Name: ______________________ SID: ______________

Collaborators: ______________________________

Week 2 Problem Set

Probability
PHW142

You must put your name and SID at the top of the page.

The exercises in this week's problem set are adapted from our textbook.

Be sure to show your work by stating intermediate steps and reasoning used.

When you upload your problem set solution, the problem set key will be unlocked for you. Because the problem set key will be released when you upload your solution, you may only upload once.

No exceptions.

This problem set is worth 24 points.

Chapter 10 exercises

unintended pregnancies

1. **(5 points)** Birth certificates show that approximately 9% all births in the United States are to teen mothers (ages 15 to 19), 24% to young-adult mothers (age 20 to 24), and the remaining 67% to adult mothers (age 25 to 44). An extensive survey of live births examined pregnancy type, defining an unintended pregnancy as one that was unwanted or mistimed by at least two years. The survey found that "only 23% of births to teen mothers are intended, and 77% are unintended. Among births to young-adult women age 20–24, 50% are intended, and at ages 25–44, 75% are intended."

   (exercise 10.12 in the 4th edition of Baldi and Moore)

1.1 Express all the percents cited here using conditional probability notation. The sample space is live births in the United States. **2 points**

1.2 Draw a tree diagram representing the age group of the mother (teen, young-adult, adult) and the pregnancy type (intended, unintended) for live births in the United States. Upload the image here.

Suggestion: Draw the tree, scan and upload. **2 points**

1.3 What is the probability that any given live birth in the United States is unintended? Your tree diagram should be helpful. **1 point**

blood types

2. **(4 points)** All human blood can be "ABO-typed" as O, A, B, or AB, but the distribution of the types varies a bit among groups of people. Here are the distributions of blood types for a randomly chosen person in China and in the United States:

	O	A	B	AB
China probability	0.35	0.27	0.26	0.12
US probability	0.45	0.40	0.11	0.04

(exercise 10.30 in the 4th edition and 10.28 in the 3rd edition of Baldi and Moore)

2.1 Choose an American individual and a Chinese individual at random, independently of each other. What is the probability that both have type O blood? **1 point**

2.2 What is the probability that both individuals have the same blood type? **3 points**

cancer-detecting dogs

3. **(4 points)** Research has shown that specific biochemical markers are found exclusively in the breath of patients with lung cancer. However, no lab test can currently distinguish the breath of lung cancer patients from that of other subjects. Could dogs be trained to identify these markers in samples of human breath, as they can be to detect illegal substances or to follow a person's scent? An experiment trained dogs to distinguish breath samples of control individuals by using a food-reward training method. After the training was complete, the dogs were tested on new breath samples without any reward or clue using a double-blind, completely randomized design. Here are the results for a random sample of 1286 breath samples:

dog result	sample from cancer subject	sample from control subject	total
positive	564	4	568
negative	10	708	718
total	574	712	1286

(adapted from exercise 10.34 in both editions Baldi and Moore)

3.1 Find the ***sensitivity*** of the test, P(positive test | disease): the test's ability to correctly give a positive result when a person tested has the disease. **1 point**

3.2 Find the ***specificity*** of the test, P (negative test | no disease): the test's ability to correctly give a negative result when a person tested does not have the disease. **1 point**

2 additional questions:

You don't need to use Bayes's Rule to find the answers; just use the values in the table.

3.3 For this special group, find the probability that a subject tested really has cancer, given the test result is positive. (This doesn't apply to the entire US population.) **1 point**

3.4 For this special group, find the probability that the subject tested really does not have cancer, given that the test is negative. (This doesn't apply to the entire US population.) **1 point**

world weight situation

4. **(6 points)** Obesity is a growing public health concern worldwide. Adults with a high body mass index (BMI) of 25 or greater are considered overweight or obese. The following table shows the number of adults (in millions) who are overweight or obese in countries with different income levels, based on data from the World Health Organization and the United Nations.

   Note: Baldi and Moore's choice of column labels is not appropriate. In this context, the opposite of high BMI is NOT low BMI! A BMI of 24 is in no sense low.

country income level	millions with high BMI	millions whose BMI is not high	total
high	549	414	963
upper middle	612	1042	1654
lower middle	288	1083	1371
low	63	357	420
all	1512	2896	4408

4.1 What is the probability that a randomly selected adult has a high BMI? **1 point**

4.2 What are the conditional probabilities that a randomly selected adult has a high BMI, given each country income level? Write a short summary of your findings in context. **2 points**

4.3 Compute compare P (high BMI lhigh-income country) and P (high-income countryl high BMI). Explain, in words, what these conditional probabilities mean and why they are not the same. **3 points**

Chapter 12

Down syndrome

5. **3 points** The state of New York reported 1484 live births in which the infants had Down syndrome (trisomy 21) between 2006 and 2010, which averages to about 5.7 cases per week. While the causes of Down syndrome are not fully understood, it is reasonable at this point to assume that live births are independent and the weekly rate is constant. Let X be the count of babies born with Down syndrome in the state of New York in each week.

   (exercise 12.15 in the 4th edition of Baldi and Moore)

   Use the Poisson functions in R and include your functions in the answers.

   5.1 What is the probability that no child is born with Down syndrome in a given week? **1 point**

5.2 What is the probability that at least 1 child is born with Down syndrome in a given week? **1 point**

5.3 What is the probability that 10 or more children are born with Down syndrome in a given week? **1 point**

vaccination at work

6. **(2 points)** Whooping cough (pertussis) is a highly contagious bacterial infection that was a major cause of childhood deaths before the development of vaccines Approximately 80% of unvaccinated children who are exposed to whooping cough will develop the infection, compared to only 5% of vaccinated children.

   A group of 20 children at a nursery school are exposed to whooping cough by playing with an infected child.

   (exercise 12.38 in the 4th edition of Baldi and Moore)

   Use the binomial functions in R and include your functions in the answers.

   6.1 If all 20 children have been vaccinated, what is the probability that no more than 2 of the 20 children will develop infections? **1 point**

   6.2 If none of the 20 of the children have been vaccinated, what is the probability that 18 or more of the 20 children will develop infections? **1 point**