

Министерство цифрового развития, связи и
массовых коммуникаций Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего
образования «Сибирский государственный университет телекоммуникаций и
информатики» (СибГУТИ)

Отчёт
по лабораторной работе №3
по дисциплине «**Прикладные задачи теории вероятностей**»

Выполнил:

студент гр. ИС-142

«__» декабря 2023 г.

/Наумов А.А./

Проверил:

профессор кафедры В.С.,

«__» декабря 2023 г.

/Родионов А.С./

Оценка « _____ »

Новосибирск 2023

ВЫПОЛНЕНИЕ РАБОТЫ

Эта лабораторная работа направлена на изучение и анализ автокорреляционных функций в контексте входящего и исходящего сетевого трафика. Основные задачи включают:

1. Анализ Внутренней Структуры Данных: Через автокорреляционные функции мы стремимся понять внутреннюю структуру данных, исследуя взаимосвязь текущих значений временного ряда с их предшественниками.

2. Выявление Периодичности: Особый акцент сделан на определении регулярности или её отсутствия в данных, что достигается путем изучения повторяющихся шаблонов в автокорреляционных функциях на разных интервалах задержки.

3. Практическое Значение: Результаты анализа будут полезны для лучшего понимания характеристик трафика и могут способствовать более эффективному управлению сетевой инфраструктурой и трафиком.

4. Разработка Методологии: Создание и демонстрация методики анализа временных рядов, применимой к аналогичным данным в будущем.

Целью исследования является не только определение текущих особенностей данных трафика, но и предоставление инструментов для интерпретации этих данных с целью прогнозирования будущих тенденций и планирования ресурсов.

Автокорреляционная функция (ACF) в анализе временных рядов иллюстрирует меру линейной статистической связи между последовательными значениями в ряду. Она рассчитывается как серия коэффициентов корреляции между исходным временным рядом и его сдвинутой копией на определенное количество временных интервалов (лаг L). Формула для ACF определяется как:

$$ACF(L) = \frac{1}{n} \sum_{i=1}^{n-L} (X_i - \bar{X})(X_{i+L} - \bar{X})$$

Здесь X_i – значение временного ряда в момент времени i , \bar{X} – среднее значение временного ряда, n – общее количество наблюдений в ряду, и L – лаг.

С увеличением лага количество элементов ряда, используемых для расчета коэффициента корреляции, уменьшается. Практически максимальный лаг не должен превышать четверти длины ряда, т.е. $L_{\max} \leq n/4$.

График зависимости ACF от лага, известный как коррелограмма, помогает анализировать динамические свойства временных рядов.

ACF максимальна при $L = 0$, когда ряд полностью коррелирован сам с собой. Значения ACF варьируются от -1 до 1. Отличие временного ряда от

случайного набора значений заключается в статистической зависимости его элементов, которую и измеряет ACF. Если на ряд влияют долгосрочные внешние факторы, это приводит к появлению в ряде трендов и циклических компонент, которые можно выявить с помощью ACF. Например, если максимум ACF достигается при лаге $L = k$, то в ряде присутствует циклическая компонента с периодом k .

Формула выглядит следующим образом:

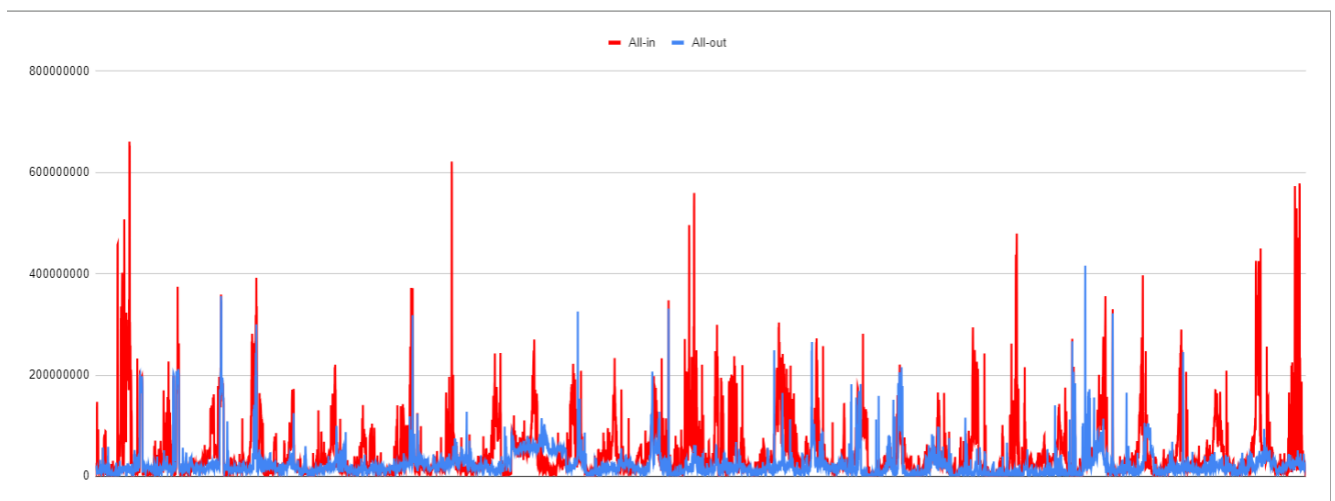
$$f(L) = \sum_{t=L}^n r_{t,t-L},$$

где:

где n — число членов (уровней) временного ряда, r — коэффициент корреляции.

Эта формула используется для вычисления коэффициентов автокорреляции, которые показывают, насколько сильно значение временного ряда в данный момент времени связано с его значениями в предыдущие моменты времени на расстоянии L временных шагов.

Визуализация исходных данных:



Этапы вычисления автокорреляционных функций:

1. Загрузка данных: Начнем с загрузки данных из текстовых файлов, в которых содержатся временные ряды. Для этого мы воспользуемся функцией `'pd.read_csv'` из библиотеки Pandas. Не будем включать заголовки, так как предполагается, что файлы содержат только числовые значения.

2. Использование Statsmodels для ACF: Для вычисления и визуализации автокорреляционных функций мы воспользуемся функцией `'plot_acf'` из библиотеки Statsmodels. Эта функция автоматически рассчитывает автокорреляционную функцию для временного ряда и представляет её в виде графика.

3. Настройка параметров ACF:

- В качестве входных данных для `plot_acf` мы передадим временные ряды из каждого файла.

- Параметр `lags` будет установлен на 40. Это означает, что функция рассчитывает автокорреляции для 40 задержек (lags). Выбор количества лагов зависит от длины временного ряда и интересующих временных интервалов.

- Для каждого временного ряда будет создан отдельный график.

4. Визуализация: Мы используем библиотеку Matplotlib для создания графиков. Графики будут отображать степень автокорреляции для различных лагов. Высокие значения на вертикальной оси указывают на сильную автокорреляцию на соответствующем лаге.

Основные аспекты при расчете автокорреляционных функций (ACF):

1. ACF и значения лагов: ACF измеряет степень соответствия значения временного ряда его собственному предыдущему значению на каждом лаге. Например, ACF с лагом 1 измеряет корреляцию между каждым значением и его непосредственным предшественником.

2. Интерпретация графиков: На графиках ACF, значения близкие к 1 указывают на сильную положительную корреляцию, значения около -1 указывают на сильную отрицательную корреляцию, а значения около 0 свидетельствуют об отсутствии корреляции.

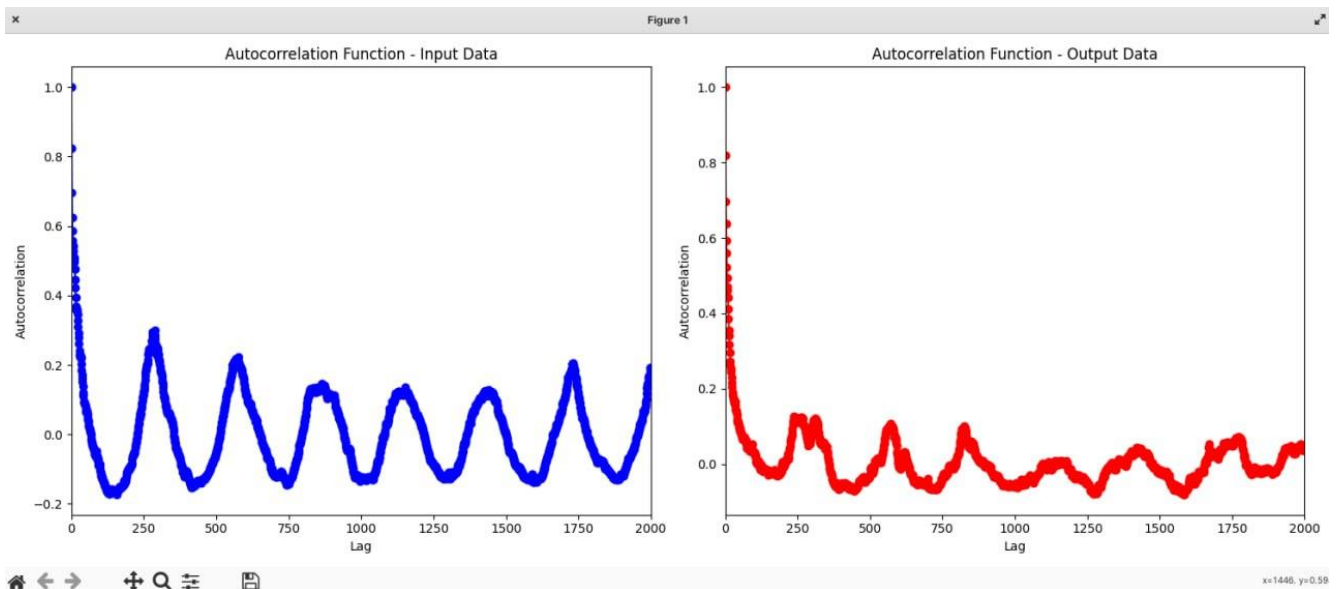
3. Периодичность: Периодичность в данных может проявляться через регулярные "пики" на графике ACF. Например, если вы замечаете пики на каждом 5-м лаге, это может указывать на периодичность в данных с периодом в 5 временных интервалов.

Для более точного определения наличия периодичности в данных, важно учитывать следующие аспекты:

- Расположение пиков: Если "пики" на графике ACF расположены с равными интервалами, это может быть хорошим индикатором наличия периодичности. Например, если "пики" наблюдаются каждые 10 лагов, это может указывать на периодичность с периодом в 10 временных интервалов.

- Высота и затухание пиков: В периодических данных обычно наблюдаются высокие "пики" на начальных лагах, которые постепенно уменьшаются с увеличением лага. Если "пики" остаются высокими и стабильными на больших лагах, это может указывать на более сложную структуру временного ряда.

Результаты вычислений:



Листинг программы:

```
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.stattools import acf

# read the contents of the uploaded files to understand the data format
with open('all-in.txt', 'r') as file_in, open('all-out.txt', 'r') as file_out:
    data_in = file_in.readlines()
    data_out = file_out.readlines()

# Displaying the first few lines of each file to understand their structure
data_in[:5], data_out[:5]

# Converting the data into numpy arrays
data_in_array = np.array([int(line.strip()) for line in data_in])
data_out_array = np.array([int(line.strip()) for line in data_out])

# Calculating the autocorrelation function for both series up to a lag of 2000
acf_in = acf(data_in_array, nlags=2000, fft=True)
acf_out = acf(data_out_array, nlags=2000, fft=True)

# Plotting the autocorrelation functions
plt.figure(figsize=(15, 6))

# Autocorrelation plot for input data
plt.subplot(1, 2, 1)
plt.plot(acf_in, marker='o', linestyle='-', color='blue')
plt.title('Autocorrelation Function - Input Data')
plt.xlabel('Lag')
plt.ylabel('Autocorrelation')
plt.xlim(0, 2000)

# Autocorrelation plot for output data
plt.subplot(1, 2, 2)
plt.plot(acf_out, marker='o', linestyle='-', color='red')
```

```
plt.title('Autocorrelation Function - Output Data')
plt.xlabel('Lag')
plt.ylabel('Autocorrelation')
plt.xlim(0, 2000)
plt.tight_layout()
plt.show()
```

Исходя из вышеуказанных графиков автокорреляционных функций (ACF) для входного и выходного трафика, можно сделать следующие выводы:

1. Автокорреляционная Функция для Входного Трафика: На графике ACF для входного трафика наблюдаются "пики" на определенных лагах. Это может свидетельствовать о наличии периодичности или зависимости данных от их предыдущих значений, примерно на расстоянии 300/600/900 единиц измерения назад.

2. Автокорреляционная Функция для Выходного Трафика: Схожая ситуация наблюдается и для выходного трафика, с "пиками" на лагах в 300/600/800 единиц.

Выводы:

1. Возможная периодичность в 24 часа (1 день): Анализ ACF указывает на выраженную периодичность в данных интернет-трафика, с периодом около 300 единиц (максимальный "пик" достигается на этом значении лага), что, вероятно, соответствует 24-часовому циклу (1 день). Эти временные зависимости могут быть связаны с дневной и ночной активностью пользователей.

2. Убывающая корреляция: С увеличением лага корреляция немного уменьшается, что означает, что данные на более дальних временных интервалах имеют менее выраженную зависимость друг от друга. Однако, если предположить, что эти "пики" соответствуют интервалам в 2,3,4... дня, то убывание корреляции может быть объяснено.

3. Отсутствие сильной корреляции на других лагах: Несмотря на высокую корреляцию на лаге 300, на других лагах (600 и 900) корреляция значительно меньше. Это может указывать на то, что влияние предыдущих данных сохраняется примерно в течение половины дня (12 часов), но становится менее значительным на более длинных временных интервалах.

4. Полезность анализа ACF: Анализ ACF позволяет выявлять периодичность и зависимости во временных рядах данных. В данном случае, он подтверждает наличие дневной периодичности в данных интернет-трафика, что может быть важной информацией при принятии решений и прогнозировании поведения пользователей или проведении

профилактических работ на серверах интернет-провайдера.

В общем, этот анализ помогает более глубоко понять характеристики данных интернет-трафика, выявить периодичность и зависимости, что может быть ключевым фактором для оптимизации ресурсов и принятия управленческих решений.