

Homework 1

Runmin Lu

September 13, 2021

Exercise 1.8

$$\binom{10}{0} \left(\frac{1}{10}\right)^{10} + \binom{10}{1} \left(\frac{1}{10}\right)^9 \left(\frac{9}{10}\right)^1 = \boxed{9.1 \times 10^{-9}}$$

Exercise 1.9

$$\begin{aligned} P(\nu \leq 0.1) &= P(|\nu - \mu| \geq 0.8) \\ &\leq 2e^{-2(0.8)^2(10)} \\ &\approx \boxed{5.52 \times 10^{-7}} \end{aligned}$$

The answer is much greater than the answer to the previous exercise but that's fine because it's an upper bound.

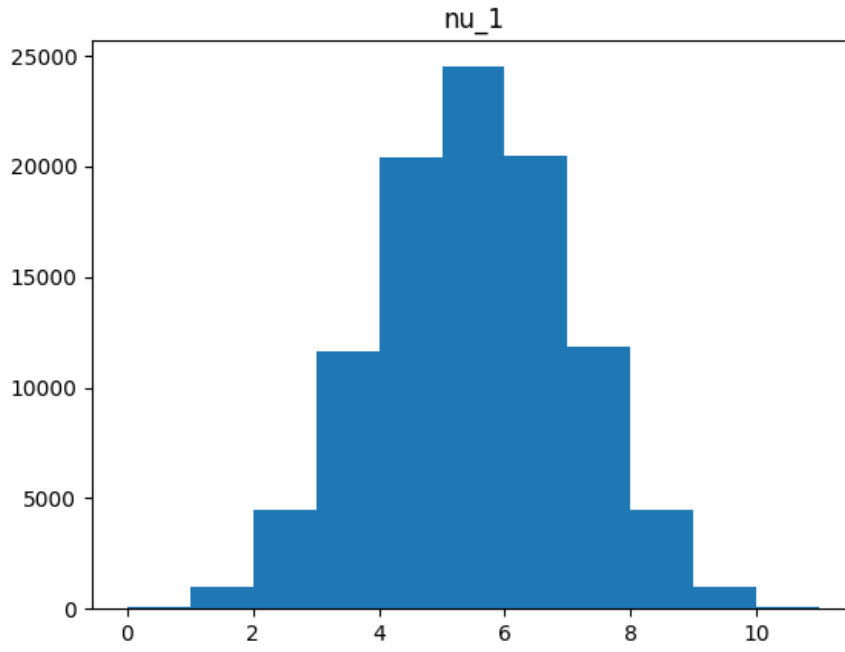
Exercise 1.10

(a)

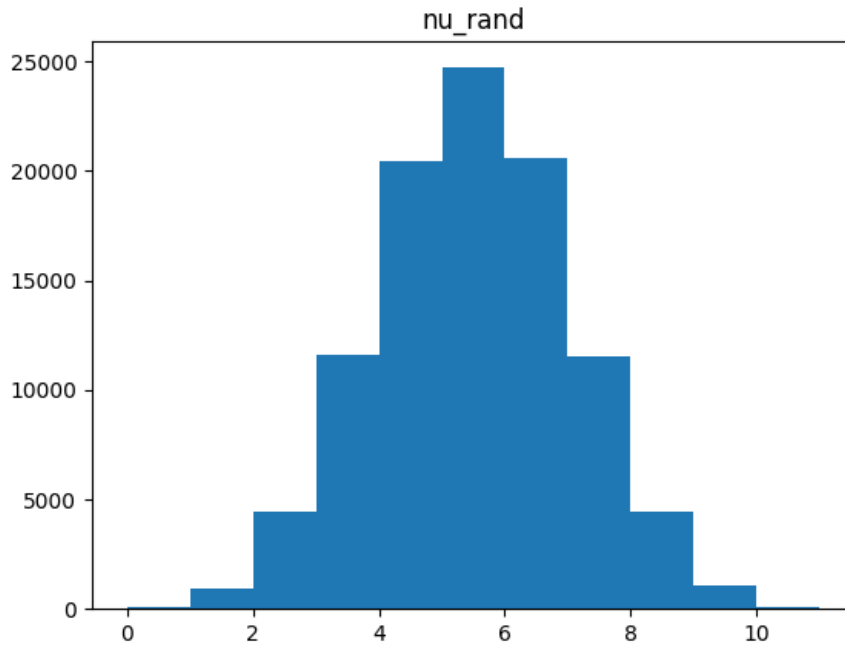
$$\mu = \frac{1}{2}$$

(b)

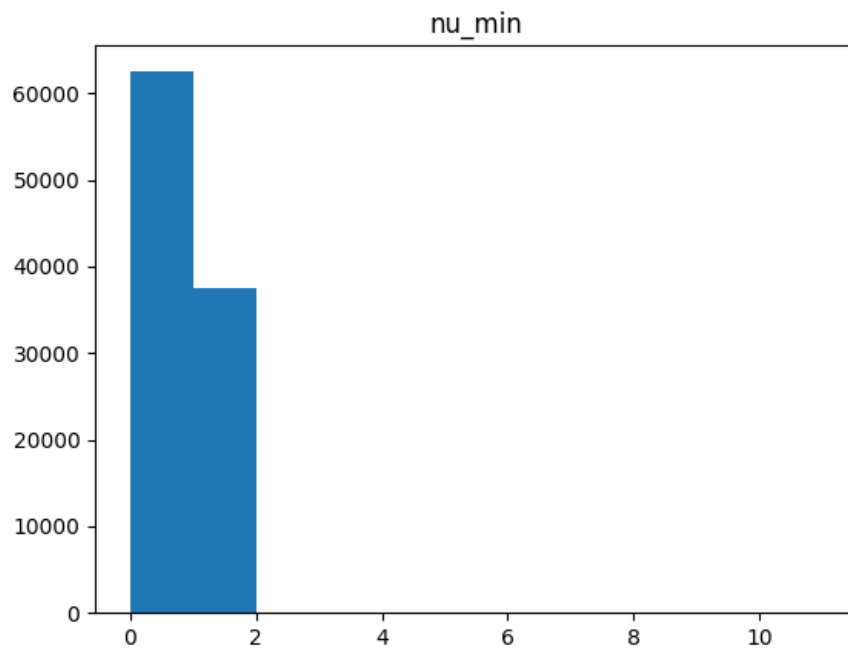
Counts of ν_1 : [101, 1020, 4472, 11673, 20430, 24504, 20470, 11830, 4437, 955, 108]



Counts of ν_{rand} : [95, 924, 4431, 11639, 20450, 24727, 20582, 11554, 4439, 1046, 113]



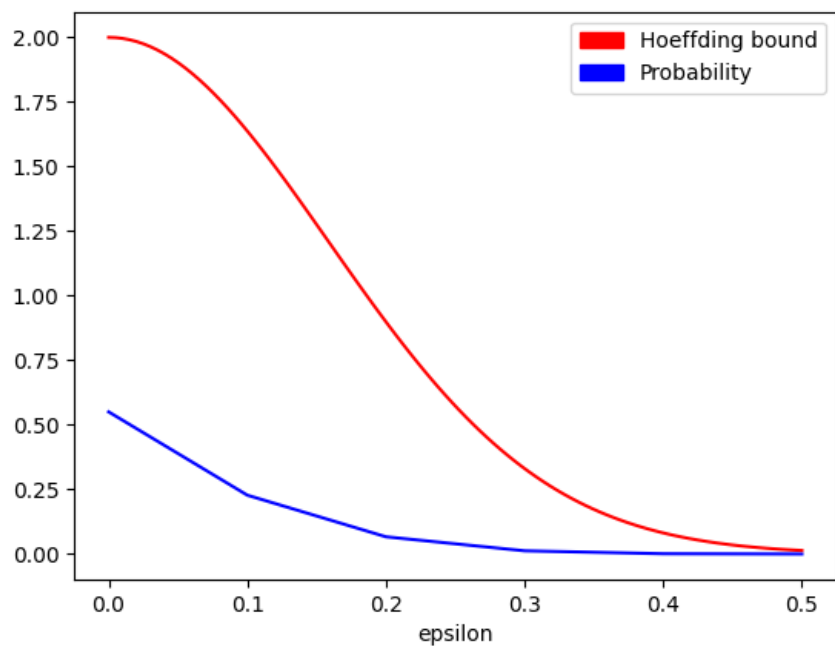
Counts of ν_{min} : [62487, 37510, 3, 0, 0, 0, 0, 0, 0, 0, 0]



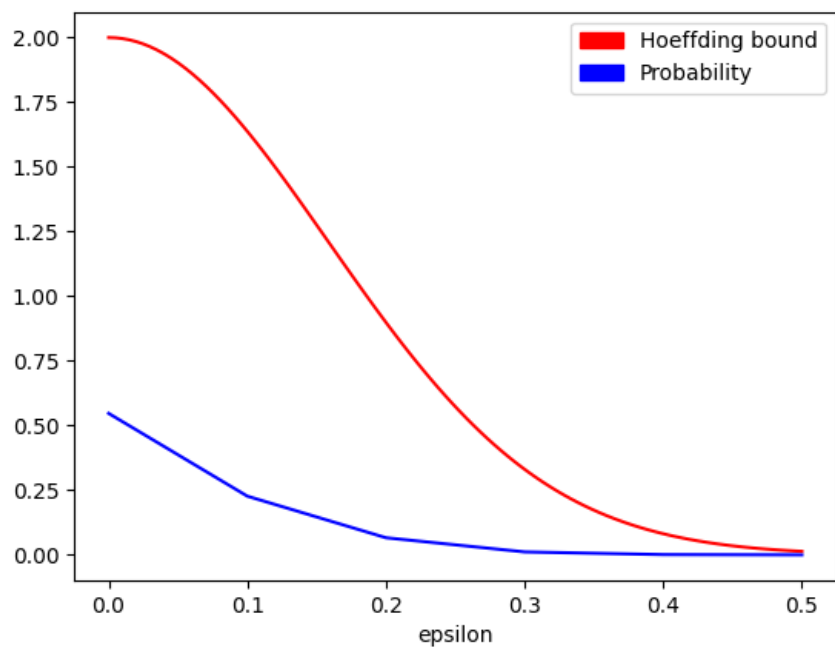
(c)

Note: the following plots were supposed to look like the hand-drawn one for the last question but I don't know how to plot step functions in python. Each line segment is supposed to stay flat starting on the left point. If you get the idea please don't take off points. Thanks.

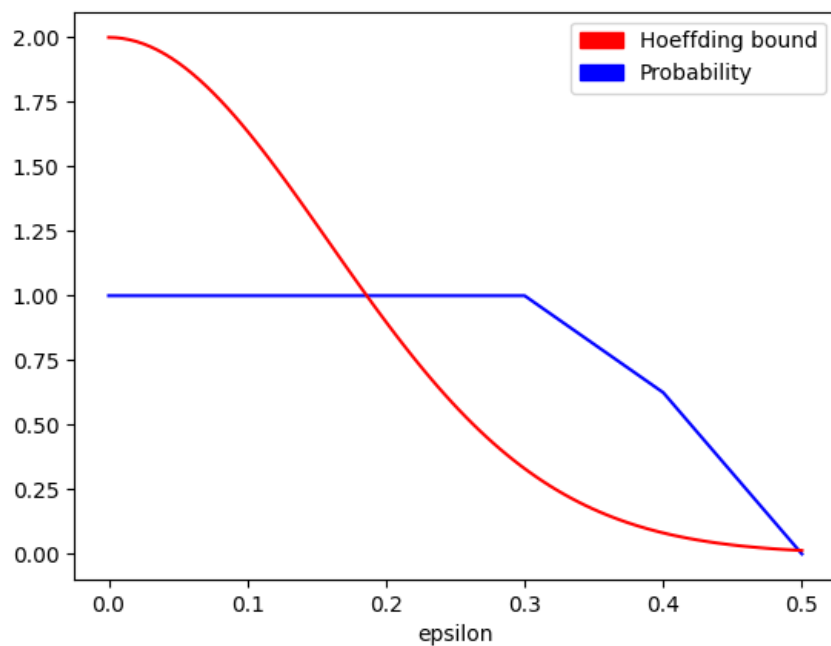
For c_1 :



For c_{rand} :



For c_{min} :



(d)

c_1 and c_{rand} obey the Hoeffding bound and c_{min} does not because for c_1 and c_{rand} the probability curve is always below the Hoeffding bound curve. This is not true for c_{min} .

(e)

c_1 and c_{rand} are selected from the bin of all the 1000 coins while c_{min} is selected from the bin of coins with minimum number of heads.

Exercise 1.11

(a)

No because we can never draw conclusion outside the data for sure.

(b)

Yes for the same reason above.

(c)

$$P(\text{more } +1\text{'s than } -1\text{'s in } D) = \sum_{n=13}^{25} \binom{25}{n} \times 0.9^n \times 0.1^{25-n} \approx \boxed{1}$$

(d)

No because D is a random sample, whose expected frequency of $+1$'s always equals p . The hypothesis produced by S will therefore always be more likely to agree with inputs out of D .

Exercise 1.12

The best I can promise is (c) because in machine learning it's always possible to fail.

Problem 1.3

(a)

Given that \mathbf{w}^* separates the data, we can conclude that $\forall n : y_n = \text{sign}(\mathbf{w}^{*T} \mathbf{x}_n)$.

Therefore, $\forall n : y_n(\mathbf{w}^{*T} \mathbf{x}_n) > 0$, which means $\min_n y_n(\mathbf{w}^{*T} \mathbf{x}_n) > 0$.

(b)

$$\begin{aligned} \mathbf{w}^T(t) \mathbf{w}^* &= (\mathbf{w}^T(t-1) + y_i \mathbf{x}_i^T) \mathbf{w}^* \text{ for some } i \\ &= \mathbf{w}^T(t-1) \mathbf{w}^* + y_i \mathbf{x}_i^T \mathbf{w}^* \\ &\geq \mathbf{w}^T(t-1) \mathbf{w}^* + \min_i y_i \mathbf{x}_i^T \mathbf{w}^* \\ &= \mathbf{w}^T(t-1) \mathbf{w}^* + \rho \end{aligned}$$

Proof of $\mathbf{w}^T(t) \mathbf{w}^* \geq t\rho$ by Induction:

Base Case $t = 0$:

$$\mathbf{w}^T(0) \mathbf{w}^* = 0 \geq 0\rho = 0$$

Inductive Step:

Assume $\mathbf{w}^T(t)\mathbf{w}^* \geq t\rho$

$$\begin{aligned}\mathbf{w}^T(t+1)\mathbf{w}^* &\geq \mathbf{w}^T(t)\mathbf{w}^* + \rho \\ &\geq t\rho + \rho \\ &= (t+1)\rho\end{aligned}$$

(c)

Given $\mathbf{w}(t) = \mathbf{w}(t-1) + y(t-1)\mathbf{x}(t-1)$

$$\|\mathbf{w}(t)\|^2 = \|\mathbf{w}(t-1)\|^2 + 2y(t-1)\mathbf{w}^T(t-1)\mathbf{x}(t-1) + y^2(t-1)\|\mathbf{x}(t-1)\|^2$$

Knowing $y(t-1)\mathbf{w}^T(t-1)\mathbf{x}(t-1) \leq 0$

$$\begin{aligned}\|\mathbf{w}(t)\|^2 &\leq \|\mathbf{w}(t-1)\|^2 + y^2(t-1)\|\mathbf{x}(t-1)\|^2 \\ &= \|\mathbf{w}(t-1)\|^2 + \|\mathbf{x}(t-1)\|^2\end{aligned}$$

(d)

Base Case $t = 0$:

$$\|\mathbf{w}(0)\|^2 = \|\mathbf{0}\|^2 = 0 \geq 0R^2 = 0$$

Inductive Step:

Assume $\|\mathbf{w}(t)\|^2 \leq tR^2$

$$\begin{aligned}\|\mathbf{w}(t+1)\|^2 &\leq \|\mathbf{w}(t)\|^2 + \|\mathbf{x}(t)\|^2 \\ &\leq tR^2 + (\max_n \|\mathbf{x}_n\|)^2 \\ &= tR^2 + R^2 \\ &= (t+1)R^2\end{aligned}$$

(e)

From part (b): $\mathbf{w}(t)\mathbf{w}^* \geq t\rho$

From part (d): $\|\mathbf{w}(t)\|^2 \leq tR^2$

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

$$\begin{aligned}\frac{\mathbf{w}^T(t)\mathbf{w}^*}{\|\mathbf{w}(t)\|} &\geq \frac{t\rho}{R\sqrt{t}} \\ &= \frac{\rho}{R}\sqrt{t}\end{aligned}$$

Cauchy-Shwartz Inequality: $\mathbf{w}^T(t)\mathbf{w}^* \leq \|\mathbf{w}(t)\| \|\mathbf{w}^*\|$

$$\begin{aligned}\frac{\mathbf{w}^T(t)\mathbf{w}^*}{\|\mathbf{w}(t)\|} &\leq \|\mathbf{w}^*\| \\ \frac{\rho}{R}\sqrt{t} &\leq \|\mathbf{w}^*\| \\ t &\leq \frac{R^2\|\mathbf{w}^*\|^2}{\rho^2}\end{aligned}$$

Problem 1.7

(a)

$\mu = 0.05$:

$$\begin{aligned}P(\text{single coin has } \nu > 0) &= 1 - \binom{10}{0} \times (1 - 0.05)^{10} \\ &\approx 0.401\end{aligned}$$

$$P(\text{some coin has } \nu = 0) = 1 - P(\text{all coins have } \nu > 0)$$

$$\text{For 1 coin: } P(\text{some coin has } \nu = 0) \approx 1 - 0.401$$

$$= \boxed{0.599}$$

$$\text{For 1000 coins: } P(\text{some coin has } \nu = 0) \approx 1 - 0.401^{1000}$$

$$\approx \boxed{1}$$

$$\text{For 1000000 coins: } P(\text{some coin has } \nu = 0) \approx 1 - 0.401^{1000000}$$

$$\approx \boxed{1}$$

$\mu = 0.8$:

$$P(\text{single coin has } \nu > 0) = 1 - \binom{10}{0} \times (1 - 0.8)^{10} \\ \approx 0.99999990$$

$$P(\text{some coin has } \nu = 0) = 1 - P(\text{all coins have } \nu > 0)$$

$$\text{For 1 coin: } P(\text{some coin has } \nu = 0) \approx 1 - 0.99999990$$

$$\approx \boxed{0}$$

$$\text{For 1000 coins: } P(\text{some coin has } \nu = 0) \approx 1 - 0.99999990^{1000}$$

$$\approx \boxed{1.02 \times 10^{-4}}$$

$$\text{For 1000000 coins: } P(\text{some coin has } \nu = 0) \approx 1 - 0.99999990^{1000000}$$

$$\approx \boxed{9.73 \times 10^{-2}}$$

(b)

$$N = 6 \implies \nu \in \{0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1\}$$

$$P(\max_i |\nu_i - \mu_i| > \epsilon) = 1 - P(|\nu_1 - \mu_1| \leq \epsilon)P(|\nu_2 - \mu_2| \leq \epsilon)$$

$$P(\max_i |\nu_i - \mu_i| > 0) = 1 - P(\nu_1 = \frac{1}{2})P(\nu_2 = \frac{1}{2})$$

$$= 1 - \left(\binom{6}{3}\left(\frac{1}{2}\right)^6\right)^2$$

$$\approx 0.902$$

$$P(\max_i |\nu_i - \mu_i| > \frac{1}{6}) = 1 - P(\nu_1 \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\})P(\nu_2 \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\})$$

$$= 1 - \left(\binom{6}{2}\left(\frac{1}{2}\right)^6 + \binom{6}{3}\left(\frac{1}{2}\right)^6 + \binom{6}{4}\left(\frac{1}{2}\right)^6\right)^2$$

$$\approx 0.390$$

$$P(\max_i |\nu_i - \mu_i| > \frac{1}{3}) = 1 - P(\nu_1 \in \{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}\})P(\nu_2 \in \{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}\})$$

$$= 1 - \left(\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5}\right)\left(\frac{1}{2}\right)^6\right)^2$$

$$\approx 0.062$$

$$P(\max_i |\nu_i - \mu_i| > \frac{1}{2}) = 0$$

