A NOTE ON COMPUTATIONAL INDISTINGUISHABILITY [1]

Oded Goldreich
Computer Science Dept.
Technion, Haifa, Israel

ABSTRACT — We show that following two conditions are equivalent:

1) The existence of pseudorandom generators.

2) The existence of a pair of efficiently constructible distributions which are *computationally indistinguishable* but *statistically very different.*

KEYWORDS: Computational Complexity, Randomness, Algorithms.

[1] Appeared as TR-602 of the CS Dept. Technion. This research was supported by grant No. 86-00301 from the United States - Israel Binational Science Foundation (BSF), Jerusalem, Israel.

1. INTRODUCTION

A fundamental notion in complexity theory is that of probability distributions which are computationally indistinguishable. This notion originates from the pioneering work of [Goldwasser, Micali 82] and was presented in full generality in the fundamental work of [Yao 82]. Loosely speaking, two probability distributions are computationally indistinguishable if no efficient algorithm can "tell them apart". Namely, the output distribution of every efficient algorithm is oblivious of whether the input is taken from the first distribution or from the second distribution.

Clearly, every two distributions which are statistically close are also computationally indistinguishable. Using a counting argument one can show that the converse does not hold [Goldreich, Krawczyk 89]; namely, there exist two distributions which are statistically very different yet are computationally indistinguishable. However, these distributions are not efficiently constructible. A fundamental question in this area is *whether there exist two efficiently constructible distributions which are computationally indistinguishable, yet statistically very different.* In the sequel we refer to a positive answer to the above question as the *non-triviality of computational indistinguishability.*

The (hypothetical) existence of pseudorandom generators, introduced and developed by [Blum, Micali 82] and [Yao 82], imply the non-triviality of computational indistinguishability. A pseudorandom generator is an efficient (deterministic) algorithm which stretches short seeds into longer output sequences such that the output distribution on a uniformly chosen seed is computationally indistinguishable from a uniform distribution. It is easy to see that the pseudorandom output distribution and the uniform distribution constitute a non-trivial case of computational indistinguishability.

We conclude that the existence of pseudorandom generators is a sufficient condition for the non-triviality of computational indistinguishability. In this note we prove that this condition is also a necessary one. Namely, the non-triviality of computational indistinguishability implies the existence of pseudorandom generator.[2]

The notion of false entropy, introduced by [Impagliazzo, Levin, Luby 89], plays a central role in our proof. This notion yields a special case of two efficiently constructible distributions which are statistically different yet computationally indistinguishable. We first show that the non-triviality of computational indistinguishability implies the existence of false entropy and conclude by employing a result of [Impagliazzo, Levin, Luby 89] which states that the existence of false entropy implies the existence of pseudorandom generators.

---

[2] We stress that it is not currently known whether pseudorandom generators do exist. Recently, it has been shown that the existence of one-way functions (another widely believed complexity assumption) is a necessary and sufficient condition for the existence of pseudorandom generators (cf. [Impagliazzo, Levin, Luby 89] and [Hastad 89]).

2. FORMAL SETTING

For a formal setting we consider sequences of probability distributions (called ensembles) instead of single probability distributions.

Definition 1: An *ensemble* $X = \{X_n\}_{n \in \mathbf{N}}$ is a sequence of random variables each ranging over binary strings. (Sometimes we omit $n \in \mathbf{N}$ from the notation.)

Definition 2: An ensemble $X = \{X_n\}_{n \in \mathbf{N}}$ is *polynomial-time constructible* if there exists a probabilistic polynomial-time algorithm, S, such that $X_n = S(1^n)$. (On input 1^n algorithm S has output distribution X_n.)

Definition 3: Two ensembles $X = \{X_n\}$ and $Y = \{Y_n\}$ are said to be *statistically different* if there exists a constant $c > 0$ and an integer N such that for all $n \geq N$

$$\sum_\alpha |\mathrm{Prob}(X_n = \alpha) - \mathrm{Prob}(Y_n = \alpha)| > \frac{1}{n^c}.$$

The ensembles $X = \{X_n\}$ and $Y = \{Y_n\}$ are *statistically close* if for every $c > 0$ there exists an integer N such that for all $n \geq N$

$$\sum_\alpha |\mathrm{Prob}(X_n = \alpha) - \mathrm{Prob}(Y_n = \alpha)| < \frac{1}{n^c}.$$

Two ensembles which are not statistically close are not necessarily statistically different. Statistically close ensembles constitute an uninteresting case of polynomially indistinguishable ensembles (see Definition 4).

Definition 4 [Goldwasser, Micali 82, and Yao 82]: Two ensembles $X = \{X_n\}$ and $Y = \{Y_n\}$ are *polynomially indistinguishable* if for every (probabilistic) polynomial-time algorithm, A, and every $c > 0$ there exists an integer N such that for all $n \geq N$

$$|\mathrm{Prob}(A(X_n) = 1) - \mathrm{Prob}(A(Y_n) = 1)| < \frac{1}{n^c}.$$

Definition 5: An ensemble $X = \{X_n\}_{n \in \mathbf{N}}$ is called *uniform* if there exists a function $l : \mathbf{N} \mapsto \mathbf{N}$, such that for every n and every $\alpha \in \{0,1\}^{l(n)}$:

$$\mathrm{Prob}(X_n = \alpha) = 2^{-l(n)}$$

If $l(n) = n$, for all n, then X is call *the uniform ensemble*. The uniform ensemble is denoted by $U = \{U_n\}$; namely, for every $\alpha \in \{0,1\}^n$: $\mathrm{Prob}(U_n = \alpha) = 2^{-n}$.

Definition 6 [Yao 82]: An ensemble $X = \{X_n\}$ is called *pseudorandom* if it is polynomially indistinguishable from some uniform ensemble.

Definition 7 [Blum, Micali 82]: A deterministic polynomial-time algorithm G is called a *pseudorandom generator* if the following two conditions hold

1) For every $s \in \{0,1\}^*$: $\quad |G(s)| \ > \ |s|$.

2) The ensemble $\{G(U_n)\}_{n \in \mathbf{N}}$ is pseudorandom.

Theorem: *The following two conditions are equivalent:*

1) *There exists a pseudorandom generator,*

2) *There exists a pair of polynomial-time constructible ensembles which are statistically different yet polynomially indistinguishable.*

3. PROOF OF THE THEOREM

To see that condition (1) implies (2), let G be a pseudorandom generator, and $l : \mathbf{N} \rightarrow \mathbf{N}$ be a function so that $\{G(U_n)\}_{n \in \mathbf{N}}$ and $\{U_{l(n)}\}_{n \in \mathbf{N}}$ are polynomially indistinguishable. Since $|G(\alpha)| > |\alpha|$, it follows that $l(n) > n$. Also, $G(U_n)$ must concentrate on strings of length $l(n)$ (i.e. $\mathrm{Prob}(|G(U_n)| = l(n)) > \frac{2}{3}$) and hence $l(n)$ can be computed with very high probability in $poly(n)$-time. Namely, there exists a probabilistic polynomial-time algorithm L such that $\mathrm{Prob}(L(1^n) = l(n)) > 1 - 2^{-n}$. Let Y_n denote the following two step random process: first assign $l - L(1^n)$ and next assign Y_n a string chosen uniformly in $\{0,1\}^l$. Then $Y = \{Y_n\}$ is polynomial-time constructible and satisfies $\mathrm{Prob}(Y_n \in \{0,1\}^{l(n)}) \ 1 - 2^{-n}$ and $\mathrm{Prob}(Y_n = \alpha | Y_n \in \{0,1\}^{l(n)}) = 2^{-l(n)}$ for every $\alpha \in \{0,1\}^{l(n)}$. Hence $\{G(U_n)\}$ and $\{Y_n\}$ (being polynomial-time constructible, statistical different, and polynomially indistinguishable) satisfy condition (2).

To see that condition (2) implies condition (1) we use the notion of false entropy introduced in [Impagliazzo, Levin, Luby 89].

Definition 8: A polynomial-time constructible ensemble $F = \{F_n\}$ is said to have *false entropy* (is a *false entropy ensemble*) if there exists a polynomial-time constructible ensemble $D = \{D_n\}$, such that F and D are polynomially indistinguishable and D has higher entropy than F. Namely, there exists a constant c 0 and an integer N such that for all $n \geq N$

$$\mathrm{Ent}(D_n) \ \geq \ \mathrm{Ent}(F_n) \ + \ \frac{1}{n^c}$$

where Ent is the entropy functional assigning to each random variable X its entropy

$$- \sum_{\alpha} \mathrm{Prob}(X = \alpha) \cdot \log_2(\mathrm{Prob}(X = \alpha)).$$

The proof follows immediately from the subsequent two lemmas.

Lemma 1: Let $\{X_n\}$ and $\{Y_n\}$ be a pair of ensembles as in condition (2) of the Theorem. Then there exists a false entropy ensemble.

Lemma 2 [Impagliazzo, Levin, Luby 89]: If there exists a false entropy ensemble then there exists a pseudorandom generator.

3.1 Proof of Lemma 1

A construction which proves the lemma is obtained by letting $F_n = (0, X_n)$ with probability $1/2$ and $F_n = (1, Y_n)$ with probability $1/2$. The ensemble $\{D_n\}$ used to demonstrate that $\{F_n\}$ has false entropy is $D_n = (B, X_n)$ with probability $1/2$ and $D_n = (B, Y_n)$ with probability $1/2$, where B is uniformly distributed over $\{0,1\}$ independently of all other random variables. It is simpler, however, to verify the validity of the more complex construction given below.

Let $c > 0$ be such that for all sufficiently large n we have

$$\sum_{\alpha} |\text{Prob}(X_n = \alpha) - \text{Prob}(Y_n = \alpha)| > \frac{1}{n^c}$$

Define $\bar{X}_n$ to be n^{2c+1} independent copies of X_n, and $\bar{Y}_n$ to be n^{2c+1} independent copies of Y_n. Clearly, $\{\bar{X}_n\}$ and $\{\bar{Y}_n\}$ are both polynomial-time constructible. Standard technique can be used to show that $\{\bar{X}_n\}$ and $\{\bar{Y}_n\}$ are polynomial-time indistinguishable (e.g., consider "hybrids" H_n^i composed of i independent copies of X_n followed by $n^{2c+1} - i$ independent copies of Y_n). $\bar{X}$ and $\bar{Y}$ are statistically very different; namely:

$$\sum_{\bar{\alpha}} |\text{Prob}(\bar{X}_n = \bar{\alpha}) - \text{Prob}(\bar{Y}_n = \bar{\alpha})| > 1 - 2^{-n}.$$

We now apply the above construction to $\bar{X}_n$ and $\bar{Y}_n$ (instead of to X_n and Y_n). Formally, $\bar{F}_n$ equals $(0, \bar{X}_n)$ with probability $1/2$ and $(1, \bar{Y}_n)$ otherwise. $\bar{D}_n$ equals $(B, \bar{X}_n)$ with probability $1/2$ and $(B, \bar{Y}_n)$ otherwise. Clearly, $\bar{F}_n$ and $\bar{D}_n$ are polynomial indistinguishable, while $\bar{D}_n$ has higher entropy than $\bar{F}_n$ (as the first bit of $\bar{D}_n$ is independent of the rest, while in $\bar{F}_n$ the first bit is determined with very high probability by the rest).

Remark: An analogous argument can be applied directly to F_n and D_n. The first bit of F_n can be predicted with non-negligible advantage (rather than almost determined) from the rest.

3.2 Proof of Lemma 2 - Sketch

The proof originates from [Impagliazzo, Levin, Luby 89]. A sketch is presented here for the sake of self-containment.

Let $F = \{F_n\}$ be a false entropy ensemble and $D = \{D_n\}$ the ensemble used to demonstrate this property. Let S be a probabilistic polynomial-time algorithm satisfying $S(1^n) = F_n$ (such an algorithm exists since F_n is polynomial-time constructible). Let $t(n)$ be a bound on the running time of $S(1^n)$. We may view $S(1^n)$ as selecting at random a sequence of $t(n)$ bits, denoted r, and then evaluating $f(r)$, where f is a polynomial-time computable function. Let R_n be uniform over $\{0,1\}^{t(n)}$, then $F_n = f(R_n)$.

Suppose $\mathrm{Ent}(D_n) > \mathrm{Ent}(F_n) + \frac{1}{n^c}$, and let $e(n) \stackrel{\text{def}}{=} \mathrm{Ent}(F_n)$. Shorthand $t = t(n)$, $e = e(n)$ and let $m = (n^c \cdot t)^{O(1)}$ and $l = m(t-e) - m^{2/3}t - n$. Consider the following three ensembles:

(1) $h,\ f(r_1)...f(r_m),\ h(r_1...r_m)$

(2) $h,\ f(r_1)...f(r_m),\ s$

(3) $h,\ D_n...D_n,\ s$ (this ensemble contains m independent copies of D_n)
where (in the appropriate ensembles) $r_1...r_m$ are uniformly selected each in $\{0,1\}^t$, the string s is uniformly selected in $\{0,1\}^l$, and h is a randomly selected function from a family of universal$_2$ hash functions mapping $m \cdot t$-bit strings to l-bit strings. Such families have the (defining) property that every two elements of the domain are mapped (uniformly and) in a pairwise independent manner by a function chosen uniformly in the family. We require the family to have succinct representation: the functions in the family should be representable by strings of length n^c, where c is a constant independent of n.

It is easy to see that the ensembles (2) and (3) are polynomially indistinguishable (as $F = \{f(R_n)\}$ and D are polynomially indistinguishable). To show that the ensembles (1) and (2) are statistically close, observe that for most sequences $r_1, r_2, ..., r_m$, the number of pre-images of $f(r_1)\cdot f(r_2)\cdots f(r_m)$ is at least $2^{m(t-e)-m^{2/3}\cdot t} = 2^{l+n}$ (as the average logarithm of pre-image size for each $f(r_i)$ is $t-e$ and m independent repetitions of an experiment are unlikely to deviate from the expectation by more than $\sqrt{m}$ times the maximal value). It can be shown that most functions chosen from the above family map a subset of size 2^{l+n} almost uniformly onto $\{0,1\}^l$ (see [Impagliazzo, Levin, Luby 89] for the non-trivial technical details). Hence, ensembles (1) and (2) are statistically close. We conclude that ensembles (1) and (3) are polynomially indistinguishable.

We now show that the entropy of ensemble (3) is greater than the number of random bits used in the construction of ensemble (1). The entropy of ensemble (3) is at least

$$|h| + m\cdot(e + \frac{1}{n^c}) + m\cdot(t-e) - m^{2/3}\cdot t - n \;>\; |h| + m\cdot t + m\cdot\frac{1}{n^c} - 2m^{2/3}\cdot t.$$

With a suitable choice of m (e.g. $m = (3n^c\cdot t)^3$) this is substantially more than $|h| + m\cdot t$ which is the number of bits used in the construction of ensemble (1).

Finally, applying a suitable hashing function on ensembles (1) and (3) yields a pseudorandom generator G. I.e.,

$$G(h', h, r_1...r_m) = h'\cdot h'(h, f(r_1)...f(r_m), h(r_1...r_m)),$$

where h' is a hashing function chosen from a suitable class. The output distribution of G is polynomially indistinguishable from the distribution $h'\cdot h'(h, D_n...D_n, s)$ and the latter is statistically close to a uniform ensemble.

ACKNOWLEDGEMENT

I am most grateful to Johan Hastad for his exposition of the proof of Lemma 2. I also wish to thank Hugo Krawczyk, Leonid Levin and Mike Luby for discussion concerning polynomial indistinguishability and pseudorandom generators. Final thanks to the anonymous referee for helpful comments.

REFERENCES

[1] Blum, M., and Micali, S., "How to Generate Cryptographically Strong Sequences of Pseudo-Random Bits", *SIAM Jour. on Computing*, Vol. 13, 1984, pp. 850-864. First version in *FOCS 1982*.

[2] Goldreich, O., and H. Krawczyk, "Sparse Pseudorandom Distributions", *Crypto89* proceedings, to appear.

[3] Goldwasser, S., and S. Micali, "Probabilistic Encryption", *JCSS*, Vol. 28, No. 2, 1984, pp. 270-299. Previous version in *STOC 1982*.

[4] Hastad, J., "Pseudo-Random Generators with Uniform Assumptions", preprint, 1989.

[5] Impagliazzo, R., L.A. Levin, and M. Luby, "Pseudorandom Generation from One-Way Functions", *Proc. of the 21st ACM Symp. on Theory of Computing* (*STOC*), 1989, pp. 12-24.

[6] Yao, A.C., "Theory and Applications of Trapdoor Functions", *Proc. of the 23rd IEEE Symp. on Foundation of Computer Science* (*FOCS*), 1982, pp. 80-91.