

# CS 107 Section B - Probability

Spring 2019, AUA

## Homework No. 06

Due time/date: 10:35AM, 15 March, 2019

**Note:** Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

**Note:** Please provide your answers in the form of a decimal number, by calculating and simplifying fractions, with the accuracy of 2 digits after the period.

**Note:** You can solve this homework either by doing the calculations by hand, or by using Microsoft Excel (or an equivalent software), or by writing a Python/R code for that. Please explain your calculations.

**Problem 1.** We want to build a toy language model. To that end, we assume that our language consists of only the following sentences:

My name is Michael. And what is your name? My name is Gohar.  
I am a Mathematician. I like talking about Mathematics.

The name of this street is Baghramyan.

You need to know Probability formulas and notions. You studied Probability for almost two months. How good do you know Probability Theory? I know Probability pretty well, and I know how solve this Probability homework.

I prefer doing homework than watching a football game. I know how to play a piano.

- a. Construct the Vocabulary of our Language
- b. We consider the Unigram Model. Calculate the probability of having the sentence  
I like Probability.
- c. Now we consider the Bigram Model. We want to construct a 3-word sequence in this model. To that end, choose the most probable first word, then "autocomplete" to 3-word sentence (i.e., after choosing the first word, choose the most probable successor, then the most probable successor of the second word).

**Problem 2.** We have the following ranking of courses (a fake one, of course, just a random selection of 0-s and 1-s 😊) - ranks are 0-1, 1 = *Like*, 0 = *Didn't Like*

Student \ Course	Calc 1	Calc 2	Calc 3	I2CS	LA	RA	OOP	Prob
S1	1	1	1	1	0	1	0	0
S2	0	1	1	0	1	1	0	1
S3	1	1	0	1	1	0	0	1
S4	1	0	1	0	1	1	1	0
S5	1	0	0	0	1	1	1	0
S6	1	0	1	1	0	1	0	0
S7	0	1	0	0	1	0	1	0
S8	0	0	0	1	0	1	0	1
S9	1	1	1	1	1	1	1	0
S10	0	0	0	1	0	1	1	0

- Student  $S$  is taking (and liking!) the course Calc3. Recommend her the next course she will like;
- Student  $S$  is taking (and liking!) the course Calc3. Recommend her 2 next courses she will like most;
- (Supplementary) Now, Student  $S$  is taking (and liking!) the courses  $Calc3$  and  $I2CS$ . Recommend her 2 other course that she would probably like.
- (Supplementary) Now assume rankings are from 0 to 5, integers (5 = "great course"). Fill randomly (or by asking your friends to rank) the table above, and think about how to recommend the next course to Student  $S$ , who ranked 5 her recent  $Calc3$  class.

**Problem 3.** Another toy example, for Naive Bayes Classification (NBC). In the attached .csv file you will find some partial data from one of my exams at YSU. Data provides two midterms grades (0-5 points), and the final result (Pass/Fail), and the Final Exam results are missing. Now assume we have a new Student, who has 2 points from the Midterm 1 and 1 points from the Midterm 2. What is the decision of the NBC, will this Student pass my exam?

**Problem 4.** For each of the following Random Experiments, construct two different random variables (one discrete and one non-discrete, if that is possible) defined on the Sample Space of that Experiment. Indicate for each r.v. if it is discrete or not.

- The experiment is tossing 3 fair coins;
- The experiment is choosing at random a participant of our Probability course;
- The experiment is choosing at random two participants of our Probability course;
- The experiment is to consider the weather tomorrow;
- The experiment is to consider the next football game of our national team.

**Problem 5.** We are given a function  $F(x)$  by its graph, see Fig. 1.

- Is this a legal CDF for some r.v.  $X$ ? Explain. If yes, continue to next tasks, otherwise go to the next problem 😊
- Calculate the probability  $\mathbb{P}(X \leq 2)$ .
- Calculate the probability  $\mathbb{P}(1.2 \leq X \leq 3)$ .
- What is the range of  $X$  (here by the range of  $X$  I mean the smallest closed set  $A$  such that  $\mathbb{P}(X \in A) = 1$ , that is,  $\mathbb{P}(X \in \overline{A}) = 0$ )?

**Note:** You do not need to prove that the set you are specifying is the smallest one. Just give your intuition behind your choice.

- e. Calculate the probability  $\mathbb{P}(X \in \{0, 1, 3, 4.5\})$ .
- f. Calculate the probability  $\mathbb{P}(X \in [3, 5])$ .
- g. Calculate the probability  $\mathbb{P}(X \in (0, 1))$ .
- h. Which is more probable:  $X > 4$  or  $X < 0$ ?
- i. If  $X$  is discrete, calculate its PMF, and if  $X$  is continuous, find the PDF of  $X$ .

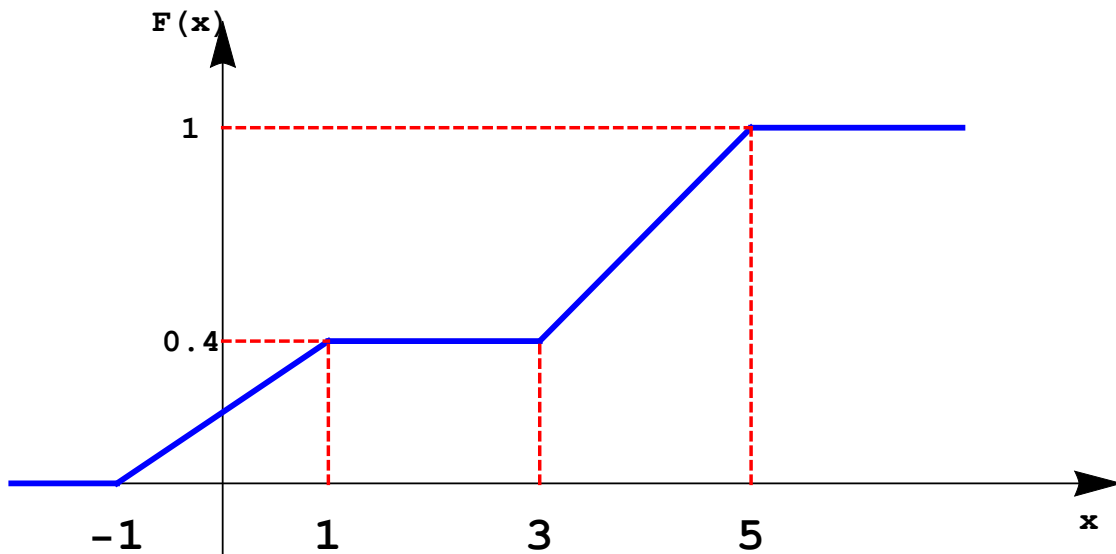


Figure 1: The graph of  $F(x)$

**Problem 6.** Fig. 2 shows the CDF of the r.v.  $X$ .

- a. Calculate, as accurate as possible, the probability  $\mathbb{P}(X \leq 5)$ .
- b. Calculate, as accurate as possible, the probability  $\mathbb{P}(X \geq 11)$ .
- c. Which is more probable:  $X \in [-5, 0]$  or  $X \in [1, 2]$ ?

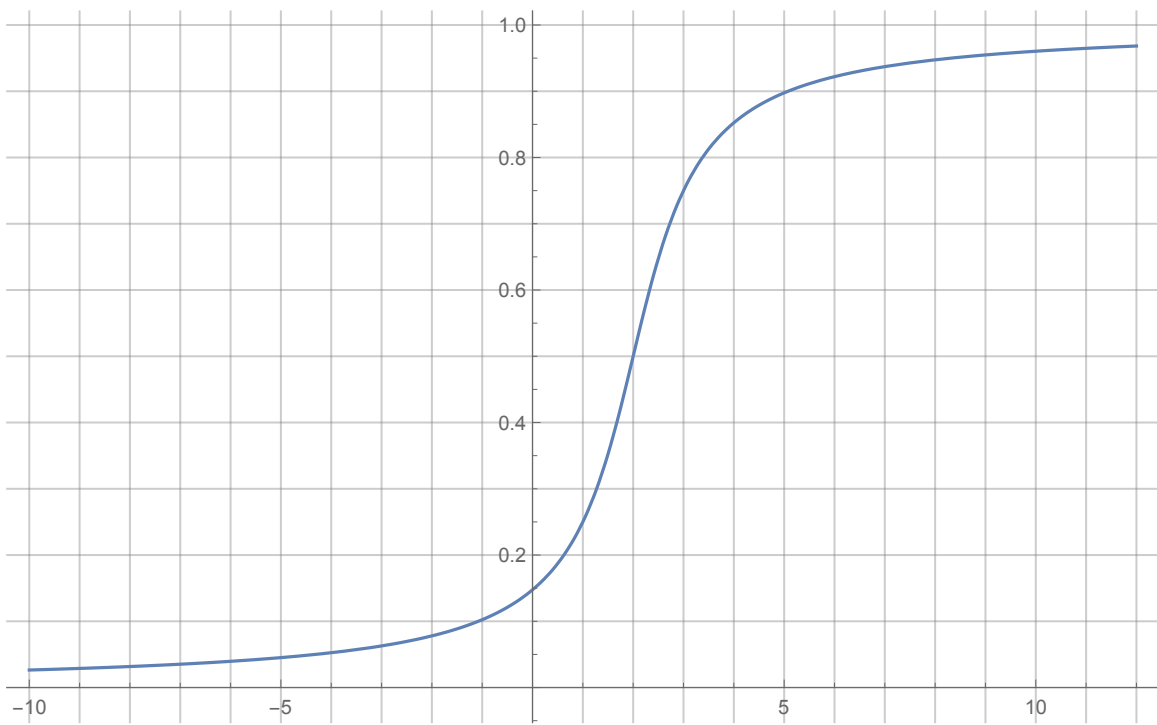


Figure 2: The graph of  $F(x)$