Group Exercise 1 Question B

The proof fails to normalize the input variables when calculating the Hoeffding Bound. Hoeffding bound expects the variables to be between 0 and 1, but in the provided proof it's calculated with variables between 1 and M. To correct this the generalized Hoeffding Bound equation should be used.

Example

Here we present an example array that fails with the suggested algorithm.

Let $A = [1, 1, 1, 1, 1, 1, 1, 1, 1, 100]$ (we fill an array of size 10 with 9 ones and one item with the value 100). In this case the probability of not finding the single element with value 100 is:

$$\left(1 - \frac{1}{10}\right)^{s} = \left(1 - \frac{1}{10}\right)^{\frac{1}{\epsilon^2}}$$

When that happens (not finding the single 100) the algorithm will give the answer $n(= 10)$. We use the left hand size of the bounds provided in the proof to explore the size of ϵ (remember A here is the average value of the array):

$$nA(1 - \epsilon) \leq sum \leq nA(1 + \epsilon)$$

$$\begin{aligned} 10\frac{109}{10}(1 - \epsilon) &\leq 10 \\ 109(1 - \epsilon) &\leq 10 \\ \frac{99}{109} &\leq \epsilon \end{aligned}$$

And when ϵ is at least $\frac{99}{109}$ the probability of error (i.e. not finding the single 100 in the array) will be at least:

$$\left(1 - \frac{1}{10}\right)^{\frac{1}{\frac{99}{109}^2}} \approx 0.88...$$

which is much bigger than the probability of error in the suggested proof ($\frac{1}{3}$).