Math 126 Problem 4: Clever Covid-19 Testing

Spring 2027

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Monday, February 8, 2027 (Pacific time).**

- You are encouraged to discuss ideas and work with your classmates. However, you *must acknowledge* your collaborators at the top of each solution on which you collaborated with others and you *must write* your solutions independently.
- Your solutions to theory questions must be written legibly, or typeset in LaTeX or markdown.
- I recommend that you write your solutions in a Python notebook using Google Colab.
- You should submit your solutions as a *single PDF* via the assignment on Gradescope.

Grading: The point of the problem set is for *you* to learn. To this end, I hope to disincentivize the use of LLMs by *not* grading your work for correctness. Instead, you will grade your own work by comparing it to my solutions. This self-grade is due the Friday *after* the problem set is due, also on Gradescope.
Original notebook: Google Colab.

Clever Covid-19 Testing

One of the most important factors in controlling diseases like COVID-19 is testing. Before at-home kits became available, testing was expensive and slow. One way to make it cheaper was to test patients in *groups*. The biological samples from multiple patients (e.g., multiple nose swabs) are combined into a single test tube and tested for COVID-19 all at once. If the test comes back negative, we know everyone in the group is negative. If the test comes back positive, we do not know which patients in the group actually had COVID-19, so further testing would be necessary. There's a trade-off here, but it turns out that, overall, group testing can save on the total number of tests run.

Part A

Consider the following deterministic "two-level" testing scheme. We divide a population of n individuals to be tested into C groups of the same size. We then test each of these groups. For any group that comes back positive, we retest all members of the group individually. Show that there is a choice for C such that, if k individuals in the population have COVID-19, we can find all of those individuals with $\leq 2(nk)^{\frac{1}{2}}$ tests. You can assume k is known in advance (often it can be estimated accurately from the positive rate of prior tests). This is already an improvement on the naive n tests when $k < 25\% \cdot n$.

Part B

We can use randomness to do better. Consider the following scheme: Collect $q = \log_2(10n)$ nose swabs from each individual (I know... not pleasant). Then, repeat the following process q times: randomly partition our set of n individuals into $C = 2k$ groups, and test each group in aggregate. Once this process is complete, report that an individual "has COVID" if the group they were part of tested positive all q times. Report that an individual "is clear" if *any* of the groups they were part of tested negative. Show that, with probability 9/10, this scheme finds all truly positive patients and reports no false positives. Thus, we only require $2k * \log_2(10n) = O(k \log n)$ tests!