=2, 64Е+21

Отклоняя нулевую гипотезу, с вероятностью 0,1 мы совершаем ошибку

Для метода максимального правдоподобия и интервального оценивания статисту критерия найдем таким же образом.

Тогда при уровне значимости равном 0,05 и 0,01 с вероятностью 0,1 и 0,4, соответственно, мы получаем ошибку при параметре, найденном методом максимального правдоподобия.

С параметрами полученные методом интервального оценивания мы получаем похожие значения.

* 1. **Гипотезы о виде распределения**

Для проверки гипотезы о виде распределения будем использовать критерий Колмогоров и критерий

Проверяем гипотезу H0 о том, что случайная величина x распределена по логнормальному закону с параметрами

При уровне значимости =0,05 найдем критическое значени с помощью функции из MS Excel ХИ2.ОБР.ПХ (вероятность; степень свободы). Критическое значение при =0,05 равно 9,49, при =0,01 равно 13,28.

Если наблюдаемое значение критерия попадет в область d0, то гипотеза о логнормальном распределении принимается, если наблюдаемое значение критерия попадет в область d1 (больше критической точки), то гипотеза отклоняется.

Построим вспомогательную таблицу, где С – интервалы, Z – стандартизованное значение для функции Лапласа, Ф – значения функции Лапласа, – функция . (Рисунок 13) Изображение выглядит как текст, стол

Автоматически созданное описание

Рисунок 13 – Вспомогательная таблица

Сумма функций больше, чем критическое значение при =0,05 и =0,01, следовательно гипотеза отвергается.

Для Критерия Колмогорова Max (\*10=4(). Табличные значения для λ:0,52 и 0,44 для уровней значимости 0.05 и 0.01 соответственно. Следовательно, при критерии Колмогорова гипотеза отвергается, так как превышено критическое значение.

* 1. **Гипотезы об однородности выборок**

В ходе работы было выяснено, что логнормальное распределение имеет общие корни с распределением Пуассона. Ниже представлены графики двух функций: логнормального распределения и распределения Пуассона (Рисунки 14 и 15).

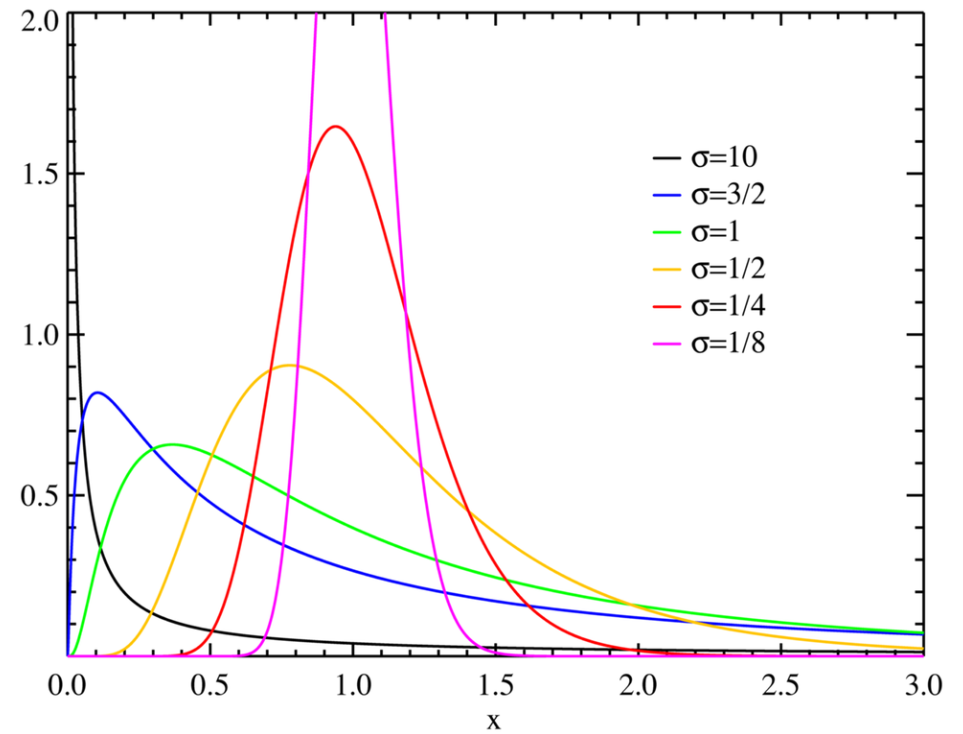


Рисунок 14 – График логнормального распределения при различных параметрах

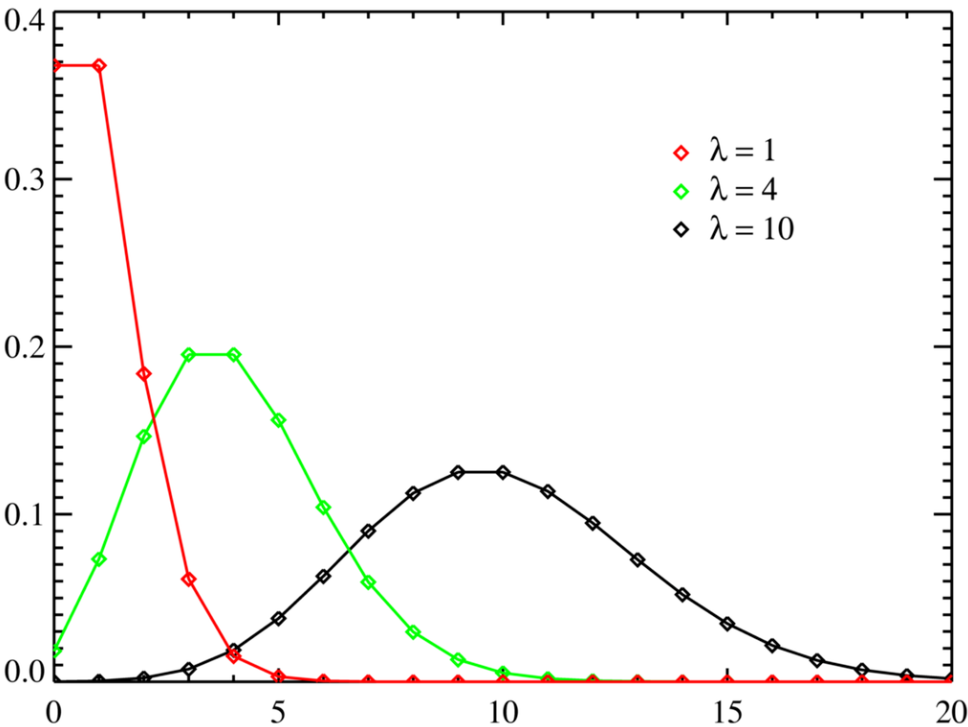


Рисунок 15 – График распределения Пуассона при различных параметрах

Методом обратного преобразования сгенерируем выборку.

Метод обратного преобразования – способ генерации случайных величин с заданной функцией распределения., путем модификации работы генератор равномерно распределенных чисел.

Рассмотрим функцию распределения интересующего нас закона распределения и попробуем выразить обратную функцию:

Прямая функция логнормального распределения:

Прямая функция распределения Пуассона:

При попытке получения значений для логнормального распределения возникли сложности. Обратная функция не может быть однозначно выражена, а потому генерация именно случайной выборки затруднена (в силу сложности работы с функцией знака). Таким образом, можно лишь предложить алгоритм оценки мощности трех-выборочного критерия, например, по Колмогорову-Смирнову.

При этом подходе статистика критерия представляет собой меру отклонения эмпирических распределений, соответствующих конкретным выборкам, от эмпирического распределения, построенного по совокупности анализируемых выборок. В k -выборочном варианте критерия Смирнова, можно рассмотреть статистику вида

В результате будет определена пара выборок, различие между которыми окажется наиболее значимым с позиций используемого k-выборочного критерия.

Из условия знаем, что k = 3.

Статистика критерия: ,

где – эмпирические функции выборок.

Уровень значимости:

Получается, что исследовать сходимость к предельным распределениям статистики выборочного критерия в данной ситуации также не представляется возможным. Оценить мощность критерия при разных объемах выборок () в данном случае мы также не можем.

1. **Источники информации**
2. Логарифмически нормальные распределения в разных науках: Ключи и подсказки: О прелестях статистики и о том, как механические модели, напоминающие игровые автоматы, дают ссылку на удобный способ характеристики логарифмически нормальных распределений, которые могут обеспечить более глубокое понимание изменчивости и вероятности - нормальные или логарифмически нормnальные: вот в чем вопрос. Экхард Лимперт, Вернер А. Шталь, Маркус Эббт. BioScience, Том 51, Выпуск 5, май 2001 года, Страницы 341-352. [Электронный ресурс] Режим доступа: <https://academic.oup.com/bioscience/article/51/5/341/243981?login=false>

### 2) [Анализ случайных величин и неопределенностей. Ен Бай, Вэй-Лян Цзинь, в книге "Проектирование морских конструкций" (Второе издание), 2016 г. 33.4. Выбор функций распределения.](https://www.sciencedirect.com/science/article/pii/B9780080999975000332) [Электронный ресурс] Режим доступа: <https://www.sciencedirect.com/topics/engineering/lognormal-distribution#:~:text=The%20lognormal%20distribution%20is%20used,rather%20than%20a%20normal%20distribution>

# 3) CFA - Логнормальное распределение вероятностей. [Электронный ресурс] Режим доступа: <https://fin-accounting.ru/cfa/l1/quantitative/cfa-lognormal-probability-distribution>

1. Лемма Неймана-Пирсона. [Электронный ресурс] Режим доступа: https://studfile.net/preview/9124450/page:4/