Three Proofs of a Simple Lemma

We give three proofs of a claim from the textbook:

Claim 2.5.2.1: There exists a set $S_n \subseteq \{0,1\}^n$ of cardinality at least $\frac{\varepsilon(n)}{2} \cdot 2^n$ such that for every $x \in S_n$, it holds that

$$s(x) \stackrel{\mathrm{df}}{=} \Pr[G(f(x), R_n) = b(x, R_n)] \geq \frac{1}{2} + \frac{\varepsilon(n)}{2}$$

This claim is stated in a rather awkward way. Instead of existentially quantifying S_n, it is simpler to just define it in terms of $s(\cdot)$, namely,

Definition:

$$S_n = \left\{ x \in \{0,1\}^* \mid s(x) \geq \frac{1}{2} + \frac{\varepsilon(n)}{2} \right\}.$$

Then the claim we are trying to prove follows from the slightly stronger

Lemma 1

$$|S_n| \geq \varepsilon(n) \cdot 2^n.$$

We will need one further fact about $s(\cdot)$.

Fact

$$E(s(X_n)) = \frac{1}{2} + \varepsilon(n).$$

This follows immediately from the definition of $\varepsilon(n)$ given in the book.

We give three proofs of the lemma—one algebraic, one geometric, and one using Markov's inequality.

1 Algebraic proof

The algebraic proof relies on the definition of expectation, namely, that

$$E(s(X_n)) = \sum_x \Pr[X_n = x] \cdot s(x) = 2^{-n} \sum_x s(x).$$

The key idea is to split the sum into two parts, those terms where $x \in S_n$ and those terms where $x \notin S_n$. Towards this end, define $\overline{S}_n = \{0,1\}^* - S_n$. We then have

$$\begin{aligned} \frac{1}{2} + \varepsilon(n) &= E(s(X_n)) = 2^{-n} \sum_x s(x) \\ &= 2^{-n} \left(\sum_{x \in S_n} s(x) + \sum_{x \in \overline{S}_n} s(x) \right) \end{aligned}$$

$$
\begin{aligned}
&\leq\ 2^{-n}\left(|S_n|+\sum_{x\in\overline{S}_n}\left(\frac{1}{2}+\frac{\varepsilon(n)}{2}\right)\right)\\
&=\ 2^{-n}\left(|S_n|+|\overline{S}_n|\left(\frac{1}{2}+\frac{\varepsilon(n)}{2}\right)\right)\\
&=\ 2^{-n}\left(|S_n|+(2^n-|S_n|)\left(\frac{1}{2}+\frac{\varepsilon(n)}{2}\right)\right)\\
&=\ \frac{1}{2}+\frac{\varepsilon(n)}{2}+2^{-n}\left(|S_n|-|S_n|\left(\frac{1}{2}+\frac{\varepsilon(n)}{2}\right)\right)\\
&\leq\ \frac{1}{2}+\frac{\varepsilon(n)}{2}+2^{-n}\left(\frac{1}{2}|S_n|\right).
\end{aligned}
$$

Subtracting $1/2+\varepsilon(n)/2$ from both sides, we have

$$\frac{\varepsilon(n)}{2}\leq\frac{1}{2}\cdot\frac{|S_n|}{2^n}$$

from which it follows that

$$|S_n|\geq\varepsilon(n)\cdot 2^n$$

as desired.

2 Geometric proof

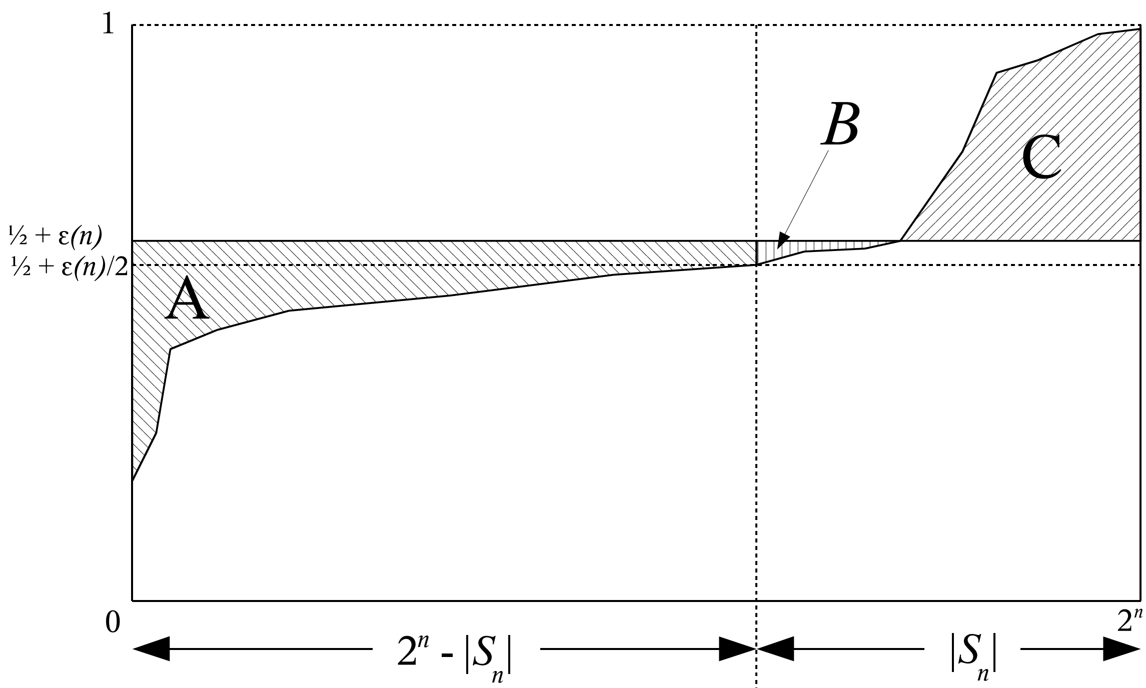

Figure 1: Graph of the function $s(x)$.

The geometric proof is based on an analysis of the graph of the function $s(x)$. Assume that the domain of $s(\cdot)$ has been ordered so as to make $s(\cdot)$ non-decreasing. Then the graph of $s(\cdot)$ looks like the diagram of Figure 1. I have drawn a solid horizontal line at $y = 1/2 + \varepsilon(n) = E(s(X_n))$. This is the average value of $s(\cdot)$ over its domain. Hence, the area above the curve and below this line (regions A and B in the diagram) is the same as the area above the line and below the curve (region C in the diagram).

I have drawn a second horizontal line at $y = 1/2 + \varepsilon(n)/2$. This is the defining threshold for the set S_n. I have drawn a vertical dashed line through the point where it intersects the curve. Values of x to the right of this line are in S_n, and those to the left are not. The goal is to prove that the line cannot be too far to the right (so that S_n isn't too small).

The proof is now fairly straightforward. First of all, as noted before, we have

$$A + B = C.$$

Clearly, region A includes the skinny rectangle between the two horizontal lines. It has height $\varepsilon(n)/2$ and width $2^n - |S_n|$. Hence,

$$A \geq \frac{\varepsilon(n)}{2}(2^n - |S_n|).$$

Region C is entirely contained within the upper right hand rectangle of height $1/2 - \varepsilon(n)$ and width $|S_n|$. Hence,

$$C \leq \left(\frac{1}{2} - \varepsilon(n)\right) \cdot |S_n|.$$

Combining these facts, we have

$$\begin{aligned}\left(\frac{1}{2} - \varepsilon(n)\right) \cdot |S_n| \quad &\geq \quad C = A + B \geq A \\ &\geq \quad \frac{\varepsilon(n)}{2}(2^n - |S_n|)\end{aligned}$$

Therefore,

$$\left(\frac{1}{2} - \frac{\varepsilon(n)}{2}\right) \cdot |S_n| \geq \frac{\varepsilon(n)}{2} \cdot 2^n.$$

Solving for $|S_n|$, we get

$$|S_n| \geq \frac{\varepsilon(n) \cdot 2^n}{2\left(\frac{1}{2} - \frac{\varepsilon(n)}{2}\right)} \geq \varepsilon(n) \cdot 2^n.$$

3 A proof using Markov's inequality

Recall Markov's inequality:

$$\Pr[X \geq v] \leq \frac{E(X)}{v}.$$

The proof using Markov's inequality applies the inequality to the random variable $1 - s(X_n)$ to obtain

$$\Pr\left[1 - s(X_n) \geq \frac{1}{2} - \frac{\varepsilon(n)}{2}\right] \leq \frac{E(1 - s(X_n))}{\frac{1}{2} - \frac{\varepsilon(n)}{2}}.$$

It uses the fact that

$$\Pr\left[s(X_n) \geq \frac{1}{2} + \frac{\varepsilon(n)}{2}\right] = \Pr[X_n \in S_n] = \frac{|S_n|}{2^n}.$$

Hence, to prove our lemma, we establish a lower bound on this quantity.

The calculation is an exercise in change of signs and negation of events.

$$\begin{aligned}\Pr\left[s(X_n) \geq \frac{1}{2} + \frac{\varepsilon(n)}{2}\right] &= 1 - \Pr\left[s(X_n) < \frac{1}{2} + \frac{\varepsilon(n)}{2}\right] \\ &= 1 - \Pr\left[1 - s(X_n) > \frac{1}{2} - \frac{\varepsilon(n)}{2}\right].\end{aligned}$$

We apply Markov's inequality to get

$$\begin{aligned}1 - \Pr\left[1 - s(X_n) > \frac{1}{2} - \frac{\varepsilon(n)}{2}\right] &\geq 1 - \frac{E(1 - s(X_n))}{\frac{1}{2} - \frac{\varepsilon(n)}{2}} \\ &= 1 - \frac{1 - \left(\frac{1}{2} + \varepsilon(n)\right)}{\frac{1}{2} - \frac{\varepsilon(n)}{2}} \\ &= \frac{\varepsilon(n)}{1 - \varepsilon(n)} \\ &\geq \varepsilon(n).\end{aligned}$$

Thus,

$$\frac{|S_n|}{2^n} \geq \varepsilon(n)$$

and the lemma follows.