
Homework 2 – Due Tuesday, February 10 in Class

Reminders: Collaboration is permitted on homeworks, but you must (a) list your collaborators on the problem set and (b) **understand and write your solutions yourself**.

Clarity is an important component of the grade on your problem sets. Poorly explained solutions will not receive full marks, even if they are correct.

Useful facts from probability:

- (*Linearity of expectation*) For every two real-valued random variables A and B , $E(A + B) = E(A) + E(B)$.
- For every two *independent*, real-valued random variables A and B , $E(AB) = E(A)E(B)$.
- (*The “union bound”*) For any collection of events E_1, \dots, E_k , the probability of their union (i.e. the probability that at least one of them occurs) is bounded above by the sum of the individual event probabilities, that is:

$$\Pr(E_1 \cup \dots \cup E_k) \leq \sum_{i=1}^k \Pr(E_i).$$

(Exercise: Use the union bound to prove that in a uniformly random string of n bits, the probability that there is some run of $\log_2(n) + 1$ or more consecutive zeros is at most $1/2$.)

Problems to be handed in:

1. (**Probability**) Suppose that $X \sim \text{Bin}(n, p)$ (for example, X may count the number of heads that occur in n independent flips of a coin, where each flip comes up heads with probability p).
 - (a) Use linearity of expectation to prove that $E(X) = np$ (here $E(\cdot)$ denotes expectation).
 - (b) Prove that the variance of X is $\text{Var}(X) = np(1 - p)$. (Recall that $\text{Var}(X) = E(X^2) - (E(X))^2$.)
 - (c) Let p and q be constants such that $0 < q < p < 1$.
 - i. Prove that there is some constant $c < 1$, depending only on p and q , such that for all $0 < i < nq$, we have $\frac{\Pr(X=i-1)}{\Pr(X=i)} < c$.
 - ii. Prove that $\Pr(X < nq) \leq \frac{1}{1-c} \Pr(X = nq)$.
 - iii. We say a function $f(n)$ is exponentially small in n if there are constants $\alpha > 0$ and n_0 such that $f(n) < \exp(-\alpha n)$ for all $n > n_0$.
Prove that $\Pr(X = nq)$ is exponentially small in n (say, using Stirling’s approximation).

iv. Prove that for any constant δ , $\Pr(|X - np| \geq \delta n)$ is exponentially small in n .

Note: Bounds of the form derived here are called *concentration inequalities*, since they show that the variable X is “concentrated” close to its mean with high probability.

(d) Suppose there is a function $G : \{0, 1\}^n \rightarrow \{0, 1\}^{n+2}$ such that for all n , with probability $1/n$ over the choice of seeds s , the substring “010” occurs at most $n/9$ times in the output $G(s)$. Prove that G is not a pseudorandom generator.

(Hint: Divide subsequences into three categories. You may assume, for simplicity, that n is a multiple of 3.)

2. (Groups)

(a) Show that the set $\{0, 1\}^n$, equipped with the binary operation \oplus is a mathematical group (Def 7.9 in KL book).

(b) Let G be a group with binary operation \circ . Suppose that group operations (\circ and inversion) can be performed efficiently. Consider the symmetric key encryption scheme with $\mathcal{M} = \mathcal{K} = \mathcal{C} = G$, where $\text{Enc}(m; k) = m \circ k$. Show that the encryption scheme is perfectly secret (KL, Definition 2.1).

3. (**Hybrid Arguments**) You are given two pseudorandom generators G_0, G_1 with expansion functions $\ell_0(n), \ell_1(n)$, and you wish to create a new pseudorandom generator G' by concatenating the outputs of G_0 and G_1 . Thus G' takes a seed s of even length n and splits it into two pieces s_0 and s_1 of length $n/2$. The output is $G'(s_0, s_1) = G_0(s_0) \| G_1(s_1)$, where “ $\|$ ” denotes concatenation of binary strings.

In this problem, you will prove that G' is indeed a pseudorandom generator, using a hybrid argument.

(a) What is the expansion function $\ell'(n)$ of G' as a function of ℓ_0 and ℓ_1 ?

(b) The first step is to define three probability distributions $\text{Exp}_0, \text{Exp}_1, \text{Exp}_2$ where Exp_0 is a random string of length $\ell'(n)$ (denoted $U_{\ell'(n)}$), Exp_2 is the output of G' on a random seed $s = s_0, s_1$ of total length n , and Exp_1 is a distribution “between” the Exp_0 and Exp_2 . Suggest an intermediate distribution Exp_1 (there is more than one natural choice).

(c) Using your suggestion for Exp_1 , prove that any algorithm D' which can distinguish $G'(s)$ from $U_{\ell'(n)}$ with advantage $\epsilon(n)$ can be used either (i) to construct a PPT algorithm D_0 that distinguishes $G_0(s)$ from random with advantage at least $\epsilon(2n)/2$, or (ii) to construct a PPT algorithm D_1 that distinguishes $G_1(s)$ from random with probability advantage at least $\epsilon(2n)/2$. Conclude that G' is a secure p.r.g. as long as both G_0 and G_1 are secure.

(d) Does the construction produce a p.r.g. if G_0 and G_1 are the same?

(e) Does the construction always produce a p.r.g. if we use the same seed for both generators, that is, if we set $G'(s) = G_0(s) \| G_1(s)$?