

CSCI-4100 Assignment 2
Yichuan Wang 661414395
EXERCISES

1.8

$v \leq 0.1$ means we get 1 red ball or 0 red ball.

P[1 red ball]: $10 * (0.1)^9 * 0.9 = 9 \times 10^{-9}$

P[0 red ball]: 10^{-10}

Answer: 9.1×10^{-9}

1.9

$u = 0.9, v \leq 0.1, |v - u| \geq 0.8$

$$P[|v - u| \geq 0.8] \leq 2e^{-2 \times 0.8^2 \times 10} = 5.52 \times 10^{-6}$$

The Hoeffding Bound result is greater than result from exercise 1.8

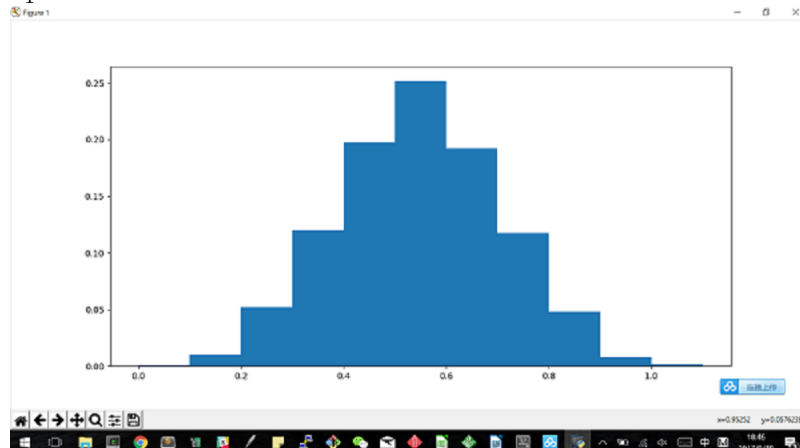
1.10

(a)

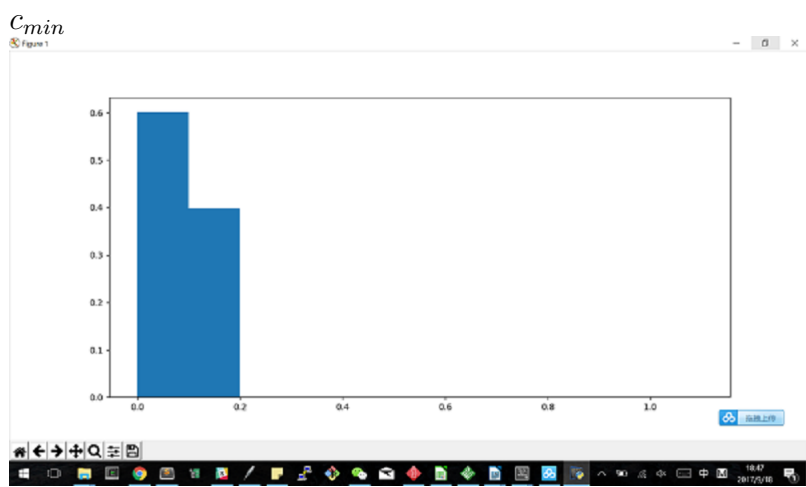
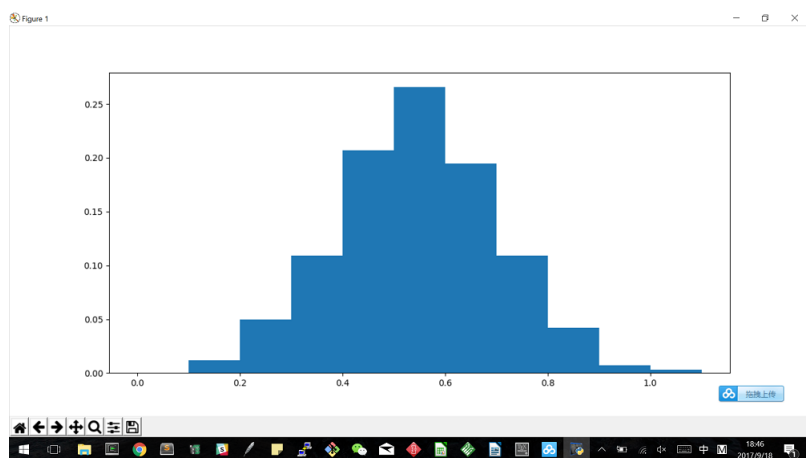
$u = 0.5$ for these three coins.

(b)

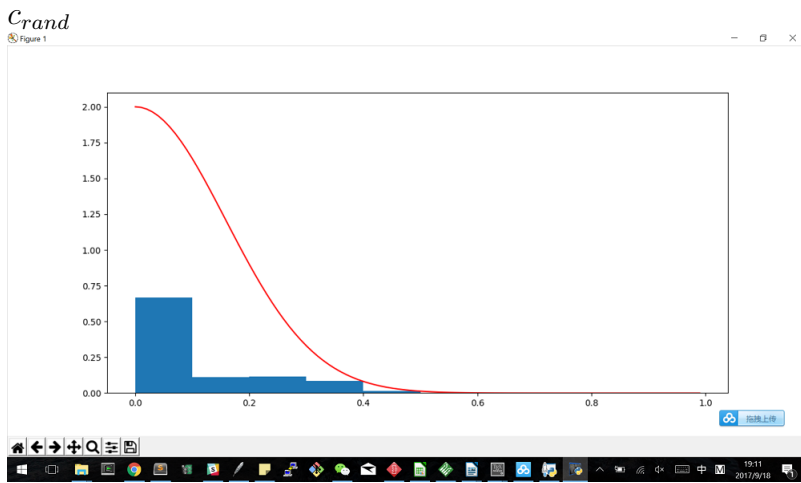
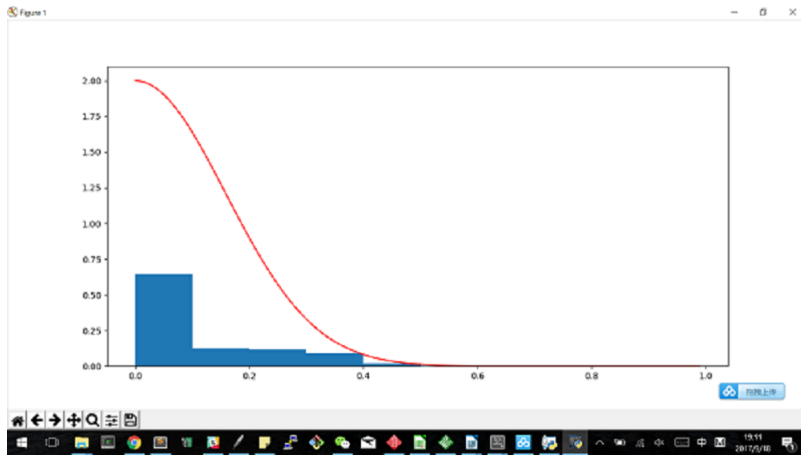
c_1



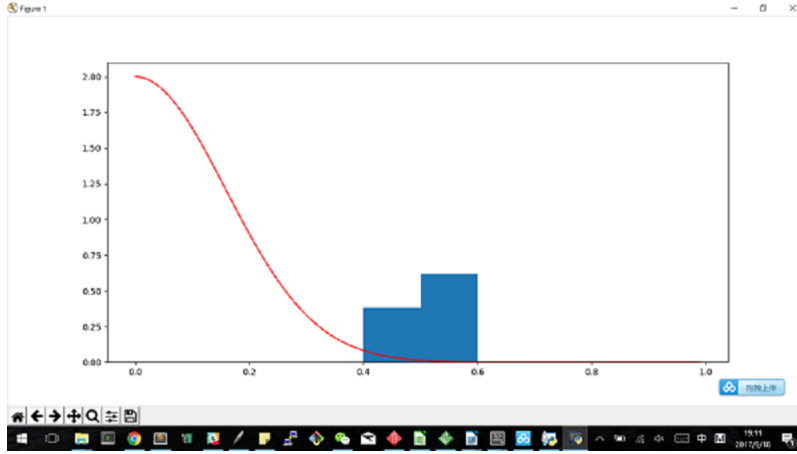
c_{rand}



(c)
 c_1



C_{min}



(d)

The first coin and the random coin obey the Hoeffding Bound because they are fixed before sampling. The minimum coin doesn't obey the Hoeffding Bound because the selected coin is changed after sampling if a smaller v is present from later sampling events.

(e)

There are 1000 bins, each having equal amount of green and red balls ($u=0.5$). We take ten balls from each bin and calculate v based on our sampling result. Looking at the first flipped coin means we always look at the sampling result of bin no.1, which won't be affected by our sampling data. Same thing for the random selected bin; at each round we look at a fixed random bin for result. In this case, these two coin cases obey the Hoeffding Bound. The minimum coin case is such that we consider 1000 bins together and take the bin with minimum result. In this case our hypothesis is not fixed during data generation since we could always jump from one bin to another if the other bin has a lower v ; so our result is not protected by Hoeffding Bound.

1.11

a) No.

b) Yes.

c) C is better than S when S choose h_1 , which will happen if there are 13 or more +1 from sampling. The probability is therefore

$$\sum_{n=13}^{25} \binom{25}{n} 0.9^n \times 0.1^{25-n} = 0.99999983791$$

d) There is no such p since the result from sample approximates the real result, and S chooses hypothesis based on sample. In this case S will always choose the better one between h_1 and h_2 , and C will always choose the worse one.

1.12

(c) is what I can promise. I cannot guarantee g 's performance, I cannot find a g that is likely to produce good performance, but I can check if the g I picked is likely to produce good performance.

PROBLEMS

1.3

(a) Since w^* separates the data, all the data points are correctly classified. In this case all of $y_n(w^{*T} x_n) > 0$, So $\rho > 0$.

(b)

$$w(t) = w(t-1) + y(t-1)x(t-1)$$

$$w^T(t)w^* = (w(t-1) + y(t-1)x(t-1))^T w^* = w^T(t-1)w^* + y(t-1)w^{*T}x(t-1)$$

Since ρ is the minimum of $y_n w^{*T} x_n$, we have

$$y(t-1)w^{*T}x(t-1) \geq \rho$$

. So we have

$$w^T(t)w^* \geq w^T(t-1)w^* + \rho$$

Induction proof of $w^T(t)w^* \geq t\rho$:

Base Case: Iteration zero has $w(0)w^* = 0$ since $w(0) = 0$, and $t\rho = 0$ since $t = 0$; we have:

$$w(0)w^* = t\rho$$

Induction Step:

For iteration N ($t = N$), we have

$$w(N)w^* \geq N\rho$$

We have proved $w^T(t)w^* \geq w^T(t-1)w^* + \rho$, by substituting in N we have

$$w(N+1)^T w^* \geq w(N)^T w^* + \rho$$

Combine this equation with $w(N)w^* \geq N\rho$ we have

$$w^T(N+1)w^* \geq (N+1)\rho$$

QED

(c)

$$\begin{aligned} w(t) &= w(t-1) + y(t-1)x(t-1) \\ \|w(t)\|^2 &= \|w(t-1)\|^2 + \|x(t-1)\|^2 + 2w^T(t-1)x(t-1)y(t-1) \\ w^T(t-1)x(t-1)y(t) &< 0 - \text{Missiclassified} \\ \|w(t)^2\| &\leq \|x(t-1)\|^2 + \|w(t-1)^2\| \end{aligned}$$

QED

(d)

Base Case:

Iteration 0:

$$\|w(t=0)\|^2 = 0 = tR$$

Induction Step:

Assume iteration N is correct:

$$\|w(N)\|^2 \leq NR^2$$

According to part (c),

$$\|w(N+1)\|^2 \leq \|w(N)\|^2 + \|x(N)\|^2$$

Combine the result with assumption of iteration N we have

$$\|w(N+1)\|^2 \leq NR^2 + \|x(N+1)\|^2$$

Since R has the maximum absolute value among all x_n , we have

$$\|x_n\|^2 \leq R^2$$

So we have

$$\|w(N+1)\|^2 \leq NR^2 + R^2$$

$$\|w(N+1)\|^2 \leq (N+1)R^2$$

QED

(e)

Take the square root of part(d)'s result:

$$|w(t)| \leq \sqrt{t}R$$

Divide result of part(d) with the equation above:

$$\frac{w^T(t)}{\|w(t)\|w^*} \geq \frac{\sqrt{t}\rho}{R}$$

By moving ρ and R to the left side and take the square of both side of the equation, we get

$$t \leq \frac{R^2 \|w^*\|^2}{\rho^2}$$

1.7

(a)

P[a coin has v=0](denoted by $P[A]$): $(1-u)^{10}$

P[at least 1 coin in n coins has v=0](denoted by $P[B_n]$): $1 - (1 - P[A])^n$

$u = 0.05$:

$$P[A] = 0.95^{10} = 0.5987$$

$$P[B_1] = 1 - (1 - P[A])^1 = 0.5987$$

$$P[B_{1000}] = 1 - (1 - P[A])^{1000} = 1$$

$$P[B_{1000000}] = 1 - (1 - P[A])^{1000000} = 1$$

$u = 0.8$:

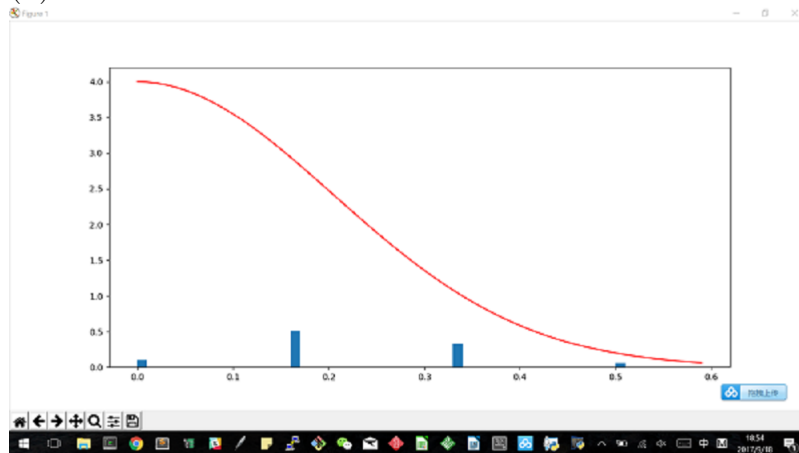
$$P[A] = 0.2^{10} = 0.0000001024$$

$$P[B_1] = 1 - (1 - P[A])^1 = 0.0000001024$$

$$P[B_{1000}] = 1 - (1 - P[A])^{1000} = 0.0001024$$

$$P[B_{1000000}] = 1 - (1 - P[A])^{1000000} = 0.0973$$

(b)



The blue histogram is the probability of maximum error over two coins; the red curve is the Hoeffding Bound.