1 Problem 1.3

a, Let $\rho = \min_{1 \leq n \leq N} y_n(w^{*T} x_n)$. Show that $\rho > 0$

Since w^* separates the data perfectly

↳ $y_n(w^{*T} x_n)$ always $> 0 \Rightarrow \rho > 0$

b, • $w^T(t) w^* \geqslant w^T(t-1) w^* + \rho$

We have: $w(t) = w(t-1) + y(t-1) x(t-1)$

$\rightarrow w^T(t) = \left(w(t-1) + y(t-1)x(t-1)\right)^T$

$\rightarrow w^T(t) w^* = \left(w(t-1) + y(t-1)x(t-1)\right)^T w^*$

$\rightarrow w^T(t) w^* = w^T(t-1) w^* + \underbrace{y(t-1)x(t-1) w^{*T}}_{\geqslant \rho}$

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

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

  - Base Case: $t = 0$

    ↳ $w^T(0)w^* \geq 0\rho = 0$

    We assume $w(0) = 0$

    $\rightarrow 0 \geq 0$: true

  - Induction: Assume (1) is true at t, need to prove (1) is true at $t+1$

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

$$\rightarrow w^T(t+1)w^* = (w(t) + y(t)x(t))^T w^*$$

$$= \underbrace{w^T(t)w^*}_{\geq t\rho (1)} + \underbrace{w^{*T}(y(t)x(t))}_{\geq \rho}$$

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

↳ true at $t+1$

$$\therefore w^T(t)w^* \geq t\rho$$

c, $\|w(t)\|^2 \leqslant \|w(t-1)\|^2 + \|x(t-1)\|^2$

We have : $\|w(t)\|^2 = \|w(t-1) + y(t-1)x(t-1)\|^2$

$$= \left(w(t-1) + y(t-1)x(t-1)\right)\left(w(t-1) + y(t-1)x(t-1)\right)^T$$

$$= \|w(t-1)\|^2 + \underbrace{(y(t-1))^2}_{1}\|x(t-1)\|^2 + \underbrace{2y(t-1)x(t-1)w^T(t-1)}_{\leqslant 0}$$

$$\Rightarrow \|w(t)\|^2 \leqslant \|w(t-1)\|^2 + \|x(t-1)\|^2$$

d, (2) $\|w(t)\|^2 \leqslant tR^2$, $R = \max_{1 \leqslant n \leqslant N} \|x_n\|$

- Base case : $t = 0$

  ↳ $\|w(0)\|^2 = 0 = 0R^2$ (assume $w(0) = 0$)

- Induction : Assume (2) is true at t, need to prove (2) is true at $t+1$

$$\left.\begin{array}{l} (t+1)R^2 = tR^2 + R^2 \geqslant \|w(t)\|^2 + R^2 \\ R \geqslant \|x_n\| \rightarrow R^2 \geqslant \|x_n\|^2 \end{array}\right\} \rightarrow$$

$$\Rightarrow (t+1)R^2 \geqslant \|w(t)\|^2 + \|x(t)\|^2 \geqslant \|w(t+1)\|^2$$

↳ $\|w(t+1)\|^2 \leqslant (t+1)R^2 \rightarrow$ (2) is true at $t+1$

e) $\bullet \dfrac{w^T(t)}{\|w(t)\|} w^* \geq \sqrt{t} \cdot \dfrac{\rho}{R}$

From b, we have: $w^T(t) w^* \geq w^T(t-1) w^* + \rho$

$\hookrightarrow \dfrac{w^T(t) w^*}{\|w(t)\|} \geq \dfrac{w^T(t-1) w^* + \rho}{\|w(t)\|}$

$\geq \dfrac{(t-1)\rho + \rho}{\|w(t)\|} = \dfrac{t\rho}{\|w(t)\|}$ A $(w^T(t) w^* \geq t\rho)$

From d, we have $\|w(t)\|^2 \leq tR^2 \rightarrow \dfrac{1}{\|w(t)\|^2} \geq \dfrac{1}{tR^2}$

$\rightarrow \dfrac{1}{\|w(t)\|} \geq \dfrac{1}{\sqrt{t} R}$ B

A, B $\rightarrow \dfrac{w^T(t) w^*}{\|w(t)\|} \geq \dfrac{t\rho}{\sqrt{t} R} = \sqrt{t} \dfrac{\rho}{R}$

- $t \leqslant \dfrac{R^2 \, \|w^*\|^2}{\rho^2}$

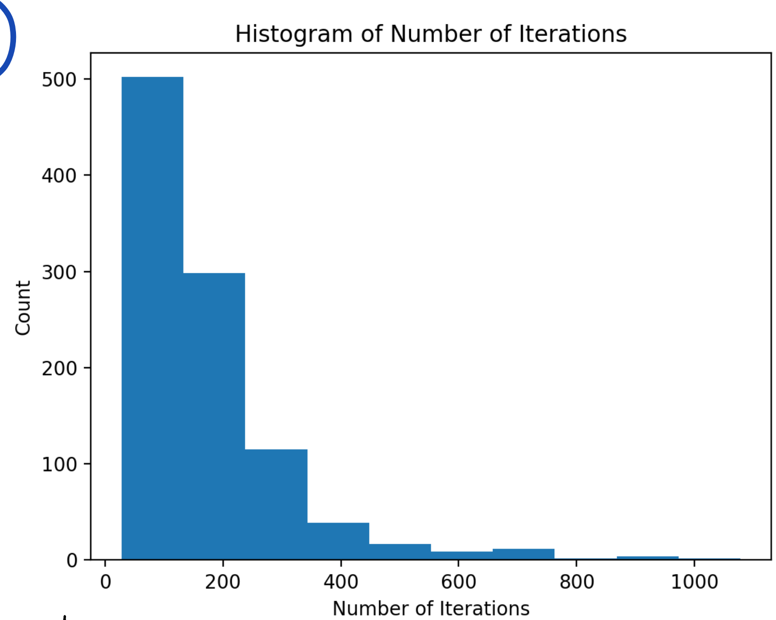
Assume θ is the angle between $w^T(t)$ and w^*

$\hookrightarrow \cos\theta = \dfrac{w^T(t)\, w^*}{\|w^T(t)\| \; \|w^*\|} \leqslant 1$

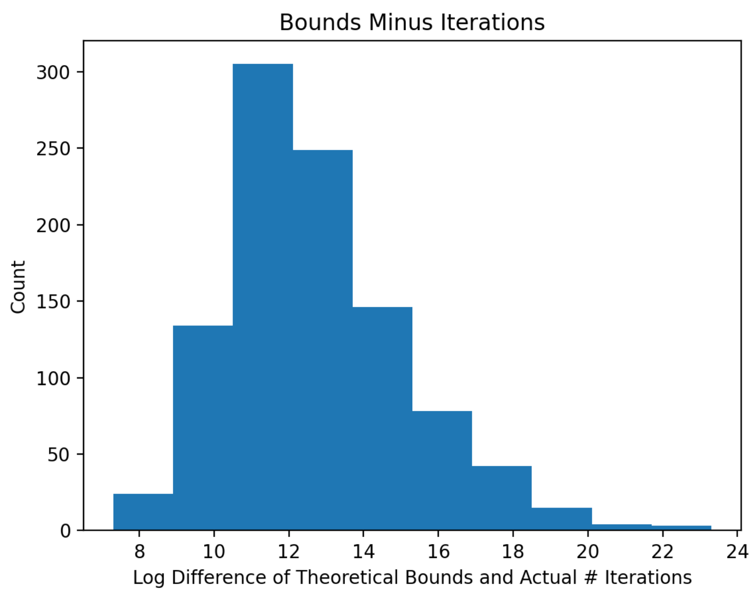
$$\Rightarrow \quad \|w^*\| \geqslant \frac{w^T(t) w^*}{\|w^T(t)\|} \geqslant \sqrt{t}\,\frac{\rho}{R}$$

$$\Rightarrow \quad \frac{\|w^*\|^2 \, R^2}{\rho^2} \geqslant t$$

2



- Histogram of Number of Iterations: looked normal with some right skew



- Bounds Minus Iterations: the data seemed normal with some right skew. The theoretical bounds are much larger than the actual number of iterations (Log Difference ≈ 11)

3 LFO Exercise 1.10

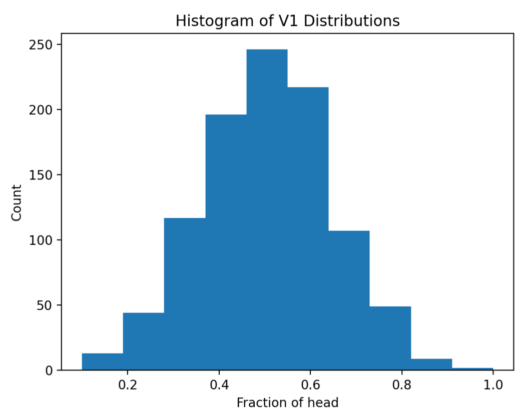
a, Since μ = true probability of flipping heads

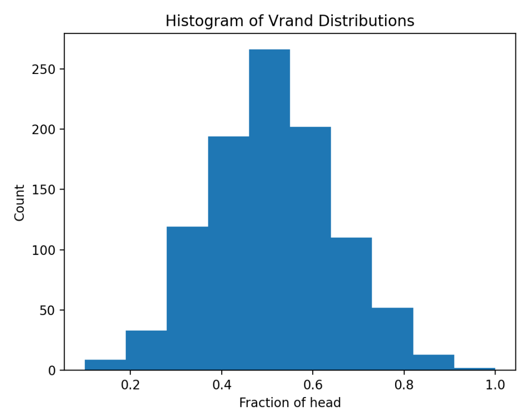
↳ P[heads] = 0.5

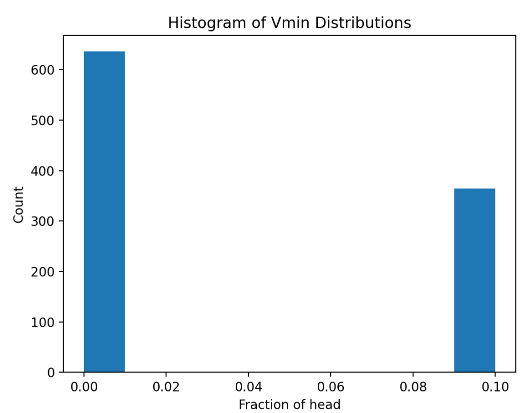
c_1, c_{rand}, c_{min} are all fair coins

↳ $\mu_1 = \mu_{rand} = \mu_{min} = 0.5$

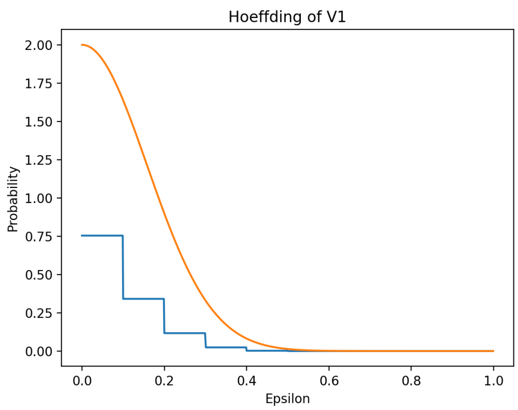
b,

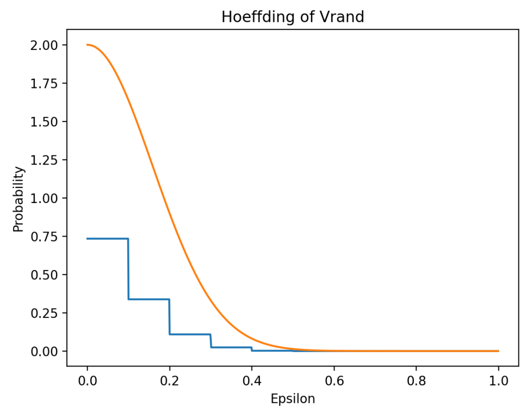


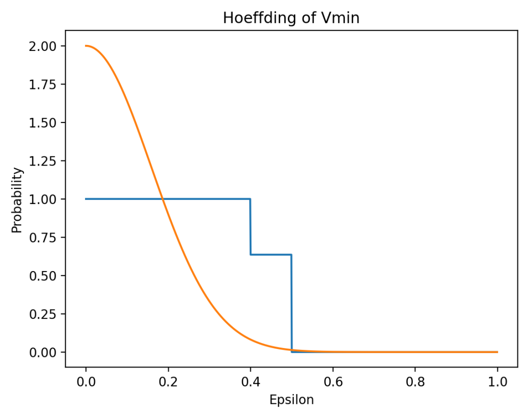




c,







d, V_1 and V_{rand} obey the Hoeffding bound since their histograms show normal distributions.

This is becaus c_1 and c_{rand} are picked randomly. Meanwhile, c_{min} is always the one with smallest value, no randomness

↳ V_{min} does not obey Hoeffding bound

4 LFD Problem 1.8

a, Prove $P[t \geqslant \alpha] \leqslant \frac{E(t)}{\alpha}$

$$E(t) = \underbrace{E(t \mid t \geqslant \alpha)}_{\geqslant \alpha \ (t \geqslant 0)} P(t \geqslant \alpha) + \underbrace{E(t \mid t < \alpha)}_{\geqslant 0 \ (t \geqslant 0)} P(t < \alpha)$$

$$\hookrightarrow E(t) = E(t \mid t \geqslant \alpha) P(t \geqslant \alpha) + 0 \, P(t < \alpha) \geqslant \alpha P(t \geqslant \alpha)$$

$$\Rightarrow P(t \geqslant \alpha) \leqslant \frac{E(t)}{\alpha}$$

b, Prove $P[(u-\mu)^2 \geqslant \alpha] \leqslant \frac{\sigma^2}{\alpha}$

From a $\rightarrow$ $$P[(u-\mu)^2 \geqslant \alpha] \leqslant \frac{E[(u-\mu)^2]}{\alpha} = \frac{E[u^2 - 2u\mu + \mu^2]}{\alpha}$$

$$= \frac{[E(u^2) - 2E(u)\mu + \mu^2]}{\alpha} = \frac{E(u^2) - \mu^2}{\alpha}$$

$$= \frac{E(u^2) - E^2(u)}{\alpha} = \frac{\sigma^2}{\alpha}$$

c, Prove $P[(u-\mu)^2 \geqslant \alpha] \leqslant \frac{\sigma^2}{N\alpha}$

$$E(u) = \frac{1}{N} \sum_{n=1}^{N} \mu = \mu \quad ; \quad \sigma^2(u) = \frac{1}{N^2} \sum_{n=1}^{N} \sigma^2 = \frac{\sigma^2}{N}$$

$$\Rightarrow P[(u-\mu)^2 \geqslant \alpha] \leqslant \frac{\sigma^2}{N\alpha}$$

5 LFD Problem 1.12

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

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

Since $\frac{d\,E_{in}(h)}{dh} = 0 \rightarrow h = \frac{1}{N}\sum_{n=1}^{N} y_n = h_{mean}$

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

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

$\rightarrow h = h_{med}$ as $\frac{d\,E_{in}(h)}{dh} = 0 \Leftrightarrow$ number of positive terms = number of negative terms

c, When $y_N \rightarrow \infty$, we get $h_{mean} \rightarrow \infty$ and h_{med} remains unchanged

6 LFD Problem 2.3

a) H = positive/negative rays $\rightarrow m_H(N) = N + 1$

- when all points are $+1$ & all points are -1 are overlap

$\Rightarrow m_H(N) = 2(N+1) - 2 = 2N$

$$\left.\begin{aligned} &\Rightarrow m_H(N) = 2N \\ &\text{Since } m_H(N) \leqslant 2^N \end{aligned}\right\} \rightarrow N = 2 \rightarrow \boxed{d_{vc} = 2}$$

b) H = positive intervals $\rightarrow m_H(N) = \frac{N^2}{2} + \frac{N}{2} + 1$

- 2 cases are counted double.

$$\underbrace{\circ\circ\circ\circ\circ\circ}_{a}\underbrace{\times\times\times\times\times\times}_{b} \qquad \underbrace{\times\times\times\times\times\times}_{a}\underbrace{\circ\circ\circ\circ\circ\circ}_{b} \qquad a+b=N$$

$\rightarrow m_H(N) = \frac{N^2}{2} + \frac{N}{2} + 1 + N - 2 = \frac{N^2}{2} + \frac{3N}{2} - 1 \leqslant 2^N$

$\rightarrow N = 3 \rightarrow \boxed{d_{vc} = 3}$

c) $a \leqslant (x_1^2 + \dots + x_d^2)^{\frac{1}{2}} \leqslant b$

This can be equivalent to positive intervals in R

$$\left.\begin{aligned} &\rightarrow m_H(N) = \frac{N^2}{2} + \frac{N}{2} + 1 \\ &m_H(N) \leqslant 2^N \end{aligned}\right\} \rightarrow N = 2 \rightarrow \boxed{d_{vc} = 2}$$