Today’s exercises

• 2.9: Events instead of clauses

• (2.10: A somewhat tighter version)

• 2*.1: Random Sampling Doesn’t Work

• 2*.2: Fewer top-level invocations

• 2*.3: Even fewer top-level invocations

• 2*.4: The procedure stops
2.9. Events instead of clauses  To prove: Let $\mathcal{A} = \{A\}$ be a set of events s.t. for all $A \in \mathcal{A}$ we have $\Pr(A) \leq p$. $A$ depends on at most $d$ other events. If $4dp \leq 1$, then $\Pr(\wedge_{A \in \mathcal{A}} \overline{A}) > 0$.

Meaning that with positive probability no event happens.
Approach: Just repeat the proof of the Local Lemma and use...

- ... events instead of clauses (the events we want to avoid are of the sort 'clause $C$ gets violated')

- ... that each event is mutually independent from all but at most $d$ of the others

We will see that this suffices to repeat the proof. This (more general) probabilistic formulation immediately implies the one we proved for SAT.
Let $A = \{A_1, A_2, \ldots, A_n\}$ with $\Pr(A_i) \leq p$. Each $A_i$ depends on at most $d$ other events. If $4dp \leq 1$, then $\Pr(A_1 \land A_2 \land \cdots \land A_n) > 0$.

**Claim:** For $B \subseteq A$ and $A \in A \setminus B$ we have

$$\Pr\left(\overline{A} \mid \bigwedge_{B \in B} \overline{B}\right) \geq 1 - 2p.$$ 

The result follows immediately from the claim. Why?

$$\Pr\left(\overline{A_1} \land \overline{A_2} \land \cdots \land \overline{A_n}\right) = \prod_{i=1}^{n} \Pr\left(\overline{A_i} \mid \overline{A_1} \land \overline{A_2} \land \cdots \land \overline{A_{i-1}}\right) \geq (1 - 2p)^n > 0.$$

Here we only need the conditional probabilities to be positive. The lower bound is only used in proof of the claim itself!
Now we prove the claim:

\[
\Pr \left( \overline{A} \mid \bigwedge_{B \in \mathcal{B}} \overline{B} \right) \geq 1 - 2p, \]

equivalently

\[
\Pr \left( A \mid \bigwedge_{B \in \mathcal{B}} B \right) \leq 2p. \]

If $A$ doesn’t depend on $\mathcal{B}$ then the above probability is just $\Pr(A)$ and hence at most $p$.

Suppose by induction that the claim holds for all sets $\mathcal{B}'$ with $|\mathcal{B}'| \leq |\mathcal{B}|$. The base case $\mathcal{B}' = \emptyset$ is trivial.
We want to prove

\[ \Pr \left( A \mid \bigwedge_{B \in \mathcal{B}} \overline{B} \right) \leq 2p. \]

Let \( \mathcal{B}' \subseteq \mathcal{B} \) be the set of all events independent from \( A \). Hence

\[ \Pr \left( A \mid \bigwedge_{B \in \mathcal{B}'} \overline{B} \right) \leq p. \]

We will show

\[ \Pr \left( \bigwedge_{B \in \mathcal{B}} \overline{B} \right) \geq \frac{1}{2} \Pr \left( \bigwedge_{B \in \mathcal{B}'} \overline{B} \right), \]

which implies the claim.
We are left to show

\[ \Pr \left( \bigwedge_{B \in \mathcal{B}} \overline{B} \right) \geq \frac{1}{2} \Pr \left( \bigwedge_{B \in \mathcal{B}'} \overline{B} \right). \]

We have \(|\mathcal{B} \setminus \mathcal{B}'| \leq d\). For all \(E \in \mathcal{B} \setminus \mathcal{B}'\), we have by induction

\[ \Pr \left( E \mid \bigwedge_{B \in \mathcal{B}'} \overline{B} \right) \leq 2p, \]

so by union bound

\[ \Pr \left( \bigvee_{E \in \mathcal{B} \setminus \mathcal{B}'} E \mid \bigwedge_{B \in \mathcal{B}'} \overline{B} \right) \leq 2pd \leq \frac{1}{2}. \]

Given that all events in \(\mathcal{B}'\) do not occur, at least one event \(E\) in \(\mathcal{B} \setminus \mathcal{B}'\) occurs with probability at most \(\frac{1}{2}\). Hence no event \(E\) occurs with probability at least \(\frac{1}{2}\).
2.10. A somewhat tighter version

Two changes to the proof:

- On Page 29, replace all occurrences of $\frac{1}{2^{k-1}} = \frac{2}{2^k}$ by $\frac{\delta}{2^k}$, that is

$$\frac{|\text{sat}(C) \cap \text{sat}(G)|}{|\text{sat}(G)|} \leq \frac{\delta}{2^k}.$$ 

The claim becomes somewhat weaker if $\delta > 2$ (in the end, $\delta \approx e$)

- On Page 30, instead of union bound, use the induction hypothesis multiple times:

Let $G' =: G_0 \subset G_1 \subset G_2 \subset \ldots \subset G_d = G$, $d = |\Gamma_F(C)|$
2.10. A somewhat tighter version (2)

Now apply induction hypothesis:

\[ \text{sat}(G_{i+1}) = |\text{sat}(G_i) \setminus \text{sat}(G_{i+1} \setminus G_i)| \geq |\text{sat}(G_i)| \cdot (1 - \frac{\delta}{2^k}) \]

And so by choice of \( \delta \)

\[ |\text{sat}(G)| \geq |\text{sat}(G')| \cdot (1 - \frac{\delta}{2^k})^d \geq \frac{1}{\delta} |\text{sat}(G')| \]

Therefore

\[ \frac{ |\text{sat}(C') \cap \text{sat}(G)|}{|\text{sat}(G)|} \leq \frac{ |\text{sat}(C') \cap \text{sat}(G')|}{|\text{sat}(G)|} \leq \frac{ |\text{sat}(C') \cap \text{sat}(G')|}{|\text{sat}(G')|/\delta} = \frac{\delta}{2^k}. \]
2*.1: Random Sampling Doesn’t Work

Trivial solution: $m$ disjoint clauses in a $k$-CNF; satisfied by a random assignment with probability $(1 - 2^{-k})^m$.

Somewhat more interesting: take any $k$-CNF formula with not too many overlaps. Now randomize by choosing the sign patterns in each clause u.a.r.

A fixed assignment satisfies the formula with probability $(1 - 2^{-k})^m$. 

--- 

Local Lemma, probabilistic and algorithmic  

Jerri Nummenpalo
2*.2: Fewer top-level invocation

If $C$ and $D$ are two clauses, both which a top level invocation is being made for, then $\text{vbl}(C) \cap \text{vbl}(D) = \emptyset$ as after correcting $C$, the whole neighborhood of $C$ is satisfied and no completed top-level invocation leaves behind new violated clauses.
2*.3. Even fewer top-level invocations

**Claim.** The average number of top-level invocations is at most \( \frac{n}{(8k)} \). This is tight.

**Proof.** No variable can occur more than \( 2^{k-3} \) times, as this would exceed the neighborhood restriction.

Therefore there are \( m \leq n \cdot \frac{2^{k-3}}{k} \) clauses in total.

Each of them is violated in the beginning with probability \( 2^{-k} \), so the expected number of violated clauses in the beginning is at most \( \frac{n}{(8k)} \).
2*.3. **Even fewer top-level invocations (2)**

For tightness, consider a formula consisting of \( \frac{n}{k} \) independent components containing \( 2^{k-3} \) clauses over \( k \) variables each.

In each component, there is a violated clause in the beginning with probability \( \frac{1}{8} \). This is the probability with which a top-level invocation has to be made for this component.
2*.4. The procedure stops

Assume $F$ is satisfied by some assignment $\alpha^*$.

If we encounter clause $C$, the variables of $C$ are set according to $\alpha^*$ with positive probability. In every such step, the distance to $\alpha^*$ decreases by at least one.

Hence with positive probability $p$, we find a satisfying assignment in time $n = |\text{vbl}(F)|$ (not necessarily $\alpha^*$, another satisfying assignment could be found first)

After $\lambda \cdot n$ steps, the probability to have found no satisfying assignment is at most $(1 - p)^\lambda$, which goes to 0 for $\lambda \to \infty$. 