بسمه تعالی

آمار و احتمال مهندسی

نیم‌سال اول ۹۶-۹۵

سیدعباس حسینی

دانشکده مهندسی کامپیوتر

پروژه‌ی نهایی

زمان تحویل: ۱۰ بهمن

1 Phase1: On the distribution of overlapping area of two random polygons

Estimating the overlapping area of two random polygons is a critical task in some fields. In this section, we want to estimate this value using sampling and in next parts, we would define how we want to do this.

1.1 Part I: N-hull overlapping area

In this part, we want to estimate the distribution of overlapping area between two random polygons. Fixing N, a sample of this area would be generated like this:

1. Considering a unit square, we choose N blue points and N red points independently random from uniform distribution on considered unit square.
2. We find convex hull of red points (red polygon) and convex hull of blue points (blue polygon).
3. We calculate overlapping area between red and blue polygons and call it a *N-hull overlapping area*. This amount obviously is between 0 and 1.

We can find the distribution of N-hull overlapping area using sampling.

1.2 Part II: N-gon overlapping area

In this part, we want to estimate the area of a random N-gon in a convex set K. We generate a random N-gon like this:

1. Considering convex set K, we choose points independently random from uniform distribution on K until the convex hull of chosen points is a N-gon. (notice that this might not be easy and might be time consuming as well)
2. We calculate the area of generated N-gon and call it a *N-gon overlapping area*.

We can estimate the expected value of N-gon overlapping area using sampling.

1.3 Your task

Your task in this phase has some parts, including:

1. Developing a program to visualize a random N-hull and a random N-gon. (20 points)

2. Estimating the distribution of N-hull overlapping area for $N \in \{5, 10, 15, 100, 200, 500, 4000\}$ using sampling and plotting it. Your plots must be as smooth as possible. (If you can't do it for large N, you will still get partial credit for the smaller choices of N) (20 points)

3. Plotting the estimated expected value of N-gon overlapping area as a function of $log(N) \in [log(5), log(5000)]$ when: (60 points)

    (a) K is the unit square (Square with side length of 1).

    (b) K is the unit circle (Circle with Diameter of 1).

    (c) K is the unit equilateral triangle (Equilateral triangle with side length of 1).

1.4 Knowledges you need to have

There are some skills and knowledges you need for doing your task, specially including:

1. What is a convex hull?

2. How to find a convex hull of a set of points? (An example algorithm is Graham Scan, which you can search about and implement).

2 Phase2: On predicting a multi-variable function

Another critical task in many fields is predicting a noisy function. For example we have some observations from a function. So we have $S = \Big\{\big(x_1, f(x_1)\big), \big(x_2, f(x_2)\big), ..., \big(x_N, f(x_N)\big)\Big\}$ in which for every x_i we have $f(x_i)$ as the value of function f on point x_i with a little noise. And the goal is finding f for applying it to new points. Note that x might be a vector and if so, f would be a multi-variable function. This operation is called regression. And in this part, we want you to do it in a specific way.

2.1 Your task

Your task in this phase has some parts, including:

1. Developing a program which gets a set of N observations in input each including a 1 dimensional point and the noisy value of f in that point and some other points on which the value of function f is requested. You must assume that f is a polynomial of degree M (M is smaller than N) and use estimators (MMSE, Method of moments) each in a separate program. Also you must plot input observations and predicted function in the given domain. (60 points)

2. Generate data sets from expotential distribution, and uniform distribution and plot the result of both methods in one graph. (the choice of M and the number of points is up to you) (40 points)

2.2 Input and output format

The input consists of one text file:

1. Observation.txt includes the observations, on the first line comes two doubles seperated by a space which specify the domain (i.e. 0.0 3.0) and then each succeeding line specifies an observation: it includes the value of the argument and the value of f in that point, seperated by a space. Your program must be able to plot the predicted function and the observations in one graph. (You must be able to set M from inside the code, or as a commandline input.)

3 Phase3: Interpolation

In this part we are going to use the first 2 parts of the project.

3.1 Your task

Your task in this phase has some parts, including:

1. Plotting the distribution of N-gon overlapping area as a function of N on the domain specified in part 1, but this time only sample 5 points and use the program you made for part 2, to generate predictions on other points to plot the distribution. The choice of the sample points is up to you. (40 points)

2. Use part 2 to redo the last task of part 1, using only 5 samples. Also survey what your function predictor predicts for large N. (60 points)

4 Phase Grading

By doing each phase completely, you would get a part of your whole mark. The percentage of mark assigned to each phase is:

1. Phase1: 30%.

2. Phase2: 40%.

3. Phase3: 30%.