Problem Set 1

Pseudorandomness, Autumn 2023, University of Chicago
Instructor: William Hoza (williamhoza@uchicago.edu)

Submission. Solutions are due **Friday, October 13** at 5pm Central time. Submit your solutions through Canvas. You are encouraged to typeset your solutions in LaTeX. (Consider using Overleaf, a popular online LaTeX editor.) If you prefer, you may submit a photo of handwritten solutions instead.

The collaboration and resource policies below can also be found on the course webpage.

Collaboration. You are encouraged to collaborate with your classmates on problem sets, but you must adhere to the following rules.

- Before discussing a problem with a classmate, you must work on the problem on your own for at least 20 minutes.
- You must ultimately write your solution on your own. While writing your solution, you may not consult any notes that you took during a discussion with another classmate.
- In your write-up, you must list any classmates who contributed to your solution through discussion. The fact that A contributed to B's solution does not necessarily mean that B contributed to A's solution.

Permitted Resources for Full Credit. Beyond discussions with me and discussions with classmates as discussed above, you may use the course texts and any notes that might be posted in the "Course Timeline" section of the course webpage. If you wish to receive full credit on a problem, you may not use any other resources.

Permitted Resources for Partial Credit. If you wish, you may use outside resources (Wikipedia, ChatGPT, Stack Exchange, research papers, etc.) to solve a problem for partial credit. If you decide to go this route, you must make a note of which outside resources you used when you were working on each problem. You must disclose using a resource even if it was ultimately unhelpful for solving the problem. Furthermore, if you consult an outside resource while working on a problem, then you must not discuss that problem with your classmates.

Problem 1 (10 points)**.** Let k be an even positive integer, and let $Y_1, \ldots, Y_t$ be k-wise independent $[0,1]$-valued random variables. Let $X = \frac{1}{t}\sum_{i=1}^{t} Y_i$, let $\mu = \mathbb{E}[X]$, and let $\varepsilon > 0$. Prove that

$$\Pr[|X - \mu| \geq \varepsilon] \leq \left(\frac{k^2}{4t\varepsilon^2}\right)^{k/2}.$$

Note: We make no assumption about the distribution of an individual Y_i variable. We merely assume that every k of the Y_i variables are independent. For hints, see Proposition 3.28 and Problem 3.8 in Vadhan's text. (This problem is the same as Vadhan's Problem 3.8, except that a typo is corrected.)

Recall that a "CNF formula" is an AND of ORs of literals. For example,

$$f(x) = (x_1 \vee x_2) \wedge (\neg x_1 \vee x_3) \wedge (x_2 \vee x_4 \vee \neg x_5).$$

Meanwhile, a "majority-of-parities formula" computes some parities of the inputs and negations of parities and then takes a majority vote. For example,

$$f(x) = \mathsf{MAJ}(x_1 \oplus x_3, 1, x_1 \oplus x_7 \oplus x_6 \oplus 1, x_2 \oplus x_5, x_1 \oplus x_3).$$

Note that the same parity function (or negated parity function) might appear multiple times.

Problem 2 (10 points)**.** Design a deterministic polynomial-time algorithm that converts a given CNF formula into an equivalent majority-of-parities formula.

Hint: Let $H_n(x, y) = \sum_{i=1}^{n} \bigoplus_{j=1}^{n} x_{ij} \cdot y_j$. What is $\mathbb{E}_y[H_n(x, y)]$?

Problem 3 (5 points)**.** Let $g, h \colon \{0,1\}^n \to \mathbb{R}$ and let $f(x) = g(x) \cdot h(x)$. Prove that $L_1(f) \leq L_1(g) \cdot L_1(h)$, where $L_1(f)$ denotes the Fourier L_1 norm of f.

An $\mathsf{AC}^0[\oplus]$ formula is a tree in which each internal node ("gate") is labeled $\wedge$ (AND), $\vee$ (OR), or $\oplus$ (PARITY), and each leaf is labeled with a literal (a variable or its negation). The formula computes a Boolean function $f \colon \{0,1\}^n \to \{0,1\}$ in the natural, inductive way: each gate applies its label-function to its children, and $f(x)$ is the output value of the root gate. Each gate is permitted to have unbounded fan-in, i.e., an unbounded number of children. The *depth* of the formula is the length of the longest path from the root to a leaf. The *size* of the formula is the number of interal nodes, i.e., the number of $\wedge$, $\vee$, and $\oplus$ gates. Traditionally, people are especially interested in constant-depth polynomial-size $\mathsf{AC}^0[\oplus]$ formulas, but in this problem, we consider the unbounded-depth sublinear-size regime.

Problem 4 (10 points)**.** Show that there is an explicit ε-PRG for size-m $\mathsf{AC}^0[\oplus]$ formulas on n variables with seed length $O(m + \log(n/\varepsilon))$.

Hint: Work out what happens to Boolean Fourier analysis when we encode the values $\{0,1\}$ as $\{\pm 1\}$. Then work out what happens to the Fourier L_1 norm when we compose functions.
