Задание по эмпирическому анализу алгоритма

Задание состоит из следующих пунктов:

1. Представить краткое описание исследуемого алгоритма, истории его разработки, области применения (по литературным данным).

2. Представить математический анализ алгоритма (на основе литературных источников).

3. Описать характеристики входных данных (выбрать их диапазон, размер и пр.), единицы измерения трудоемкости при проведении эксперимента (время выполнения программы или количество базовых операций), способ измерения.

4. Создать и представить программу, реализующую алгоритм (допускается использование литературных и достоверных источников).

5. Описать способ генерации входных данных, создать и представить программную реализацию генератора.

6. Провести вычислительный эксперимент в исследуемом диапазоне размеров входных данных (при необходимости провести многократное измерение трудоемкости f с последующим усреднением полученных результатов).

7. Проанализировать полученные результаты, сопоставить с теоретическими оценками (рассмотреть отношение измеренной трудоемкости к функции $g(n)$, задающей класс временной сложности $\Theta(g(n))$, рассмотреть отношение значений измеренной трудоемкости при удвоении размера входных данных)

Рекомендуемая форма отчета

Титульный лист с указанием задания, ФИО исполнителя.

Содержание отчета.

Основные результаты (по пунктам 1–7).

Список использованных литературных источников и информационных материалов.

Характеристики использованной вычислительной среды и оборудования.

Отчет оценивается по следующим критериям:

- Точность формулировок и грамотность изложения.
- Логичность и структурированность текста.
- Обоснованность и адекватность результатов и выводов.

Оценка **“отлично”** ставится при условии соблюдения критериев в каждом из пунктов 1–7 и в отчете в целом. Если критерии выполнены хотя бы в шести пунктах, то ставится оценка **“хорошо”**, хотя бы в 4 пунктах — оценка **“удовлетворительно”**, в других случаях ставится оценка **“неудовлетворительно”**.

***Задание повышенной сложности**

Задание повышенной сложности обусловлено необходимостью повысить точность результатов эмпирического анализа, который в обычном случае позволяет получить лишь точечные оценки среднего времени выполнения программы и лежащего в ее основе алгоритма (см. например, Эмпирический анализ алгоритмов, стр. 127 в книге: А.В. Левитин Алгоритмы: введение в разработку и анализ алгоритмов. М: Издательский дом “Вильямс”, 2006.
`https://books.google.ru/books?id=HIaf7DSPtl0C&pg=PA127`).

Применение классического подхода математической статистики, связанного с построением доверительных интервалов оцениваемой величины трудоемкости с заданной доверительной вероятностью, приводит к понятию доверительной трудоемкости (см. М.В. Ульянов, В.Н. Петрушин, А.С. Кривенцов Доверительная трудоемкость — новая оценка качества алгоритмов // Информационные технологии и вычислительные системы. 2009, №2, с. 23–37.
`http://www.isa.ru/jitcs/images/stories/2009/02/23_37.pdf`).

Анализ алгоритма в рамках этого подхода представляет собой достаточно большое исследование, поэтому оценка за задание повышенной сложности, при условии его выполнения, засчитывается, как один вопрос в экзаменационном билете.

I. Этап предварительного исследования (проверка гипотезы о законе распределения)

1. Выбрать размер входных данных, при котором будет проведено исследование.

2. Определить необходимое число экспериментов m с программной реализацией для получения гистограммы относительных частот значений трудоемкости.

3. Провести экспериментальное исследование и получить значения трудоемкости $f_i = f_A(D_i)$, $i = \overline{1..m}$ для случайно генерируемых входов выбранного размера.

4. Получить нижнюю и верхнюю границы значений теоретической функции трудоемкости алгоритма для лучшего и худшего случаев, либо использовать оценки этих границ по выборке из экспериментальных данных.

5. Выбрать количество полусегментов для построения гистограммы частот значений трудоемкости.

6. Нормировать значения экспериментальной трудоемкости и построить на основе полученных данных гистограммы относительных частот в полусегментах.

7. Вычислить выборочное среднее и выборочную дисперсию по экспериментальным данным.

8. Сформулировать гипотезу и вычислить параметры аппроксимирующего закона распределения.

9. Вычислить теоретические частоты по функции плотности.

10. Вычислить наблюдаемое значение критерия Пирсона.

11. Проверить гипотезу о законе распределения: если нет оснований отвергнуть нулевую гипотезу, то перейти к основному этапу исследования; в противном случае —- выбрать другой закон распределения и повторно проверить гипотезу.

II. Этап основного исследования

1. Определить диапазон значений длин входа, на котором будут получены интервальные оценки.

2. Определить значения длин входа, для которых будут проводиться экспериментальные исследования.

3. Выбрать шаг изменения длины входа в экспериментальном исследовании.

4. Определить необходимое количество экспериментов с программной реализацией алгоритма при фиксированной длине входа для определения выборочной средней и дисперсии.

5. Вычислить значения выборочной средней и выборочной дисперсии для каждого значения n на основе экспериментальных данных.

6. Регрессионный анализ экспериментальных данных — построить уравнение регрессии для выборочной дисперсии.

7. Вычислить параметры $\alpha(n)$, $\beta(n)$ аппроксимирующего бета-распределения, как функции длины входа на основе полученных результатов.

8. Выбрать значения доверительной вероятности и вычислить значения γ–квантиля бета-распределения как функции длины входа: $x_\gamma(n) = B^{-1}(\gamma, \alpha(n), \beta(n))$.

9. Вычислить значения функции доверительной трудоемкости $f_\gamma(n) = f^\vee(n) + x_\gamma(n) \cdot (f^\wedge(n) - f^\vee(n))$ в исследуемом диапазоне длин входа.