Appendix A

Statistical Distinguishers

As usual in these lecture notes we restrict ourselves to finite probability spaces.

A.1 Distingishing Distributions by a Test

Let A be a finite probability space with two probability distributions P_0 and P_1. Accordingly for a real valued function $\Delta\colon A \longrightarrow \mathbb{R}$ we have the mean values (or expectations)

$$\mu_i = \sum_{a \in A} \Delta(a) \cdot P_i(a).$$

For $\varepsilon > 0$ we call Δ an ε**-distinguisher** of P_0 and P_1 if

$$|\mu_1 - \mu_0| \geq \varepsilon.$$

That is, the expectations of Δ with respect to P_0 and P_1 differ considerably.

> Note the analogy with the common statistical test scenario where we decide whether a sample deviates from an assumed distribution by comparing mean values.

This notion has an obvious analogue for bit valued functions (or binary attributes) $\Delta\colon A \longrightarrow \mathbb{F}_2$. Here

$$\mu_i \;=\; \sum_{a \in \Delta^{-1}(1)} P_i(a) \;=\; P_i(\Delta^{-1}(1))$$

is the probability that $\Delta(a) = 1$ for a randomly chosen $a \in A$. Thus

$$\mu_1 - \mu_0 = P_1(\Delta^{-1}(1)) - P_0(\Delta^{-1}(1)).$$

The "test" Δ ε-distinguishes between the distributions P_1 and P_0 if the probabilities for $\Delta(a) = 1$ with respect to these two distributions differ by at least ε.

> Note that the notion "test" just means "function". However in the present context it suggests a role that this function plays. A similar remark also holds for the notion "randomize".

We may "randomize" our test by more generally considering a function

$$\Delta\colon A \times \Omega \longrightarrow \mathbb{F}_2$$

where Ω is a finite probability space from which we take an additional random input ω, and then consider the probabilities μ_i that $\Delta(a, \omega) = 1$,

$$\mu_i = \frac{1}{\#A \cdot \#\Omega} \cdot \#\{(a, \omega) \in A \times \Omega \mid \Delta(a, \omega) = 1\}.$$

A.2 Testing Bitsequences

A statistical test for bitsequences of length r is simply a Boolean function $\Delta\colon \mathbb{F}_2^r \longrightarrow \mathbb{F}_2$, a probabilistic statistical test is a function

$$\Delta\colon \mathbb{F}_2^r \times \Omega \longrightarrow \mathbb{F}_2$$

where Ω is a finite probability space.

We want to distinguish between random bitsequences $u \in \mathbb{F}_2^r$, and bitsequences that arise from a "generator map"

$$G\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^r$$

that transforms a randomly chosen $x \in \mathbb{F}_2^n$ (called "seed") to a bitsequence $G(x) \in \mathbb{F}_2^r$. This sequence $G(x)$, if it passes our tests, may qualify as a pseudorandom sequence. In this test scenario the reference distribution P_0 is the uniform distribution on $\mathbb{F}_2^r$,

$$P_0(u) = \frac{1}{2^r} \quad \text{for all } u \in \mathbb{F}_2^r\,.$$

We want to compare it with the induced distribution

$$P_1(u) = \frac{1}{2^n} \cdot \#\{x \in \mathbb{F}_2^n \mid G(x) = u\}\,.$$

Or, somewhat more generally, if G is defined on a subset $A \subseteq \mathbb{F}_2^n$ only,

$$P_1(u) = \frac{1}{\#A} \cdot \#\{x \in A \mid G(x) = u\}\,.$$

A probabilistic statistical test $\Delta : \mathbb{F}_2^r \times \Omega \longrightarrow \mathbb{F}_2$ ε-distinguishes between random bitsequences $u \in \mathbb{F}_2^r$ and sequences generated by $G\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^r$ if

$$|\mu_1 - \mu_0| \geq \varepsilon$$

where

$$\mu_0 \;=\; \frac{1}{2^r \cdot \#\Omega} \cdot \#\{(u,\omega) \in \mathbb{F}_2^r \times \Omega \mid \Delta(u,\omega) = 1\}$$

is the probability that the test assigns the value 1 to a random bitsequence $u \in \mathbb{F}_2^r$, and

$$\mu_1 \;=\; \frac{1}{2^n \cdot \#\Omega} \cdot \#\{(x,\omega) \in A \times \Omega \mid \Delta(G(x)),\omega) = 1\}$$

is the probability that the test yields the value 1 for a bitstring generated by a random seed $x \in A$.

Examples

We want to distinguish sequences generated by a map $G\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^r$ from random sequences (by deterministic tests, that is $\#\Omega = 1$).

Example 1

First an extremely simple example with the test function

$$\Delta\colon \mathbb{F}_2^r \longrightarrow \mathbb{F}_2, \quad \Delta(u) = \begin{cases} 1 & \text{if } \#\{i \mid u_i = 1\} \geq \frac{r}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

That is Δ decides on the majority of ones in the sequence u. Then obviously $\mu_0 = \frac{1}{2}$.

Case 1a: Let $n = 1$ and $G\colon \mathbb{F}_2 \longrightarrow \mathbb{F}_2^r$ be defined by

$$\begin{aligned} G(0) &= (0,0,0,\ldots), \\ G(1) &= (1,1,1,\ldots). \end{aligned}$$

Then also $\mu_1 = \frac{1}{2}$, yielding $\mu_1 - \mu_0 = 0$. Thus Δ is not an ε-distinguisher for any $\varepsilon > 0$.

Case 1b: We keep the definition of $G(1)$ but change the definition of $G(0)$ to

$$G(0) = (1,0,1,0,1,\ldots).$$

Then $\Delta(G(0)) = \Delta(G(1)) = 1$, hence $\mu_1 = 1$, yielding $\mu_1 - \mu_0 = \frac{1}{2}$. Thus Δ is an ε-distinguisher for $0 < \varepsilon \leq \frac{1}{2}$.

Example 2

For a serious example we consider sequences generated by a linear feedback shift register $G\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^r$ of length n where $2n < r \leq 2^n - 1$. We know that the output of G is distinguished by a low linear complexity $\lambda(u) \leq n$. Therefore we use

$$\Delta\colon \mathbb{F}_2^r \longrightarrow \mathbb{F}_2, \quad \Delta(u) = \begin{cases} 1 & \text{if } \lambda(u) < \frac{r}{2}, \\ 0 & \text{if } \lambda(u) \geq \frac{r}{2}, \end{cases}$$

as test . Since $n < \frac{r}{2}$ this yields

$$\mu_1 \;=\; \frac{1}{2^n} \cdot \#\{x \in \mathbb{F}_2^n \mid \Delta(G(x)) = 1\} \;=\; 1.$$

For arbitrary sequences $u \in \mathbb{F}_2^r$ we know from Theorem 3 that we may expect $\lambda(u) \approx \frac{r}{2}$. A more precise statement follows from the frequency count in Proposition 11:

$$k \;:=\; \#\{u \in \mathbb{F}_2^r \mid \lambda(u)) \leq \frac{r-1}{2}\} \;=\; 1 + \sum_{l=1}^{\lfloor\frac{r-1}{2}\rfloor} 2^{2l-1} \;=\; \frac{1}{2} + \frac{1}{2} \cdot \sum_{l=0}^{\lfloor\frac{r-1}{2}\rfloor} 4^l.$$

Case 2a: Let r be even. Then $\lfloor\frac{r-1}{2}\rfloor = \frac{r}{2} - 1$, and

$$k \;=\; \frac{1}{2} + \frac{1}{2} \cdot \frac{4^{r/2} - 1}{3} \;=\; \frac{1}{2} + \frac{1}{6} \cdot (2^r - 1) \;=\; \frac{1}{3} + \frac{1}{6} \cdot 2^r,$$

$$\mu_0 \;=\; \frac{1}{2^r} \cdot k \;=\; \frac{1}{6} + \frac{1}{3 \cdot 2^r} \;\leq\; \frac{1}{3} \quad \text{for } r \geq 1.$$

Case 2b: Let r be odd. Then $\lfloor\frac{r-1}{2}\rfloor = \frac{r-1}{2}$, and

$$k \;=\; \frac{1}{2} + \frac{1}{2} \cdot \frac{4^{(r+1)/2} - 1}{3} \;=\; \frac{1}{2} + \frac{1}{6} \cdot (2^{r+1} - 1) \;=\; \frac{1}{3} + \frac{1}{3} \cdot 2^r,$$

$$\mu_0 \;=\; \frac{1}{2^r} \cdot k \;=\; \frac{1}{3} + \frac{1}{3 \cdot 2^r} \;\leq\; \frac{1}{2} \quad \text{for } r \geq 1.$$

Hence in any case we have

$$\mu_1 - \mu_0 \;\geq\; \frac{1}{2} \quad \text{for } r \geq 1.$$

Thus Δ is an ε-distinguisher for $0 < \varepsilon \leq \frac{1}{2}$, distinguishing between LFSR sequences and random sequences.