1. The AUGMENTED INDEX problem (abbreviated as AUGIND) is the following. Alice receives a string $x \in \{0,1\}^n$ and Bob receives an index $i \in [n]$. In addition, Bob's input is augmented with the substring $x_1 \ldots x_{i-1}$ (denoted $x_{<i}$). Their goal is to compute x_i. In the randomized bounded error one-way communication model, where communication is allowed only from Alice to Bob, the communication complexity of AUGIND is $\Omega(n)$. In this question, we aim to prove this claim using information theory.

   One useful information theoretic tool to prove the lower bound on AUGIND is Fano's inequality stated as follows.

   Theorem 1. *Let X be a binary random variable. Let Y be another random variable and g be a function such that $g(Y) = X$ with probability at least $1-\delta$ for $\delta < 1/2$. Then, $H(X|Y) \leq H_2(\delta) := \delta \cdot \log_2(1/\delta) + (1-\delta)\log_2(1/(1-\delta))$.*

   (a) Show that if Bob is provided the substring $x_{i+1} \ldots x_n$ in addition to i and $x_{<i}$, then the one-way communication complexity of computing x_i is $O(1)$. **5 points**

   (b) Let X denote a bitstring of length n, where each bit is sampled uniformly and independently at random from $\{0,1\}$. Let I denote a uniformly random value from $[n] := \{1,2,\ldots,n\}$. Consider an arbitrary deterministic protocol π for AUGIND that has error probability at most 1/3 over this input distribution.

      i. Let M denote the message sent by Alice on input X. Justify the following equalities.

      $$I(X;M) = \sum_{i \in [n]} I(X_i; M | X_{<i}) = \sum_{i \in [n]} H(X_i|X_{<i}) - H(X_i|X_{<i}, M)$$

      15 points

      ii. Show that $I(X;M) \geq n \cdot (1 - H_2(2/3))$ and that $I(X;M) \leq |M|$. **20 points**
      Hint: Use Fano's inequality above.

      iii. Conclude that the distributional one-way communication complexity of AUGIND over the uniform distribution is $\Omega(n)$.
      What does this say about the randomized one-way communication complexity of AUGIND when the error bound is at most 1/3? **10 points**

2. We know (from the midterm and the second homework) that the function greater than (GT) has randomized communication complexity $O(\log n \log\log n)$ and that it has one way communication complexity $\Omega(n/\log n)$. In this problem, we better our understanding about this fundamental problem. Recall that GT: $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as $\text{GT}(x,y) = 1$ if and only if $x > y$.

   (a) Show that the greater than problem has randomized bounded error one-way communication complexity $\Omega(n)$. **20 points**

   (b) Show that the communication complexity of GT is $\tilde{O}(\sqrt{n})$ when the parties are allowed 2 rounds of communication. **20 points**

      You may assume that n is a perfect square.

      Hint: Recall that EQ has a randomized bounded error protocol with communication complexity $O(1)$. How many rounds does this protocol take?

3. Alice and Bob are given sets $X \subseteq [n]$ and $Y \subseteq [n]$ respectively. There is an additional promise that $|X \setminus Y| = 1$ and $Y \subset X$. Alice and Bob need to determine the unique element in $X \setminus Y$. Give an $O(\log n)$ bit communication protocol for this problem. **20 points**