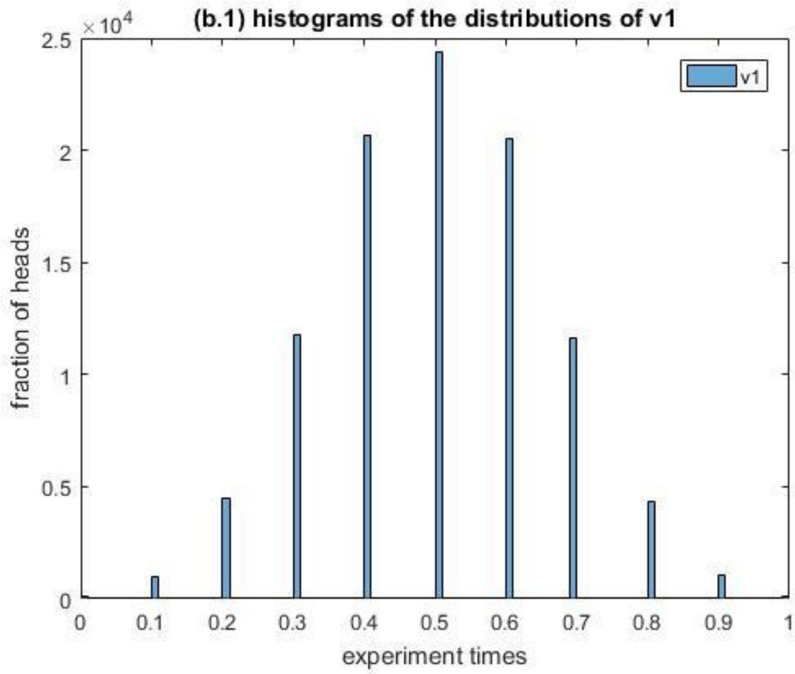
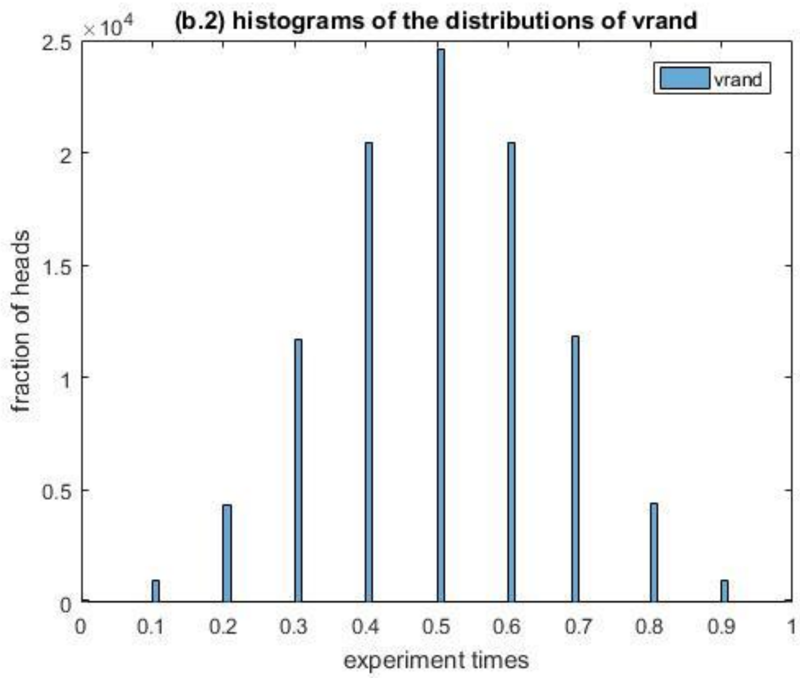
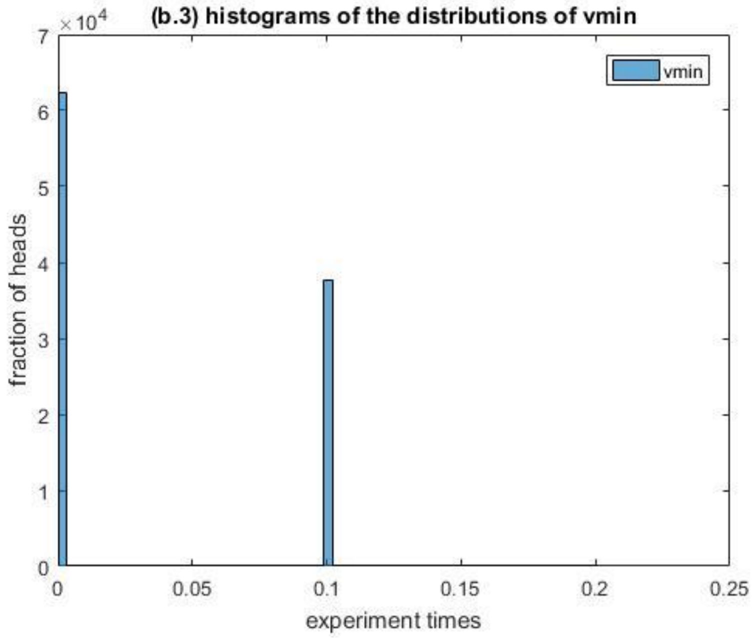
Collaboration with Jizhou Huang

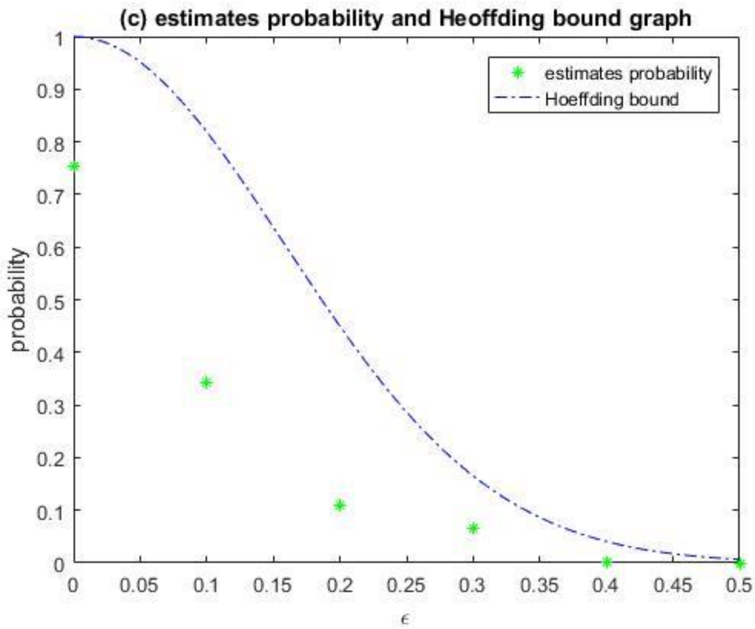
1. Exercise 1.10
   (a) The μ for the three coins selected is 0.5 because they are fair coin, that is, the possibility of head and tail are both equal 0.5.
   The one-line code command for the experiment:
   V = sum(randi([0 1], 1000, 10), 2) / 10;
   Where V is 1000 by 1 matrix containing vi, i = 1~1000. vi = heads times / 10
   (b)







(c)



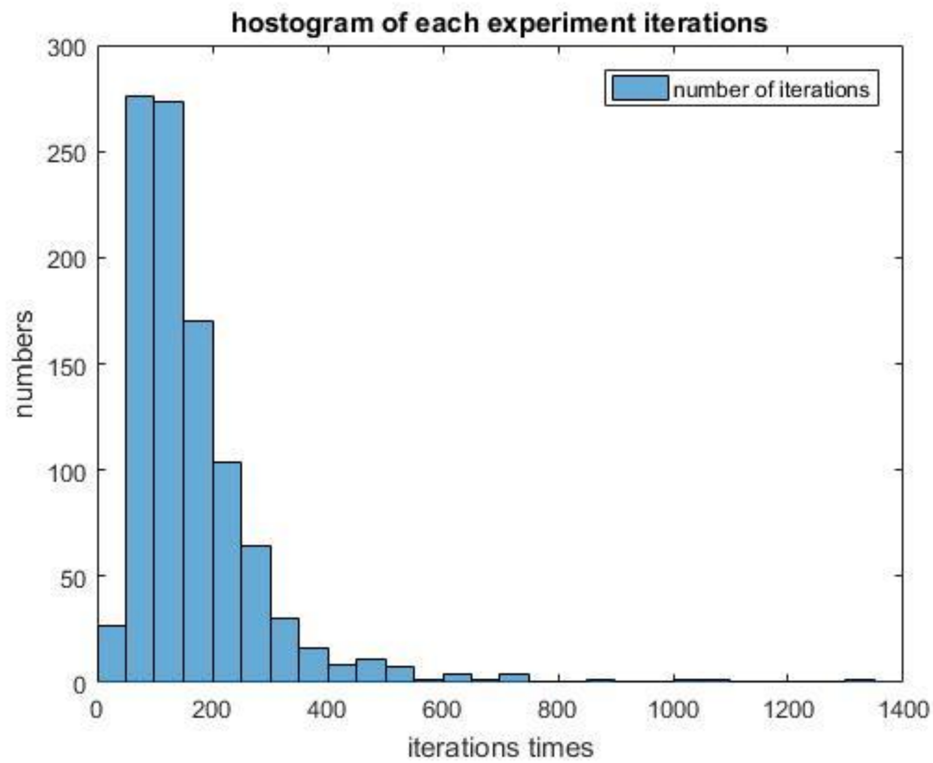
(d) v1 and vrand obey the Hoeffding bound, since they're in a likely normal distribution. However, the vmin doesn't obey the Hoeffding bound, because the estimates probability 0 heads happen P($|v - \mu| > \varepsilon$) = P($|0 - 0.5| > \varepsilon$) is approximately 0.6. On the contrary, the Hoeffding bound at v = 0 would be 0.00678 which is far less than vmin's case. The reason why v1 and vrand obey the rule yet the vmin doesn't is that v1 and vrand are selected "randomly", hence are one of random distribution. On the other hand, the vmin is "targeted" for the smallest value, which is not a random process. Hence, vmin doesn't obey the

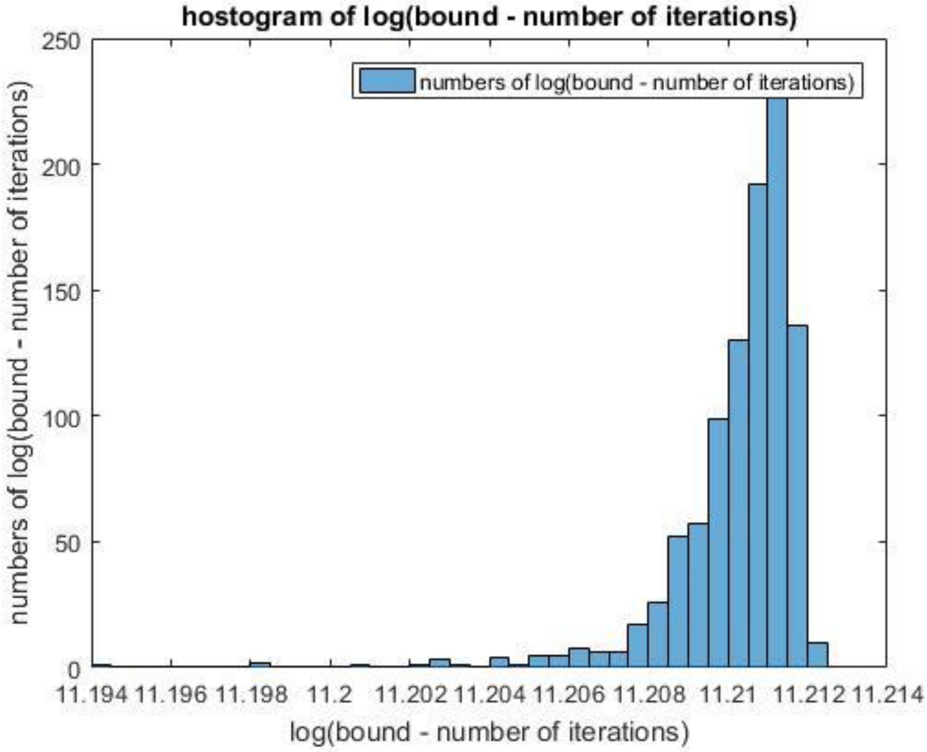
Hoeffding's rule.

(e) From (d) we can notice that the case of random distribution will obey the Hoeffding bound rule. The multiple bins in Figure 1.10 of LFD textbook is actually random distribution, because it contains only two elements, red bin and green bin, also it has random unknown μ. Just like flipping coin, the multiple bins problem would be random distribution, which will obey the Hoeffding bound rule.

2. Perceptron Learning experiment

```
[N, d, num_samples] = [100, 10, 1000]
= [ number of training examples, examples' dimension, experiments times]
```





Bound: t = $\frac{R^2\|\omega^*\|^2}{P^2}$, where $R = \max\|X_n\|$, P = $\min(y_n \cdot (\omega * X))$, X includes all training examples. For example, when x's dimension is 10 and there are 100 training examples, then X is a 10 by 100 matrix. Bound t is related to x's dimension, which in code is 'd', and related to ω's dimension, since x's and ω's norm is square root of it's all elements' summation. P is obtained by selecting the minimum value of the product of y and (ω*x), which has N elements.

3. Problem 1.7

(a) P(at least one coin equal 0) = 1 – P(all coins equal 0)
For $\mu = 0.05, 1$ coin: $(1 - C_0^{10} 0.05^{10})^1 = 1 - 0.5987 = 0.4013$
For $\mu = 0.05, 1000$ coins: $(1 - C_0^{10} 0.05^{10})^{1000} \cong 1$
For $\mu = 0.05, 1000000$ coins: $(1 - C_0^{10} 0.05^{10})^{1000000} \cong 1$
For $\mu = 0.8, 1$ coin: $(1 - C_0^{10} 0.8^{10})^1 = 1 - 1.024 \cdot 10^{-7} \cong 1$
For $\mu = 0.8, 1000$ coins: $(1 - C_0^{10} 0.8^{10})^{1000} \cong 1$
For $\mu = 0.8, 1000000$ coins: $(1 - C_0^{10} 0.8^{10})^{1000000} \cong 1$

(b) $max_{1,2} \mid v_i - \mu_i \mid > \epsilon$
Case 1: $(v_1 - \mu_1) > (v_2 - \mu_2) > \epsilon$

Case 2: $(v_1 - \mu_1) > \epsilon > (v_2 - \mu_2)$

Case3: $(v_2 - \mu_2) > (v_1 - \mu_1) > \epsilon$

Case4: $(v_2 - \mu_2) > \epsilon > (v_1 - \mu_1)$

~P(X) = 1 - (X^c)

$\Rightarrow 1 - (\text{caseA} \ \& \ \text{caseB})$, where

Case A: $(v_1 - \mu_1) < (v_2 - \mu_2) < \epsilon$

Case B: $< (v_2 - \mu_2) < (v_1 - \mu_1) < \epsilon$

$\Rightarrow$ case C: $((v_1 - \mu_1) < \epsilon)$ X $((v_2 - \mu_2) < \epsilon)$

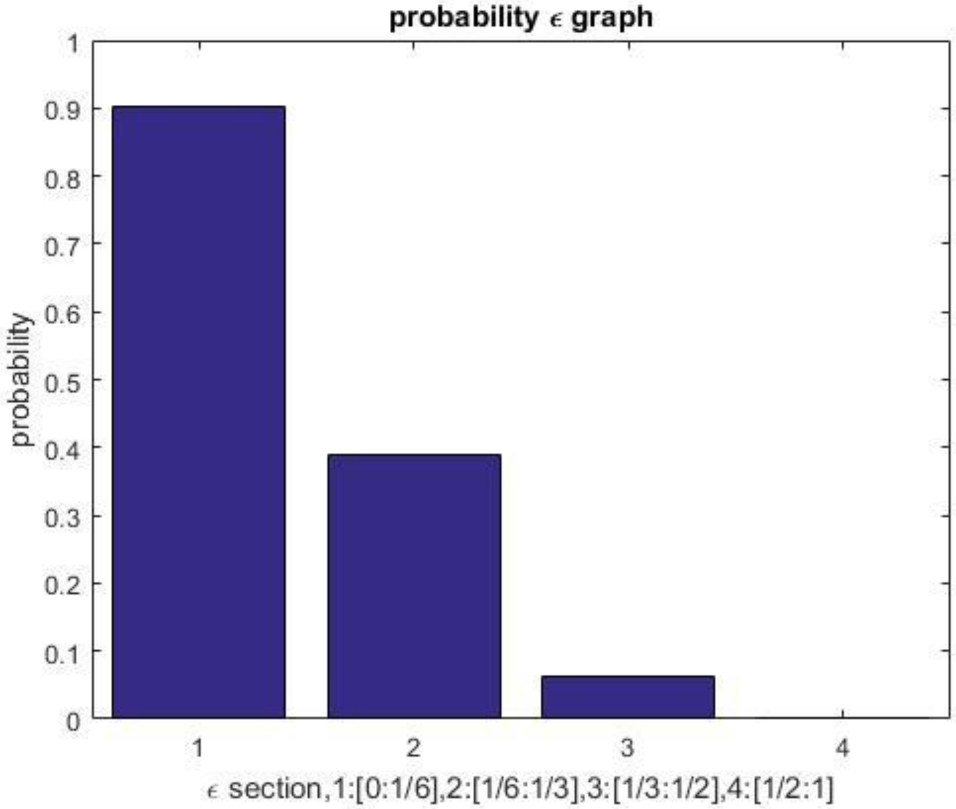
k	$\mid v_i - \mu_i \mid$	$P_k = C_k^6 (\frac{1}{2})^6$
0	1/2	0.015626
1	1/3	0.09375
2	1/6	0.234375
3	0	0.3125
4	1/6	0.234375
5	1/3	0.09375
6	1/2	0.015626

$1 - P_3 = 0.9023$

$1 - [P_2 + P_3 + P_4] = 0.3896$

$1 - [P_1 + P_2 + P_3 + P_4 + P_5] = 0.0615$

$1 - [P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6] = 0$



4. Problem 1.8

(a) $P[t \geq \alpha] \leq \frac{E(t)}{\alpha}$

$$\int_{\alpha}^{\infty} p(t)dt \ ? \ \int_{0}^{\infty} t\frac{p(t)dt}{\alpha}$$

$$\int_{\alpha}^{\infty} \alpha \ p(t)dt \ ? \ \int_{0}^{\alpha} tp(t)dt + \int_{\alpha}^{\infty} tp(t)dt$$

$$Because \ \int_{0}^{\alpha} tp(t)dt \geq 0 \ and$$

$$t \geq \alpha \Rightarrow \int_{\alpha}^{\infty} \alpha \ p(t)dt \ \leq \ \int_{\alpha}^{\infty} tp(t)dt$$

Therefore $P[t \geq \alpha] \leq \frac{E(t)}{\alpha}$

(b) $P[(u-\mu)^2 \geq \alpha] \leq \frac{\sigma^2}{\alpha}$, from (a)

$$\Rightarrow P[(u-\mu)^2 \geq \alpha] \ \leq E[(u-\mu)^2]/\alpha$$
$$= E[u^2 - 2u\mu + \mu^2]/\alpha$$

$= (\mathrm{E}(u^2) - 2\mu E(u) + \mu^2)/\alpha$
$= (\mathrm{E}(u^2) - \mu^2)/\alpha$
$= (\mathrm{E}(u^2) - E^2(u))/\alpha$
$= \sigma^2/\alpha$

(c) Suppose σ^{*2} is variance of u, and from (b)

$\Rightarrow \mathrm{P}[(\mathrm{u} - \mu)^2 \geq \alpha] \leq \frac{\sigma^{*2}}{\alpha}$

Because $\sigma^{*2} = \frac{\sigma^2}{N}$

$\Rightarrow \mathrm{P}[(\mathrm{u} - \mu)^2 \geq \alpha] \leq \frac{\sigma^{*2}}{\alpha \mathrm{N}}$

5. Problem 1.12

(a) $E_{in}(h) = \sum_{n=1}^{N}(h - y_n)^2$,

find min$\Longrightarrow$

$$\frac{d}{dh}E_{in}(h) = \frac{d}{dh}\sum_{n=1}^{N}(h - y_n)^2 = 0$$
$$= 2\sum_{n}^{N}(h - y_n) = 0$$
$$\Longrightarrow \sum_{n=1}^{N} h - \sum_{n=1}^{N} y_n = 0$$
$$\Longrightarrow \mathrm{h} = \frac{1}{N}\sum_{n=1}^{N} y_n$$

(b)

(c)

Reference:

Wikipedia

Mean: https://en.wikipedia.org/wiki/Mean

Expected Value: https://en.wikipedia.org/wiki/Expected_value