CSCI 4100 Assignment 2

Boliang Yang 661541863

1. Exercise 1.8 in LFD

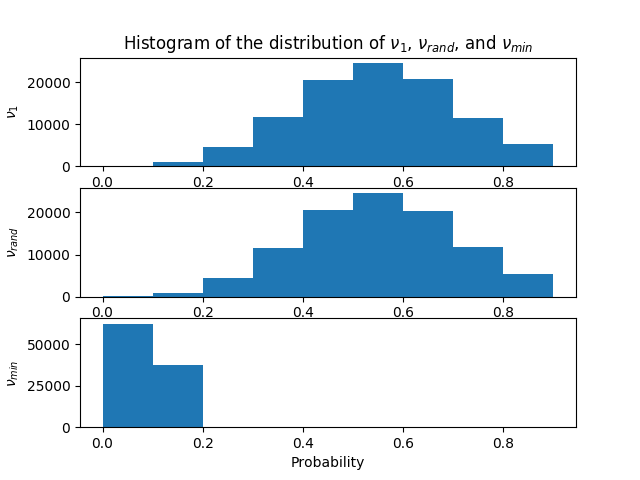
We have occurrence of red to be 0 or 1 since and there are 10 marbles in the sample, the red occurrence is .

2. Exercise 1.9 in LFD

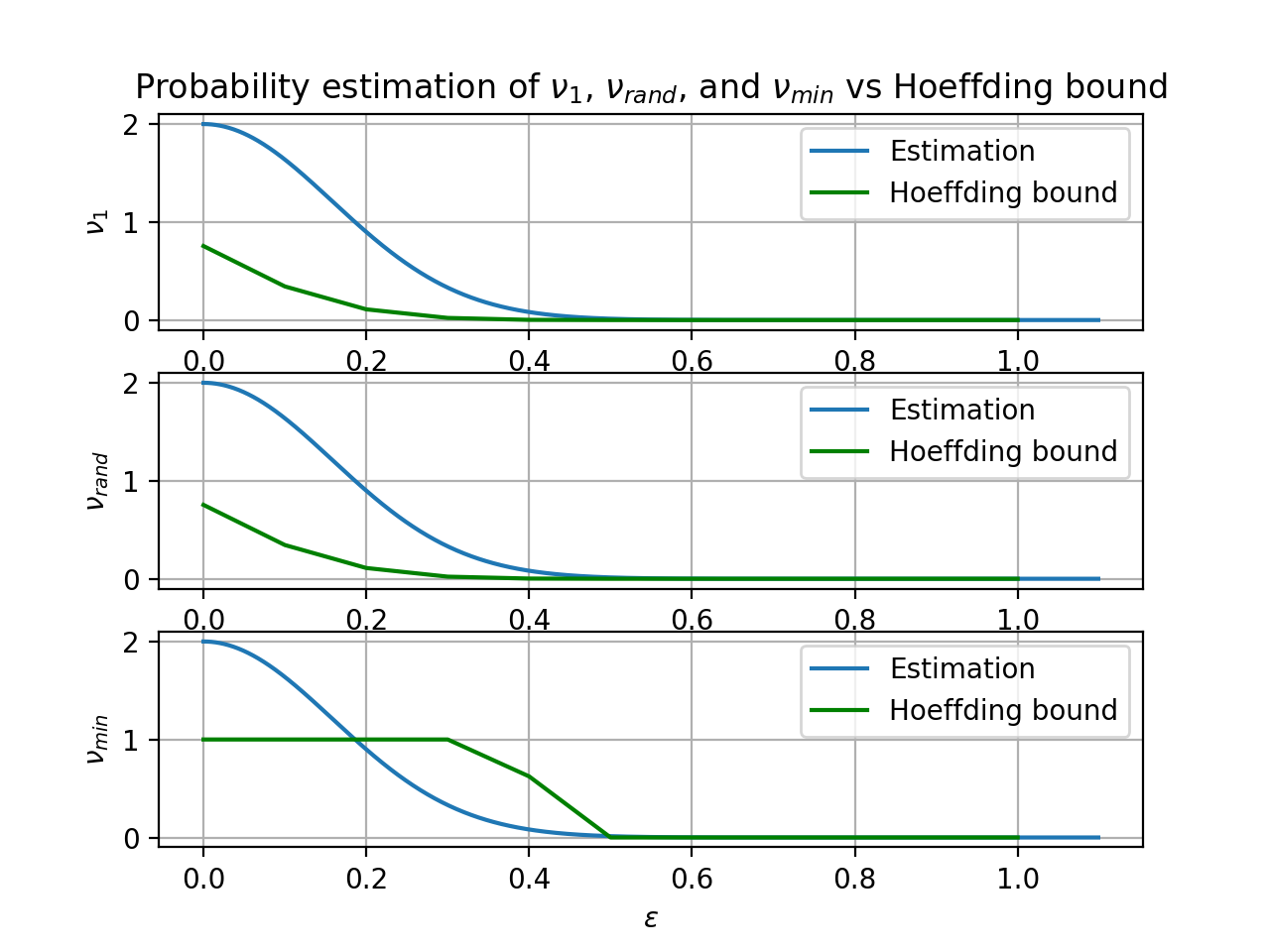
Since and , then

The answer here is large comparing to the answer in 1.8.

3. Exercise 1.10 in LFD

 (a) The for all three coins is 0.5.

(b)

 (c)

(d) and obey the Hoeffding bound, doe not.

Essentially and are randomly chosen coins, and was not randomly selected but specifically chosen. To be applicable to the Hoeffding bound, the selection must be random. Therefore, and obey the Hoeffding bound, but doe not.

(e) If we assume that we do not know the probability of heads as in Figure 1.10 that we don’t know the fraction of red marbles, different coins are just like different bins and our hypothesis set are all 1000 coins, then the problem is similar to the bins problem. We pick out the bins for , then the hypothesis are , respectively. Because and are randomly chosen and independent, then they obey the Hoeffding bound. And does not obey because we always choose it that has the least error and it is not independent.

4. Exercise 1.11 in LFD

(a) No. Because the given dataset D could contain all the training examples with the same value, for example y = +1 for all in D, then it cannot guarantee to perform better than random on any point outside D. Nothing is guaranteed outside the data set D.

(b) Yes, it is possible. It is possible that all points outside D have y = -1 then C would work better than S. Again, nothing is guaranteed outside the D.

(c) If , then D should contain more than examples with y = +1:

(d) If , then S would choose . If C works better outside D, then . Then and .

If , then S would choose . If C works better outside D, then . Then and .

Then and since, we have . Therefore, and .

5. Exercise 1.12 in LFD

The best I can promise her is (c). Nothing is guaranteed outside the data set. Depending on the problem, we may need a very large hypothesis set to deal with the training data. Because the data set is fixed to 4000 points, a good hypothesis will lead to high in . Then we declare that we failed. Or we may find the function by just trying a few times, then is approximate to . And if approximates to 0 then we can return a hypothesis with high probability that it will approximate well out of sample.

6. Problem 1.3 in LFD

(a) For separable data, the optimal set of weights will separate the data such that . Then for all , we have . Thus, .

(b) Since , then .

Since ,

then . Now we show by induction.

Base case:

And

Induction step:

Thus,.

By induction, we show that .

(c)

Since

Then together with , we have .

Thus, .

(d) Prove by induction.

Base case: When t = 0,

Induction step: Assume for t = k where .

At , .

Since , then . Then . Thus, holds for .

By induction, we prove that , where .

(e) Since , then and it follows .

Since , then , and it follows . Then . Since , then . Thus, we have

7. Problem 1.7 in LFD

(a) For : The probability of getting no head with 1 coin is

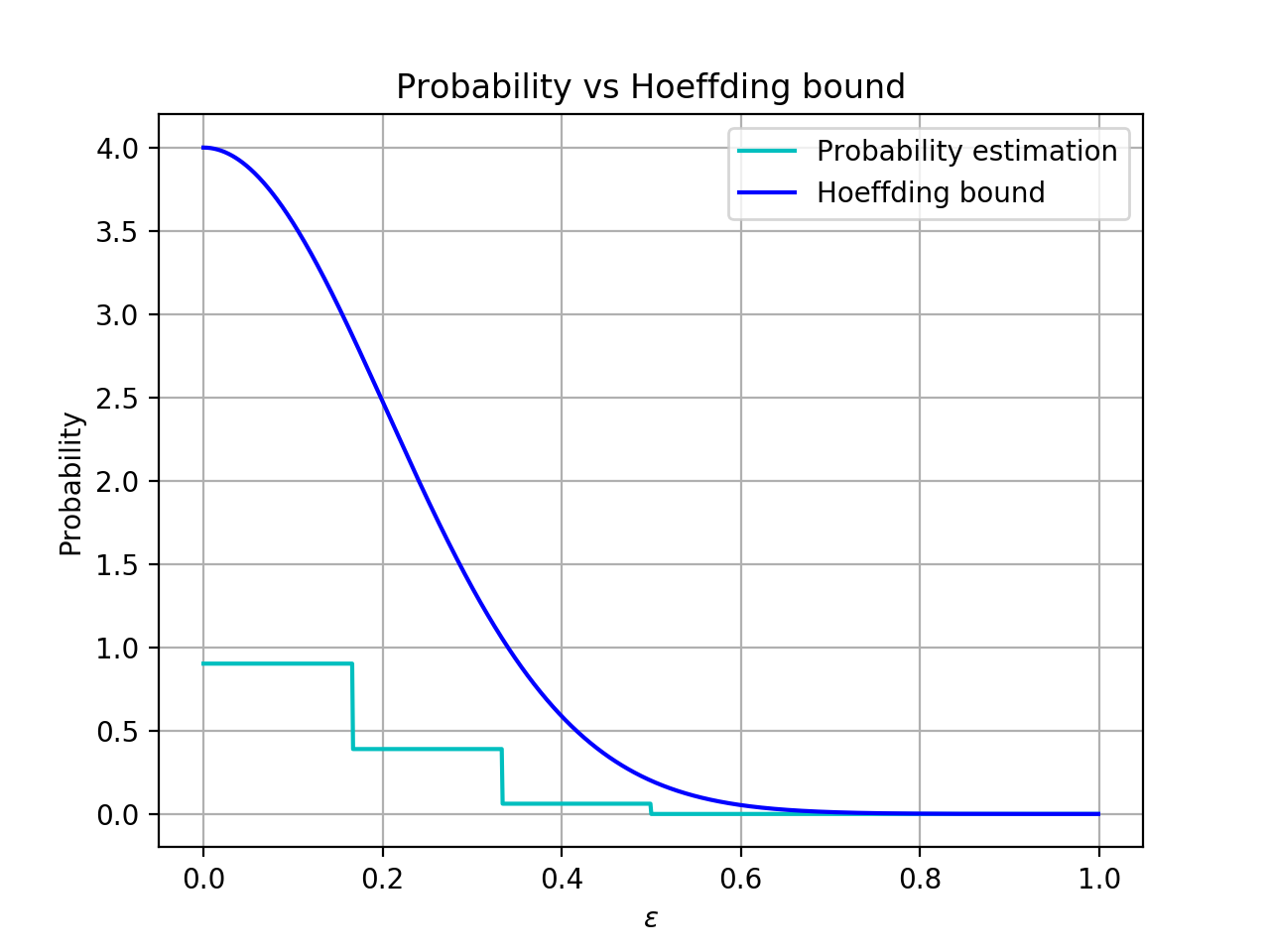
For 1000 coins:

For 1000000 coins:

For : The probability of getting no head with 1 coin is .

For 1000 coins:

For 1000000 coins:

 (b) Since the two coins are independent, then using and N =6 we have: for hoeffding bounds,