The questions marked by ‘UG’ are meant to be submitted only by undergraduate students, those marked by ‘G’ are meant to be submitted only by graduate students, and the rest are common to both.

Question 1. (a) (UG) Let $X := X_1 X_2 \cdots X_n$ be a random variable, with $X_i$ being mutually independent. Let $M$ be jointly distributed with $X$, and $S \subseteq [n]$ be a random variable independent of $XM$ such that for every $i \in [n]$, we have $\Pr(i \in S) \leq p$. Prove that

$$I(X_S : M | S) \leq p I(X : M).$$

(b) Let $\Pi$ be any private-coin two-party communication protocol for computing the logical AND of two bits $x, y$ with worst-case error at most $\epsilon$, when Alice is given bit $x$ and Bob is given bit $y$. Let $XY$ be uniformly distributed over $\{00, 01, 10\}$. Prove that

$$I(X : M | Y) + I(Y : M | X) \in \Omega((1 - 2\epsilon)^2),$$

where $M$ is the message transcript of $\Pi$ on input $X, Y$.

Hint: Give a proof by contradiction.

Question 2. In this question, we show that linear codes achieve the capacity of a binary erasure channel with erasure probability $\alpha$.

For a linear code $C = \{Gw : w \in \mathbb{Z}_2^k\}$ generated by an $n \times k$ matrix $G$ over $\mathbb{Z}_2$, let $D : \{0, 1, \perp\}^n \to C \cup \{\perp\}$ be the following decoding function:

$$D(y) = \begin{cases} x & \text{if } y \text{ agrees with exactly one } x \in C \text{ on the unerased coordinates} \\ \perp & \text{otherwise} \end{cases}.$$

(a) For positive integers $l, m$ with $l \leq m$, let $M$ be an $m \times l$ matrix over $\mathbb{Z}_2$, chosen uniformly at random. Show that the probability that $M$ has rank less than $l$ over $\mathbb{Z}_2$ is at most $2^{l-m}$.

Hint: Imagine that $M$ is generated by sequentially picking the columns uniformly at random from $\mathbb{Z}_2^m$.

(b) Suppose we pick the generator matrix $G$ of the code $C$ as above, uniformly at random. For any set $J \subseteq [n]$, prove that the probability that decoding fails on some codeword, conditioned on the erasures being specified by set $J$, is less than $2^{k+|J|-n}$.

(c) (G) Suppose we pick the generator matrix $G$ of the code $C$ as above, uniformly at random. Show that when $k = \lfloor Rn \rfloor$ for any $R < 1 - \alpha$, the probability that decoding fails on some codeword is exponentially small in $n$. 
Question 3. Let $\mathbb{F}_q$ be the finite field with $q$ elements.

(a) Consider an $[n, k, n-k+1]_q$ Reed-Solomon code $C$. Let $W \in \mathbb{F}_q^k$ be uniformly random, and consider the random codeword $X := C(W)$. Prove that for any subset $S \subseteq [n]$ of size $k$, the random variable $X_S$ is uniformly random over $\mathbb{F}_q^k$.

(b) A secret sharing scheme is a probabilistic protocol for sharing a secret $a \in \mathbb{F}_q$ among $m$ parties in such a way that no subset of less than $r$ parties can infer any information about the secret by colluding, and any subset of at least $r$ parties can jointly recover the secret from their parts.

Formally, an $(m, r)$-secret sharing scheme for $\mathbb{F}_q$ is a function $f : \mathbb{F}_q \times [m] \times \Omega \rightarrow \mathbb{F}_q$, where $\Omega$ is the sample space of some random variable $X$, such that

(i) for any two “secrets” $a_1, a_2 \in \mathbb{F}_q$, and any subset $S := \{i_1, i_2, \ldots, i_l\} \subseteq [m]$ of size $l < r$, and any vector $(b_1, b_2, \ldots, b_l) \in \mathbb{F}_q^l$, $\Pr(\forall j \in [l], f(a_1, i_j, X) = b_j) = \Pr(\forall j \in [l], f(a_2, i_j, X) = b_j)$.

(ii) for every subset $S := \{i_1, i_2, \ldots, i_l\} \subseteq [m]$ of size $l \geq r$, and any vector $(b_1, b_2, \ldots, b_l) \in \mathbb{F}_q^l$, there is at most one $a \in \mathbb{F}_q$ such that $\Pr(\forall j \in [l], f(a, i_j, X) = b_j) > 0$.

In other words, the function $f$ maps a secret $a$ to a random vector $(f(a, 1, X), f(a, 2, X), \ldots, f(a, m, X))$ consisting of $m$ shares. The first condition ensures that the shares of any subset of at most $r - 1$ parties reveal no information about the secret. Furthermore, the second condition guarantees that any subset of at least $r$ shares uniquely determines the secret.

Use a Reed-Solomon code with appropriate parameters to design an $(m, r)$ secret sharing scheme.