# Lecture 11. Information theoretical argument

An interesting technique to prove lower bound is to use some information theoretical argument. Since introduced by [CSWY01], *information complexity* has been studied and applied to prove strong lower bounds of communication complexity. In this lecture, we will first introduce/review some basic notions such as entropy and mutual information, and then show how to use them to prove a linear lower bound for randomized communication complexity of Disjointness function [BYJKS04], one of the most important functions in communication complexity.

## Review of basic notions in information theory

Let’s first review some basic concepts in information theory. All these can be found in the standard textbook such as [CT06] (Chapter 1). Suppose that is a discrete random variable with alphabet X and probability mass function . Following the standard notation in information theory, we sometimes also use the capital letter to denote the distribution. By writing , we mean to draw a sample from the distribution . The entropy of a discrete random variable is defined by

(Here we use the convention that .) One intuition is that entropy measures the amount of the uncertainty of a random variable. The more entropy has, the more uncertain it is. In particular, if then is fixed/deterministic. The maximum entropy for any random variable on X is :

with upper bound achieved by the uniform distribution.

Sometimes we have more than one random variable. If is a pair of random variables on , distributed according to a joint distribution , then the total entropy is simple:

We can also define the conditional entropy , namely the expected entropy of the conditional distribution . Thus, . Another basic fact is the *chain rule*:

The chain rule also works with conditional entropy:

And if we have more variables, it also works:

,

where the inequality is usually referred to as the subadditivity of entropy. The relative entropy of two distributions and is

For a joint distribution , the mutual information between and is

It is a good exercise to verify that the above equality holds. But the latter one has a clear explanation: the mutual information is the amount of uncertainty of minus that when is known. In this way, it measures how much contains the information of . It is not hard to verify that it’s symmetric: , and that it’s nonnegative:

. That is,

The conditional mutual information is defined by

It is not hard to see that

and

**Lemma 1.1**. If , , and , and ’s are independent, then

*Proof*.

// : independence of ’s

// : subadditivity of entropy

## Linear lower bound for the Disjointness function

The information complexity is an approach to prove lower bounds for communication complexity. The basic idea is that for some functions, any small-error protocol has to contain enough information about the input. Thus the communication cost, namely the length of the protocol, is also large.

Now we focus on Disjointness function, where both Alice and Bob have input set and iff . We sometimes use subscript to denote the set . Denote by the -error private-coin communication complexity of . We want to prove the following.

**Theorem 2.1**.

First, we define a distribution on inputs. Again, we’ll use capital letters to denote both the random variables and the distribution. We also introduce a third random variable with the sample space . The joint distribution is defined by , where each with probability . That is, ’s are independent for different ’s, and each is distributed according to .

Now take a private-coin protocol with minimum worst-case communication cost among all protocols that have -error in the worst case. Consider the messages over all rounds, namely the entire transcript of the conversation. Denote it by , and its length (i.e. the number of bits) by . Note that since the input is a random variable, and the protocol is randomized, the induced transcript is also a random variable. By the relation among mutual information, entropy and the size of sample space, we have

Note that both and can be decomposed into n bits: and . Each and are also random variables, and note that for different *i* are independent. Thus we can apply Lemma 1.1 and have

Since , we have

For each fixed , we design a worst-case ε-error private-coin protocol , for the function , as follows. Note that the protocol should work for all possible inputs , not only those in the support of .

: On input ,

1. Generate from .
2. Run protocol on and output the answer.

**Lemma 2.2.**

1. is private-coin,
2. has -error in the worst case.

*Proof*.

1. Note that once is fixed, then each is a product distribution (over Alice and Bob’s spaces). Indeed, if , then where is the uniform distribution on . If , then .
2. Since only puts weight on 0-inputs, , and thus is correct on iff is correct on . Thus the error probability of on is the average (over and ) error prob of Γ on . Since Γ has error probability at most on all inputs, so is on all its possible inputs .
3. does nothing but invoking on the same input distribution .

Now we have

Next we show that any worst-case ε-error private-coin protocol for has to contain a constant amount of information about (conditioned on ).

**Lemma 2.3**. For any worst-case ε-error private-coin protocol for with transcript ,

*Proof*. Denote the transcript for input by . First, we show that **if the protocol doesn’t contain enough information, then is close to both and . Thus is close to** . Consider the Hellinger distance:

We have

// For

// Cauchy-Schwartz and Triangle

Then, we will show that **any communication protocol enjoys the property that for any** . **Thus**  **is close to .**Theproperty is proven by writing down and for any fixed . It’s not hard to see that is product of some function of and some function of . So by switching the function for and that for , we have that

Then by the multiplicative nature of the definition of , one gets .

**Now the contradiction comes: , so should be far from .** The best probability gap to distinguish and is , which is upper bounded in terms of by the following fact: .

Putting everything together, we get

## References

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar, **An information statistics approach to data stream and communication complexity**. Journal of Computer and System Sciences, 68(4), pp. 702-732, 2004.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao, **Informational complexity and the direct sum problem for simultaneous message complexity**, in Proceedings of the 42nd Annual Symposium on Foundations of Computer Science, pp. 270-278, 2001.

[CT06] Thomas Cover, Joy Thomas. **Elements of Information Theory**, Second Edition, *Wiley InterScience*, 2006.