

Assignment 2

Darwin Ding

September 14, 2016

Exercise 1.8

The probability of picking a marble out of the bag and getting red is 0.9, and we are looking for the probability that we sample 10 random marbles and get 1 or 0 red marbles.

The probability of getting 0 red marbles is $(.1)^{10} = 10^{-10}$, and the probability of getting 1 red marble is $(.1)^9 * .9 * 10$, because we pulled 9 green marbles (with probability .1) and 1 red marble (probability .9), and there were 10 ways to do that.

The probability of doing either of those things is the sum of those two probabilities is **$9.1 * 10^{-9}$**

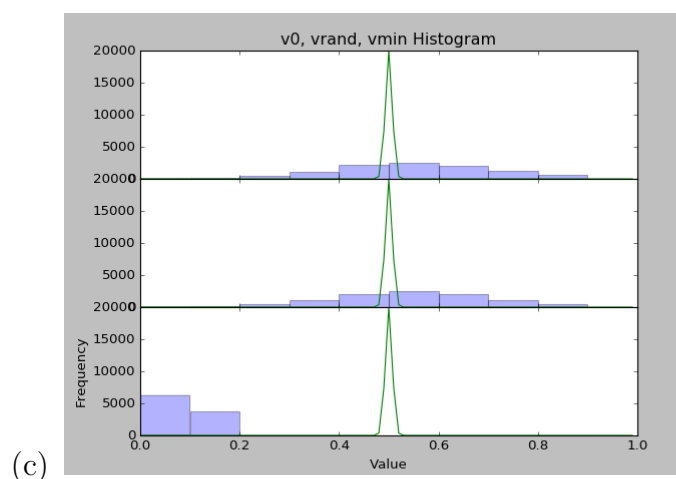
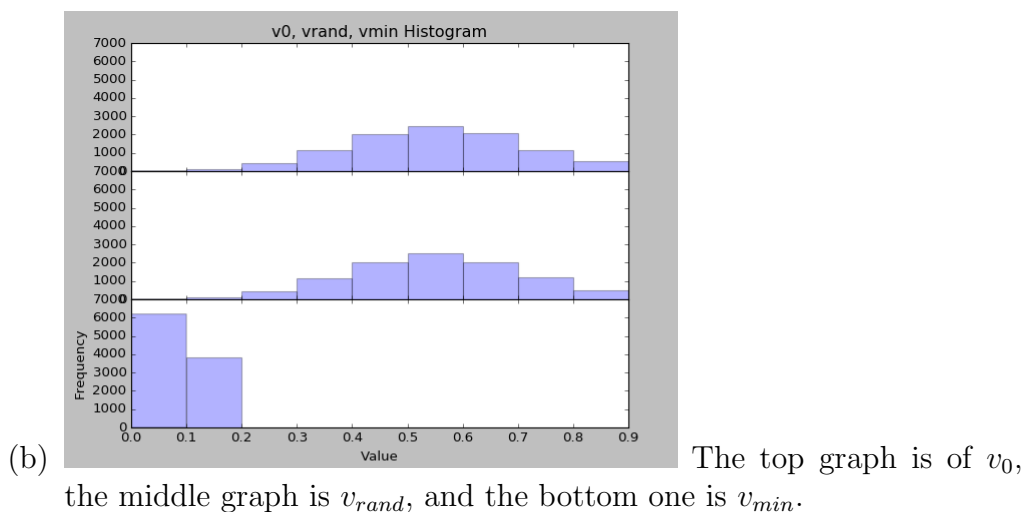
Exercise 1.9

Using Hoeffding's Inequality, $\mu = .9$. Since we're looking for $v \leq .1$, $\epsilon \geq .9 - .1 = .8$. By definition, also, $N = 10$.

$$\begin{aligned} P[|v - \mu| \geq \epsilon] &\leq 2e^{-2\epsilon^2 N} \\ &\leq 2e^{-2 * .8^2 * 10} \\ &\leq \mathbf{5.52 * 10^{-6}} \end{aligned}$$

Exercise 1.10 (needs checking)

- (a) After running the experiment a single time, μ for $v_0 = \mathbf{0.3}$, μ for $v_{rand} = \mathbf{0.5}$ and μ for $v_{min} = \mathbf{0}$



- (d) v_0 and v_{rand} both obey the Hoeffding bound, while v_{min} does not. This makes sense, because v_0 and v_{rand} are essentially randomly selected coins. There's nothing special about the first coin or any random particular coin. The problem arises when you specifically seek out the minimum coin and then try to retroactively perform the Hoeffding bound. Hoeffding's inequality doesn't support this.
- (e) If we pretend we don't know the true nature of coin flipping, and that μ is unknown to us, this problem is very similar to the bins problem. Let's say we are trying to predict μ using our data (the experiments we've performed and their results), and our hypothesis set is $[v_0, v_{rand}, v_{min}]$.

If we fix our hypothesis set and then perform the experiment, we find that the Hoeffding bounds work simply because they work for $[v_0, v_{rand}]$. We cannot, however, claim that the bounds don't work simply because they retroactively don't work for v_{min} .

Exercise 1.11 (check d)

- (a) No, not necessarily. If $p = 0.5$, then even if E_{in} is approximately equal to E_{out} , $E_{in} = 0.5$ which is just as good as random.
- (b) Yes. Just because the smart algorithm has the smallest in-sample error (even if it's 0) doesn't mean there's a chance that the out-of-sample data is completely different from the in-sample. In this case, all of the points out of D might be -1, and the crazy learning algorithm will have worked better.
- (c) If $p = 0.9$, S will pick the "correct" hypothesis unless $E_{in}(h_1) > 0.5$. The probability of that is the probability that $\epsilon > .4$ ($E_{out}(h_1) = .1$). $M = 2$, so this probability is bounded by $2 * 2 * e^{-2\epsilon^2 N} = .0013$. Therefore the probability that S correctly hypothesizes from the set is at least $1 - .0013 = \mathbf{0.99866}$
- (d) For any p , $E_{in}(h_1) = 1 - p$. Without loss of generality, let's assume that $p > 0.5$ (this could all be flipped and applied to the opposite case). $E_{out}(h_1)$ needs to be greater than 0.5 with probability greater than 0.5 for this to be true.

Solving for $2 * 2 * e^{-2(p-0.5)^2 * 25} > 0.5$ gets us $\mathbf{0.296 < p < 0.704}$

Exercise 1.12

The best you can provide is (c). Depending on the problem, you may need to have a very large hypothesis set, which would decrease your out-of-bound prediction accuracy. In that scenario, you may have to announce that you have failed. However, if you find that E_{out} is not close to 0.5, then you can announce that you have found a function that approximates well out-of-sample.