

УДК 004

**ПРОВЕРКА ГИПОТЕЗЫ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ПОСТУПЛЕНИЯ
ЗАПРОСОВ НА САЙТ МОРДОВСКОГО ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА им. Н.П. ОГАРЁВА****Варюхин Владимир Алексеевич**

студент

Лещанкин Константин Александрович

кандидат технических наук

Мордовский государственный университет им. Н.П. Огарёва, Саранск

author@apriori-journal.ru

Аннотация. В данной статье осуществляется проверка гипотезы о законе распределения, по которому поступают запросы на официальный сайт Мордовского государственного университета им. Н.П. Огарёва. На основе имеющихся данных, о количестве запросов, выводятся выборочные функции плотности и распределения. Подтверждение гипотезы осуществляется по критерию Пирсона.

Ключевые слова: закон распределения, проверка гипотезы, выборочная функция плотности, выборочная функция распределения, критерий Пирсона, MATLAB.

**CHECK OF THE HYPOTHESIS OF THE LAW OF DISTRIBUTION
OF RECEIPT OF REQUESTS FOR THE SITE
OF THE MORDOVIAN STATE UNIVERSITY****Varyukhin Vladimir Alekseevich**

student

Leshchankin Konstantin Aleksandrovich

candidate of technical sciences

Mordovian state university, Saransk

Abstract. In this article check of a hypothesis of the law of distribution under which requests for the official site of the Mordovian state university of arrive is carried out. On the basis of the available data, about amount of inquiries, sample functions of density and distribution are removed. Confirmation of a hypothesis is carried out by Pearson's criterion.

Key words: law of distribution, check of a hypothesis, sample function of density, sample function of distribution, Pearson's criterion, MATLAB.

В настоящее время Интернет занимает очень большое место, как у отдельных людей, так и в организациях и субъектах страны [1].

Различные сайты имеют различную популярность в интернете. Данную оценку можно получить, проанализировав количество запросов на данный ресурс [1].

Во многих случаях на практике встречаются ситуации, когда у исследователя есть основания предположить, что случайная величина подчиняется какому-либо закону распределения (нормальному, показательному, Гамма-распределению и т.д.).

Основной метод решения данной задачи – это построение вариационных рядов для больших совокупностей. Большую значимость для этого имеет правильное построение самой выборки: определения количества групп и размера интервала варьирующего признака случайной величины.

Говоря о характере, типе закона, имеется в виду отражение в нем общих условий, определяющих вариацию генеральной совокупности признака. Они познаются с помощью анализа сущности явления, тех его характеристик, которые и определяют вариацию изучаемого признака. Следовательно, должна быть выдвинута какая-то гипотеза, обосновывающая определенный тип теоретической кривой распределения [2].

Часть выборки о количестве поступающих в день требований на сайт Мордовского государственного университета им. Н.П. Огарёва представлена в табл. 1.

Таблица 1

Выборка запросов с 01.06.2015 по 23.10.2015

Day Index	Sessions
01.06.2015	2589
02.06.2015	2851
03.06.2015	2896
...	...
21.10.2015	2614
22.10.2015	2516
23.10.2015	479

На основе полученных данных из Интернет центра Мордовского госуниверситета, выполним расчет выборочных функций плотности и распределения.

Выборочная или эмпирическая функция распределения в математической статистике – это приближение теоретической функции распределения, построенное с помощью выборки из него.

Необходимо произвести определение наблюдаемых частот. Для вычисления количества подынтервалов воспользуемся формулой Стерджесса [3]:

$$\Delta = [1 + \log_2 N], \quad (1)$$

где N – размер исходного массива поступающих требований.

Произведя подсчет наблюдаемых частот, получим выборочную функцию плотности $f(x)$ как отношение определенных частот к сумме всех частот (рис. 1). Затем получим выборочную функцию распределения $F(x)$ как отношение частичных сумм выборочной функции плотности к сумме всех значений выборочной функции плотности (рис. 2).

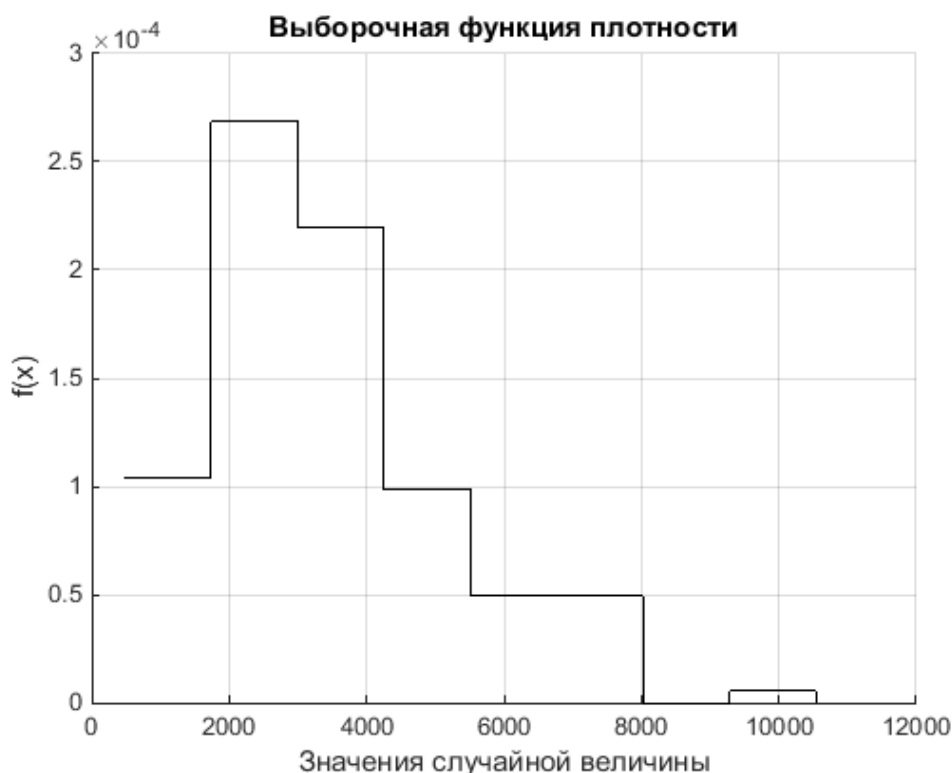


Рис. 1. Выборочная функция плотности

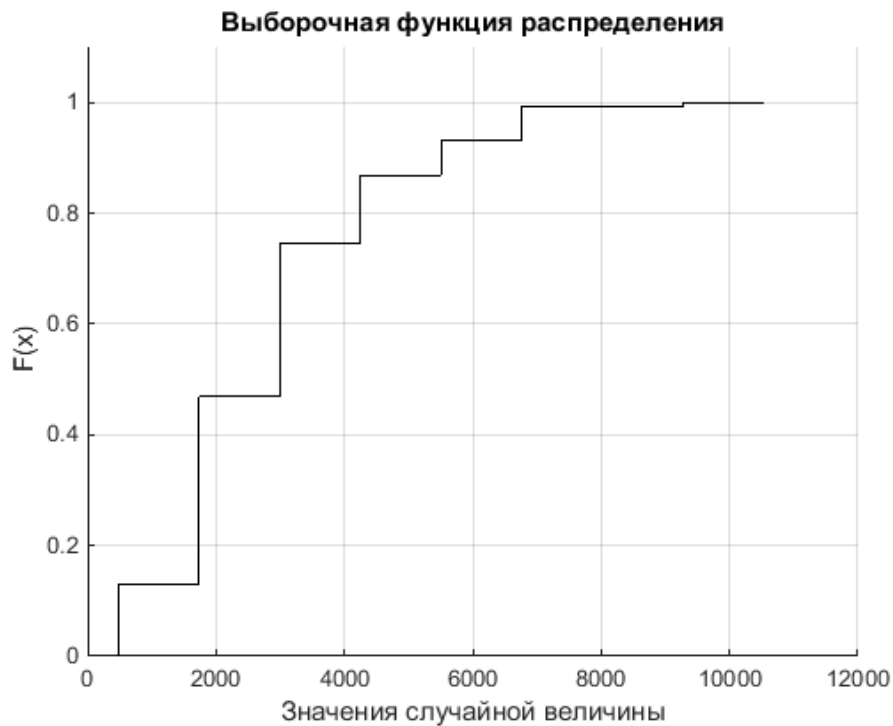


Рис. 2. Выборочная функция распределения

По исходным данным находим среднее арифметическое $M(x)$ и дисперсию $D(x)$: $M(x) = 3.4917e+003$, $D(x) = 2.7840e + 006$.

На основе анализа рис. 1 выдвигаем гипотезу, что выборочная функция плотности соответствует закону распределения потока Эрланга k -го порядка:

$$f_k(t) = \frac{\lambda (\lambda t)^k}{k!} e^{-\lambda t}, \quad (2)$$

$$F_k(t) = 1 - \sum_{j=0}^k \frac{(\lambda t)^j}{j!} e^{-\lambda t}, \quad (3)$$

Поставим задачу аппроксимацию данных распределением Эрланга k -го порядка [4; 5]. Для этого требуется проверить условие замены и последующим определением параметров закона [6]:

$$M^2(x) \geq D(x), \quad (4)$$

$$\lambda = \frac{M(x)}{D(x)}, \quad (5)$$

$$k = \lfloor \lambda M(x) - 1 \rfloor, \quad (6)$$

Произведем расчет: $\lambda = 0.0013$, $k = 3$. Значит, аналитические формулы для расчета функции плотности и функции распределения примут вид

$$f_3(t) = \frac{0.0013(0.0013t)^3}{3!} e^{-0.0013t}, \quad (7)$$

$$F_3(t) = 1 - \sum_{j=0}^3 \frac{(0.0013t)^j}{j!} e^{-0.0013t}, \quad (8)$$

Теоретическая функция плотности закона Эрланга 3-го порядка, с полученными параметрами, показана на рис. 3.

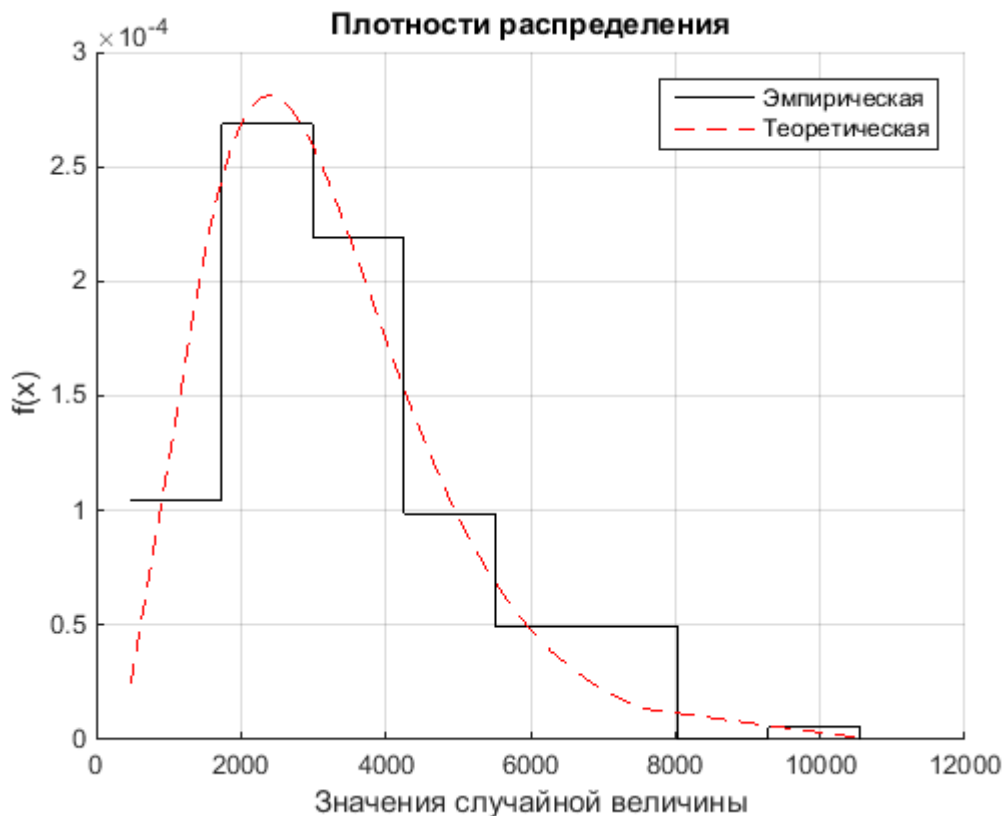


Рис. 3. Сравнение функций плотности

По графикам рис. 3 становится видно, что выборочная и теоретическая функции плотности близки друг другу.

Теперь требуется сделать заключение: согласуются ли наблюдения с истиной. Для решения этой задачи следует выдвинуть две гипотезы:

H_0 : Случайная величина X поступающих запросов на сайт подчиняется распределению Эрланга 3-го порядка;

H_1 : Случайная величина X поступающих запросов на сайт не подчиняется распределению Эрланга 3-го порядка (альтернативная гипотеза);

Проверка гипотез будет производиться в среде автоматизации технических расчетов MATLAB, т.к. данная система предоставляет удобный и гибкий функционал для решения задач подобного характера.

Для проверки будем использовать критерий согласия Пирсона [2]. Для этого используем функцию `chi2gof` системы MATLAB, которая и даст нам окончательный ответ [7]. Также эту задачу можно решить и в Microsoft Excel, а также с помощью любого языка программирования, например C [8]. Зададим доверительную вероятность $p = 0.95$, следовательно уровень значимости $\alpha = 0.05$. и найдем количество степеней свободы предполагаемого закона:

$$\mathcal{D} = \Delta - i - 1, \quad (9)$$

где i – количество параметров, характеризующее теоретическое распределение, у распределения Эрланга их 2: λ, k .

Функция `chi2gof` возвращает переменную h , которая принимает два значения 0 или 1. В случае 0 подтверждается гипотеза H_0 . В случае 1 подтверждается гипотеза H_1 . В нашем случае гипотеза на уровне значимости 5 % подтвердилась (функция `chi2gof` вернула 0).

Расчетным путем было подтверждено, что запросы, поступающие на сайт Мордовского государственного университета им. Н.П. Огарёва в период с 01.06.2015 по 23.10.2015 подчинялись закону распределения Эрланга – 3 порядка. А это значит, что данный факт можно использовать в задачах телекоммуникации для моделирования входящего потока требований поступающих на сайт университета. Предложенная методика может быть использована для анализа сайтов других учебных заведений республики и других субъектов Российской Федерации.

Список использованных источников

1. Лещанкин К.А. Региональный интернет // Мордовия: Мир Internet. 1999. № 3 (30). С. 34-37.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989. 530 с.
3. Sturges H. The choice of a class-interval // J. Amer. Statist. Assoc. 1926. 21. P. 65-66.
4. Афонин В.В., С.А. Федосин. Моделирование систем [Электронный ресурс]. Режим доступа: <http://www.intuit.ru/studies/courses/627/483/info>
5. Афонин В.В. Основы анализа систем массового обслуживания [Электронный ресурс]. Режим доступа: <http://elibrary.ru/item.asp?id=19581660>
6. Афонин В.В. Основы анализа систем массового обслуживания: Учеб. пособие / В.В. Афонин, С.М. Мурюмин, С.А. Федосин. Саранск: Изд-во Мордов. ун-та, 2003. 236 с.
7. Афонин В.В. Моделирование систем: Учебно-практическое пособие / В.В. Афонин, С.А. Федосин. 2010. 231 с.
8. Александров Э.Э. Программирование на языке С в Microsoft Visual Studio 2010: учеб. пособие / Э.Э. Александров, В.В. Афонин. Саранск: Изд-во Мордов. ун-та, 2010. 423 с.