

Lecture Note 02

Date: 27.09.2007

1 Introduction to indistinguishability

1.1 Mathematics recalls

Definition 1 : Two random variables X, Y are **independent** if for all x in the image of X and all y in the image of Y , the events $X = x$ and $Y = y$ are independent :

$$Pr[X = x \wedge Y = y] = Pr[X = x]Pr[Y = y]$$

Equivalently, X and Y are independent if and only if their joint distribution is equal to the product of their individual distributions.

Definition 2 : Let X_1, \dots, X_n be a collection of random variables. We say that X_1, \dots, X_n are **pairwise independent** if $\forall i, j = 1, \dots, n (i \neq j)$, the variables X_i and X_j are independent.

Definition 3 : We say that X_1, \dots, X_n are **mutually independent** if $\forall x_1 = X_1, \dots, x_n = X_n$, we have

$$Pr[X_1 = x_1 \wedge \dots \wedge X_n = x_n] = \prod_{i=1}^n Pr[X_i = x_i]$$

More generally, for $k = 2, \dots, n$, we say that X_1, \dots, X_n are k -wise independent if any k of them are mutually independent.

Definition & Notations 4 : Let I be a countable index set. An ensemble indexed by I is a sequence of random variable indexed by I . Namely, any $X = \{X_i\}_{i \in I}$, where each X_i is a random variable, is an ensemble indexed by I .

- $X = \{X_n\}_{n \in \mathbb{N}}$ has each X_n ranging over strings of length $poly(n)$.
- $X = \{X_w\}_{w \in \{0,1\}^*}$ has each X_w ranging over strings of length $poly(|w|)$.

Notation : Thereafter, we will use the following notation :

$$Pr[A(X_n) = 1] = \sum_x Pr[X_n = x]Pr[A(x) = 1]$$

1.2 Definition of indistinguishability

Definition 5 : Two distributions $D_1 = (U, P_1)$ and $D_2 = (U, P_2)$ are said to be computationally indistinguishable if no efficient procedure can differentiate them.

More concretely, considering an adversary A , we say that A does not distinguish the two distributions D_1 and D_2 if and only if

$$\forall x \in U, |Pr(A(x_1) = 1) - Pr(A(x_0) = 1)|$$

is negligible.

where $Pr(A(x_b) = 1)$ ($b \in 0, 1$) is the probability for A to guess correctly that the element x is taken from D_b .

Application to cryptography : For a perfect encryption scheme we wish:

$$|Pr[Enc(1) = 1]Pr[Enc(0) = 1]|$$

is negligible.

2 Definitions

2.1 Polynomial-Time Indistinguishability

2.1.1 Definition

Now that we know the notion of a set of random variables, let us introduce the notion of indistinguishability. Indistinguishability characterizes the quantity of Information of a set, that we can deduce from the knowledge of the other set.

Let give us the formal definition:

- Two sets, $X := \{X_n\}_{n \in N}$ and $Y := \{Y_n\}_{n \in N}$, are indistinguishable in polynomial time if for every probabilistic polynomial-time algorithm D , every positive polynomial $p(\cdot)$, and all sufficiently large $n \in N$,

$$|Pr[D(X_n, 1^n) = 1] - Pr[D(Y_n, 1^n) = 1]| \leq \frac{1}{p(n)}$$

This means in more mathematical statement:

$$\forall D \in PPT \forall p \in R^+[X] \exists n_0 \in N \forall n \in N$$

$$n > n_0 \Rightarrow |Pr[D(X_n, 1^n) = 1] - Pr[D(Y_n, 1^n) = 1]| \leq \frac{1}{p(n)}$$

PPT : Set of the probabilistic polynomial-time algorithms

$R^+[X]$: Set of the positive polynomials

- Two ensembles, $X := \{X_w\}_{w \in S}$ and $Y := \{Y_w\}_{w \in S}$, are indistinguishable in polynomial time if for every probabilistic polynomial-time algorithm D , every positive polynomial $p(\cdot)$, and all sufficiently long $w \in S$,

$$|Pr[D(X_w, w) = 1] - Pr[D(Y_w, w) = 1]| \leq \frac{1}{p(|w|)}$$

2.1.2 Examples

See the exercises in the slides of the course on Indistinguishability.

2.1.3 Properties

Transitivity Let $X := \{X_n\}_{n \in N}$, $Y := \{Y_n\}_{n \in N}$ and $Z := \{Z_n\}_{n \in N}$, three sets of random variables. If X and Y are indistinguishable and Y and Z too, then X and Z are indistinguishable.

Proof: use the triangular inequality

2.2 Indistinguishability by Repeated Sampling

The next definition, is a generalisation of the previous one:

Two sets, $X := \{X_n\}_{n \in N}$ and $Y := \{Y_n\}_{n \in N}$ are indistinguishable by polynomial-time sampling if for every probabilistic polynomial-time algorithm D , every positive polynomials $m(\cdot)$ and $p(\cdot)$, and all sufficiently large $n \in N$:

$$|Pr[D(X_n^1, X_n^2, \dots, X_n^{m(n)}) = 1] - Pr[D(Y_n^1, Y_n^1, \dots, Y_n^{m(n)}) = 1]| \leq \frac{1}{p(n)}$$

where X_n^1 through $X_n^{m(n)}$ and Y_n^1 through $Y_n^{m(n)}$ are independent random variables, with each X_n^i identical to X_n and Y_n^i identical to Y_n .

2.3 Efficiently Constructible Ensembles

An ensemble $X := \{X_n\}_{n \in N}$ is said to be polynomial-time-constructible if there exists a probabilistic polynomial-time algorithm S such that for every n , the random variables $S(1^n)$ and X are identically distributed.

3 Hybrid Technique

In this section, we will demonstrate a very important theorem establishing a strong link between the notion of Indistinguishability in polynomial-time and Indistinguishability by Repeated sampling.

3.1 Theorem

Let $X := \{X_n\}_{n \in N}$ and $Y := \{Y_n\}_{n \in N}$ be two polynomial-time-constructible ensemble, and suppose that X and Y are indistinguishable in polynomial time. Then X and Y are indistinguishable by polynomial-time sampling.

3.2 Proof

Idea of the proof We will use the classical proof method of the contraposition. Indeed, to show that:

$$A \Rightarrow B$$

We prove that:

$$\neg B \Rightarrow \neg A$$

since the two assertions are equivalent.

In this case, we suppose that we have an efficient algorithm that distinguishes two sets of random variables (X and Y) using several samples. We will show then, that we are able to build a probabilistic polynomial-time algorithm that distinguish X and Y .

Let us introduce H_n^k In order to help us, we introduce $m + 1$ random variables:

$$\forall 0 \leq k \leq m H_n^k = (X_n^1, \dots, X_n^k, Y_n^{k+1}, \dots, Y_n^m)$$

where X_n^1 through $X_n^{m(n)}$ and Y_n^1 through $Y_n^{m(n)}$ are independent random variables, with each X_n^i identical to X_n and Y_n^i identical to Y_n .

We have of course:

$$H_n^m = (X_n^1, \dots, X_n^m)$$

$$H_n^0 = (Y_n^1, \dots, Y_n^m)$$

The proof As we said, we begin with the choice of an algorithm D such that, for every $n \in N$:

$$\delta(n) = |Pr[D(X_n^1, X_n^2, \dots, X_n^{m(n)}) = 1] - Pr[D(Y_n^1, Y_n^1, \dots, Y_n^{m(n)}) = 1]| > \frac{1}{p(n)}$$

We can rewrite this by using the variables H_n^k , indeed:

$$\delta(n) = |Pr[D(H_n^m) = 1] - Pr[D(H_n^0) = 1]| > \frac{1}{p(n)}$$

So, D distinguish H_n^m and H_n^0 . Set $m = m(n)$. At this point, we define D' the distribution depending on the parameter α , an integer k randomly taken in $[0, m]$, k samples of X_n and $m - 1 - k$ samples of Y_n :

$$D'(\alpha) = (X_n^1, \dots, X_n^k, \alpha, Y_n^{k+2}, \dots, Y_n^m)$$

We see that D' is very near from H_n^k . Of course, D' is the distribution which we want to show that distinguish X and Y

In order to prove that D' distinguish X and Y , we have to calculate: $Pr[D'(X_n, 1^n) = 1]$

$$\begin{aligned} Pr[D'(X_n) = 1] &= Pr[(X_n^1, \dots, X_n^k, X_n, Y_n^{k+2}, \dots, Y_n^m) = 1] \\ &= \sum_{k=0}^{m-1} Pr[X_n = k] Pr[(X_n^1, \dots, X_n^k, X_n, Y_n^{k+2}, \dots, Y_n^m) = 1] \end{aligned}$$

Now, the position of X_n (k in fact) is uniformly choosen in $[0, m - 1]$ so:

$$\forall k \in [0, m - 1] Pr[X_n = k] = \frac{1}{m}$$

and by using the definition of H_n^k , we have finally:

$$Pr[D'(X_n) = 1] = \frac{1}{m} \sum_{k=0}^{m-1} Pr[H_n^{k+1} = 1]$$

We have also a similar result for Y :

$$Pr[D'(Y_n) = 1] = \frac{1}{m} \sum_{k=0}^{m-1} Pr[H_n^k = 1]$$

It's now time to calculate the difference:

$$\begin{aligned} |Pr[D'(X_n) = 1] - Pr[D'(Y_n) = 1]| &= \frac{1}{m} \sum_{k=0}^{m-1} Pr[H_n^{k+1} = 1] - Pr[H_n^k = 1] \\ &= \frac{1}{m} (Pr[H_n^m = 1] - Pr[H_n^0 = 1]) = \frac{\delta(n)}{m} \end{aligned}$$

and since $\delta(n) > \frac{1}{p(n)}$, we proved that X and Y are distinguishable. We achieve the contraposition and the theorem at the same time.