Question 1

Let (Enc, Dec) be a (deterministic) encryption scheme with key length k and message length m. Suppose that $Enc(K, M)$ and $Enc(K, M')$ are more than $1/2$-statistically close for every two messages M, M'.

(a) Show that $Enc(K, M')$ is a possible encryption of M with probability more than $1/2$.

Solution: Let D be the distinguisher that accepts only those ciphertexts C that are possible encryptions of M. Then $D(Enc(K, M)) = 1$ with probability one. By statistical closeness, $D(Enc(K, M')) = 1$ with probability more than $1/2$.

(b) Fix a message M. Show that there exists a key K for which $Enc(K, M')$ is a possible encryption of M for more than half the messages M'.

Solution: In part (a) we showed that for every M', $Enc(K, M')$ is a possible encryption of M with probability more than $1/2$. This is therefore true for a random M', namely

$$\Pr_{K,M'}[Enc(K, M') \text{ is a possible encryption of } M] > 1/2.$$

Since this inequality holds for a random K, it must also be true for a specific K.

(c) Show that if $m > k$ then (Enc, Dec) is not an encryption scheme.

Solution: If (Enc, Dec) was an encryption scheme all 2^m encryptions of different messages under K would be distinct. By part (b), more than 2^{m-1} of them are possible encryptions of M. But there are at most 2^k possible encryptions of m. Therefore k has to be larger than $m - 1$, so it has to be at least as large as m.

Question 2

In Lecture 2 we showed that if $G\colon \{0,1\}^k \to \{0,1\}^n$ is an (s, ε)-pseudorandom generator of size t then

$$G'(K) = (\text{first } n-k \text{ bits of } G(K), G(\text{last } k \text{ bits of } G(K)))$$

is an $(s - t, 2\varepsilon)$-pseudorandom generator. Assuming that pseudorandom generators (with sufficiently good parameters) exist, show that there is a $G\colon \{0,1\}^k \to \{0,1\}^n$ that is an (s, ε)-pseudorandom generator but such that G' is not a $(\omega(n), 1.99\varepsilon)$-pseudorandom generator.

Solution: Let G^* be an (s, ε^*)-pseudorandom generator. We'll figure out what ε^* should be shortly but think of it as much smaller than ε. Let

$$G(X, R) = \begin{cases} \text{all zeroes}, & \text{if } R \text{ is the all zero string} \\ G^*(X), & \text{if not.} \end{cases}$$

Then G is an $(s, \varepsilon^* + 2^{-r})$-pseudorandom generator where r is the length of R: Unless an event of probability 2^{-r} happens, $G^*(X)$ and $G(X, R)$ are identically distributed. Let $\varepsilon = \varepsilon^* + 2^{-r}$.

We now estimate the advantage of the distinguisher Z for G' that accepts when the last n bits of the output are zeroes. This is a distinguisher of size 1. Z accepts $G'(X, R)$ in case the last r bits of $G(X, R)$ are zeroes. This happens either when R is the all zero string, or when R is not the all zero string but $G^*(X)$ ends with r zeroes. Since G^* is (s, ε^*) pseudorandom the probability of this event cannot be smaller than $2^{-r} - \varepsilon^*$ (as long as $s \geq 1$).

In summary, Z accepts $G'(X, R)$ with probability at least

$$2^{-r} + (1 - 2^{-r}) \cdot (2^{-r} - \varepsilon^*) \geq 2 \cdot 2^{-r} - (2^{-2r} + \varepsilon^*) = 2\varepsilon - (3\varepsilon^* + \varepsilon^2),$$

while it accepts a truly random string with probability 2^{-n}. If we choose $\varepsilon^* = \varepsilon^2$ the difference is at least 1.99ε as desired, as long as ε is less than $1/500$ and greater than $2^{-n/2}$.

Question 3

Let F_K be a pseudorandom function. Are these functions also pseudorandom? Assume the key length, input length, and output length are all equal to the security parameter n.

(a) The function $F'_K(x) = F_K(F_K(x))$.

Solution: Yes. We'll show that if F is an (s, q, ε)-PRF then F' is an $(s, q/2, \varepsilon + O(q^2 \cdot 2^{-n}))$-PRF.

First we show that when $R\colon \{0,1\}^n \to \{0,1\}^n$ is a random function, the function $(R \circ R)(x) = R(R(x))$ is statistically $O(q^2 \cdot 2^{-n})$-indistinguishable from a random function with respect to any q-query distinguisher.

$R \circ R$ is *not* a truly random function: Even when $n = 1$ it can never be the *NOT* function, while a truly random function can. A distinguisher that makes two distinct queries X_1, X_2 has nonzero advantage already. For a random function, the probability of a collision (i.e., $F(X_1) = F(X_2)$) is exactly 2^{-n}. For the function $R \circ R$ it is about twice that much: $R(X_1)$ and $R(X_2)$ are equal with probability 2^{-n}, but even if they are different $R(R(X_1))$ and $R(R(X_2))$ have another 2^{-n} chance of being equal.

To develop some intuition let's consider an arbitrary adaptive distinguisher that makes two distinct queries X_1 and X_2. The answer $R(R(X_1))$ to the first query is random. How about the second query X_2? As long as X_2 and $R(X_2)$ are both distinct from X_1 and $R(X_1)$, we would expect to see an independent random answer because it only uses evaluations of R at inputs that were not seen by the distinguisher before. Let's try to upper bound the probability that this happens.

Let's start with the probability that $X_2 = R(X_1)$. X_2 may depend on $R(R(X_1))$. However, as long as X_1 and $R(X_1)$ are distinct, $R(R(X_1))$ (and therefore X_2) is independent of $R(X_1)$, so $X_2 = R(X_1)$ are equal with probability 2^{-n}. Assuming this, $R(X_2)$ is a value of R at an "new" input so it is independent of X_1, $R(X_1)$, and X_2; the probability that it equals any of them is at most $3 \cdot 2^{-n}$.

We now make this argument formal. Let E_i be the event that $X_1, \ldots, X_i, R(X_1), \ldots, R(X_i)$ are all distinct (assuming that all queries X_i are distinct) and B_i be the event that X_i equals one of $R(X_1), \ldots, R(X_{i-1})$. By the rule for conditional probabilities,

$$\begin{aligned}\Pr[\text{NOT } E_i \mid E_{i-1}] &= \Pr[B_i \mid E_{i-1}] + \Pr[\text{NOT } E_i \text{ AND NOT } B_i \mid E_{i-1}] \\ &\leq \Pr[B_i \mid E_{i-1}] + \Pr[\text{NOT } E_i \mid \text{NOT } B_i, E_{i-1}].\end{aligned}$$

Conditioned on E_{i-1}, $R(R(X_1)), \ldots, R(R(X_{i-1}))$ are independent of $R(X_1), \ldots, R(X_{i-1})$, so X_i is also independent of them. The probability that it equals any of them is at most $i \cdot 2^{-n}$. Conditioned on any fixing of the values $X_1, \ldots, X_{i-1}, R(X_1), \ldots, R(X_{i-1})$ that are different from X_i, the probability that $R(X_i)$ equals any of them is at most $(2i-1) \cdot 2^{-n}$. Therefore

$$\Pr[\text{NOT } E_i \mid E_{i-1}] \leq i \cdot 2^{-n} + (2i-1) \cdot 2^{-n} = (3i-1) \cdot 2^{-n}.$$

Adding up these probabilities we get that

$$\begin{aligned}\Pr[\text{NOT } E_q] &= \Pr[\text{NOT } E_1] + \Pr[\text{NOT } E_2 \text{ AND } E_1] + \cdots + \Pr[\text{NOT } E_q \text{ AND } E_{q-1}] \\ &\leq \Pr[\text{NOT } E_1] + \Pr[\text{NOT } E_2 \mid E_1] + \cdots + \Pr[\text{NOT } E_q \mid E_{q-1}] \\ &= O(q^2 \cdot 2^{-n}).\end{aligned}$$

Conditioned on E_q, the view of the distinguisher consists of q independent random values. so $R \circ R$ is a $(\infty, q, O(q^2 \cdot 2^{-n}))$-pseudorandom function.

It remains to show that $F'_K = F_K \circ F_K$ is pseudorandom. Suppose it is not (s, q', ε')-pseudorandom. Then $F_K \circ F_K$ is $(s, q', \varepsilon' - O(q'^2 \cdot 2^{-n}))$-distinguishable from $R \circ R$ by some D'. But then F_K are R must be $(s, 2q', \varepsilon' - O(q'^2 \cdot 2^{-n}))$-distinguishable by this D: Simulate D', and when it makes a query to $F \circ F$, make a query to F, record the answer and query F on it again.

(b) The function $F'_{K,K'}(x, y) = F_K(x) + F_{K'}(y)$, where K and K' are independent.

Solution: No. F' satisfies the identity $F'(x, y) + F'(x', y) + F'(x, y') + F'(x', y') = 0$ for any x, y, x', y' As long as x, x' and y, y' are distinct, for a truly random function these four values are independent so the probability a random function satisfies the same identity is 2^{-n}. So the two can be distinguished with very high advantage by this 4-query distinguisher.

(c) **(Optional)** The function $F'_K(x) = F_K(x + K)$.

Solution: F' may not be a pseudorandom function. Let H be an $(s, q+1, \varepsilon)$ pseudorandom function and

$$F(x) = \begin{cases} H(x), & \text{if } x \neq K \\ 0^n, & \text{if } x = K. \end{cases}$$

We will show that F is then $(s - O(qn), q, 2\varepsilon + 2q2^{-n})$ pseudorandom, but F' is not even $(2, 1 - 2^{-n})$ pseudorandom.

The second part is easy: The distinguisher makes a single query at zero and accepts if the output is all zeros. This is always true for F but has probability 2^{-n} for a random function.

We now argue that F is a PRF if H is. Suppose that F and H are distinguishable. The only way a distinguisher can tell the two apart is if it queries K. But then the distinguisher learns the secret key, so it can distinguish H from a random function.

More precisely, let D be a q-query distinguisher that tells F from a random function with advantage more than $2\varepsilon + 2q2^{-n}$. If H is (s, q, ε)-pseudorandom, D must distinguish F from H with advantage more than $\varepsilon + 2q2^{-n}$. Conditioned on D not querying K, D^F and D^H are identically distributed, so D^H must query K with at least this probability.

Now consider the distinguisher D' that simulates D, remembering all queries $K_1, \ldots, K_q$ made by it and accepts if $H_{K_i}(Y)$ matches the value of its oracle at Y for some i and for a random Y. Assuming D'^H queries K it must accept, so D'^H accepts with probability at least $\varepsilon + 2q2^{-n}$. On the other hand D'^R accepts if and only if $H_{K_i}(Y) = R(Y)$ for some i. Since Y is independent of the queries made by D^R, Y is different from all queries of D except with probability $q2^{-n}$. If this is the case then $R(Y)$ is independent of $K_1, \ldots, K_q$ and Y so the probability that $H_{K_i}(Y) = R(Y)$ for some i is at most $q2^{-n}$. In summary, D' distinguishes H from a random function with advantage ε.

Question 4

In our setup of private-key encryption we assumed that Alice and Bob share identical copies of the random key. Now suppose that Alice's and Bob's copies of the key are noisy. Specifically, the keys K_A, K_B are elements of the group $\mathbb{Z}_{2^k}$ (i.e., integers modulo 2^k) that are individually uniformly distributed such that the difference $K_A - K_B$ is in the range from $-2^b + 1$ to 2^b modulo 2^k (where $b < k$).

(a) Give a definition of a noisy key encryption scheme.

Solution: A noisy-key encryption scheme is a pair of circuits Enc, Dec such that for every pair of keys K_A and K_B such that $K_A - K_B$ is between $-2^b + 1$ and 2^b modulo 2^k, $Dec(K_B, Enc(K_A, M)) = M$ for every message M.

The scheme is perfectly secure if for a random K_A, and every pair of messages M, M', $Enc(K_A, M)$ and $Enc(K_A, M')$ are identically distributed.

(b) Show that if the message length is less than $k - b$ then there exists a perfectly secure noisy key encryption scheme.

Solution: Messages are represented in $\mathbb{Z}_q$ by padding them with zeroes. The encryption is $K_A + M$ and the decryption of a message C is obtained by rounding $C - K_B$ to the closest possible message, rounding down in case of a tie. Then $Dec(K_B, Enc(K_A, M)) = M + (K_A - K_B)$, which is in the range between $M - 2^b + 1$ and $M + 2^b$. If the message length is less than $k - b$ all these intervals are disjoint so decryption always recovers M.

The scheme is perfectly secure because when K_A is random encryptions are distributed as uniform random strings.

(c) Show that if the message length is $k - b$ or more then perfect security is no longer possible. Show how to construct a message-simulatable (computationally secure) scheme assuming the existence of a pseudorandom generator. Provide a proof of security.

Solution: We show that if perfect security with these parameters was possible, we could use it to construct an ordinary perfectly secure encryption with shorter keys than messages. To implement the ordinary scheme, Alice and Bob interpret their shared key K as the $k-b-1$ most significant bits of an element of $\mathbb{Z}_{2^k}$. Alice sets the other bits of her key randomly and subtracts 2^b-1 to obtain K_A. Bob pads K with zeroes to obtain K_B. This distribution of keys is consistent with the requirements in part (a), so if a perfectly secure noisy key encryption scheme with these parameters existed so would an ordinary scheme with message length $k-b$ and key length $k-b-1$. We showed this was impossible in lecture 1.

If a pseudorandom generator was available, Alice could use the noisy scheme to communicate a $(k-b)$-bit long key K to Bob and then use it to encrypt messages using any scheme for long messages (which can be based on a pseudorandom generator). Let $(NEnc, NDec)$ be the noisy scheme from part (b) and (Enc, Dec) be any message-simulatable scheme for long messages, such as the one from Lecture 2. To encrypt M, Alice chooses a random key K and sends $(NEnc(K_A, K), Enc(K, M))$ to Bob. To decrypt a ciphertext of the form (N, C), Bob first recovers $K = NDec(K_B, N)$ then computes $M = Dec(K, C)$. By the message-simulatability of $(NEnc, NDec)$, the first part of a ciphertext is distributed as a random string for every K, so it is independent of K. Security then follows from the simulatability of (Enc, Dec) (with the same parameters).