Notes on Pairwise Uniform Bits

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

Definition 1 (k-junta)**.** Let f be a function on $\{0,1\}^n$. We say that f is a *k-junta* if there exist indices $i_1, \dots, i_k \in [n]$ and there exists a function g on $\{0,1\}^k$ such that for every $x \in \{0,1\}^n$, we have

$$f(x) = g(x_{i_1}, \dots, x_{i_k}).$$

We will present a PRG that fools 2-juntas (with error zero). The correctness of the PRG is based on the following lemma.

Lemma 1. *Let Y and Z be $\{0,1\}$-valued random variables. Assume that Y, Z, and $Y \oplus Z$ are uniformly distributed over $\{0,1\}$. Then (Y,Z) is uniformly distributed over $\{0,1\}^2$.*

Proof. For each $a, b \in \{0,1\}$, let $p_{ab} = \Pr[(Y,Z) = (a,b)]$. Then

$$p_{00} + p_{01} = \Pr[Y = 0] = 1/2 \qquad \text{because } Y \text{ is uniform} \qquad (1)$$
$$p_{00} + p_{10} = \Pr[Z = 0] = 1/2 \qquad \text{because } Z \text{ is uniform} \qquad (2)$$
$$p_{01} + p_{11} = \Pr[Z = 1] = 1/2 \qquad \text{because } Z \text{ is uniform} \qquad (3)$$
$$p_{00} + p_{11} = \Pr[Y \oplus Z = 0] = 1/2 \qquad \text{because } Y \oplus Z \text{ is uniform.} \qquad (4)$$

Subtracting (3) from (1) gives $p_{00} = p_{11}$. Plugging into (4), this implies $p_{00} = p_{11} = 1/4$. Plugging $p_{00} = 1/4$ into (1) and (2) gives $p_{01} = p_{10} = 1/4$. □

Now we are ready to present the PRG.

Theorem 1. *There is a PRG $G \colon \{0,1\}^s \to \{0,1\}^n$ that fools 2-juntas with error 0 and seed length $s = \lfloor \log n \rfloor + 1$.*

Proof. Let $I_1, \dots, I_n$ be distinct, nonempty subsets of $[s]$. (Such sets exist because $2^s > n$.) The PRG is given by

$$G(x) = \left(\bigoplus_{i \in I_1} x_i, \bigoplus_{i \in I_2} x_i, \dots, \bigoplus_{i \in I_n} x_i \right).$$

For the analysis, consider any two output coordinates of G, say $j, k \in [n]$ where $j \neq k$. Sample $X \sim U_n$ and let Y and Z be the j-th and k-th output bits of G, namely

$$Y = \bigoplus_{i \in I_j} X_i$$
$$Z = \bigoplus_{i \in I_k} X_i.$$

Because the sets $I_1, \dots, I_n$ are nonempty, each individual output bit such as Y or Z is uniformly distributed over $\{0,1\}$. Now let us look at the XOR of two output bits:

$$Y \oplus Z = \left(\bigoplus_{i \in I_j} X_i \right) \oplus \left(\bigoplus_{i \in I_k} X_i \right) = \bigoplus_{i \in I_j \Delta I_k} X_i,$$

where $I_j \Delta I_k$ denotes the "symmetric difference" of I_j and I_k, namely $I_j \Delta I_k = (I_j \setminus I_k) \cup (I_k \setminus I_j)$. Since I_j and I_k are distinct, the symmetric difference $I_j \Delta I_k$ is nonempty, and therefore $Y \oplus Z$ is uniformly distributed over $\{0,1\}$. Therefore, by the lemma, (Y,Z) is uniformly distributed over $\{0,1\}^2$. □

Terminology: Let $X_1, \ldots, X_n$ be random variables. We say that $X_1, \ldots, X_n$ are *pairwise independent* if every two of them are independent, i.e., for every two indices $i, j \in [n]$ with $i \neq j$, the two random variables X_i and X_j are independent. We say that $X_1, \ldots, X_n$ are *pairwise uniform* if they are pairwise independent and each X_i is distributed uniformly over its domain. A PRG that fools 2-juntas with error 0 is called a *pairwise uniform generator*.

Remark 1. *People are not always careful to distinguish the concept of pairwise independence from the concept of pairwise uniformity. Sometimes people say something like "sample pairwise independent bits* $X_1, \ldots, X_n$*" when they technically mean "sample pairwise* uniform *bits* $X_1, \ldots, X_n$*."*