On the Distributional Complexity of Disjointness

A.A.Razborov

Steklov Mathematical Institute,
Vavilova, 42, GSP-1, 117966,
Moscow, USSR

Abstract

We prove that the distributional communication complexity of the predicate "DISJOINTNESS" with respect to a very simple measure on inputs is $\Omega(n)$.

1 Introduction

The following concept of the *ε-error probabilistic communication complexity* $C_\varepsilon(A)$ of a binary predicate $A(x,y)$ was introduced by A.Yao [Y1]. Assume that two infinitely powerful computers want to evaluate the predicate $A(x,y)$ in the situation when the first computer possesses x and the second possesses y (x and y are binary strings of length n). They do this interchanging messages between each other. Both computers are allowed to flip a coin. At the end of the communication for each x and y they must output the correct value of $A(x,y)$ with probability $\geq 1-\varepsilon$. The complexity is measured by the expected number of communications in the worst case. For more details we refer to [Y1] or [BFS].

In the paper [Y2] A.Yao suggested an approach to estimating $C_\varepsilon(A)$ from below and gave an application of this approach. It is based upon the notion of the ε-error distributional communication complexity $D_\varepsilon(A)$ of a binary predicate $A(x,y)$. This notion, in turn, was generalized in [BFS] to the concept of the *ε-error distributional communication complexity* $D_\varepsilon(A|\mu)$ *under a probabilistic measure* μ *on inputs* ($D_\varepsilon(A)$ is just $D_\varepsilon(A|\mu)$ with uniform μ). This concept is somewhat dual to $C_\varepsilon(A)$: now the computers run a *deterministic* protocol and are required to output the correct value of $A(x,y)$ everywhere except for at most ε-fraction (with respect to the measure μ) of inputs. It was proved in [Y2] for uniform μ and generalized in [BFS] to arbitrary μ that $C_\varepsilon(A)\geq\frac{1}{2}D_{2\varepsilon}(A|\mu)$ for any A,μ and $\varepsilon>0$.

Several authors studied these complexity measures for the predicate "DISJOINTNESS". Let DIS_n denote this predicate (we will remind its definition below). L.Babai, P.Frankl and J.Simon [BFS] proved that $D_\varepsilon(DIS_n|\mu)\geq\Omega(n^{1/2})$ where μ is some measure on inputs and $\varepsilon>0$ is sufficiently small. This implies $C_\varepsilon(DIS_n)\geq\Omega(n^{1/2})$ for any $\varepsilon<1/2$. The measure μ in [BFS] is product of a measure on columns and a measure on rows. In comparison with the lower bound it was also proved in [BFS] that $D_\varepsilon(DIS_n|\mu)\leq O(n^{1/2}\log n)$ for any μ which is *product* measure in the above sense i.e. is product of a measure on columns and a measure on rows ($\varepsilon>0$ is arbitrary). Then B.Kalyanasundaram and G.Schnitger [KS] established the best possible lower bound $C_\varepsilon(DIS_n)\geq\Omega(n)$ ($\varepsilon<1/2$) for the ε-error probabilistic communication complexity of "DISJOINTNESS".

Probably all lower bound for $C_\varepsilon(A)$ known prior to the paper [KS] were actually lower bounds for the distributional complexity $D_\varepsilon(A|\mu)$ with some suitable measure μ. But the proof in [KS] involves complicated arguments related to the Kolmogorov complexity and this results in the fact that the measure μ "meant" in the proof depends on the protocol given by "the adversary".

The aim of this note is to show that the "random coupling" arguments of B.Kalyanasundaram and G.Schnitger can be carried over to yield the lower bound $D_\varepsilon(DIS_n|\mu)\geq\Omega(n)$ for a very simple measure μ described below (this does *not* contradict to the result from [BFS] since our μ is *not* a product measure). The proof involves only classical probabilistic arguments and does not appeal to the Kolmogorov complexity.

2 The result

We will identify throughout binary predicates and their characteristic 0-1 matrices. Given a predicate $A(x,y)$ ($x\in X, y\in Y$), *the ε-error distributional complexity $D_\varepsilon(A|\mu)$ under a probabilistic measure μ on inputs* (i.e on $X\times Y$) is the minimal possible length of a deterministic communication protocol which, given the random input $(\mathbf{x},\mathbf{y})$ according to the measure μ, outputs $a_{\mathbf{xy}}$ with probability $\geq 1-\varepsilon$ [Y2,BFS]. Fix the notation DIS_n for the so-called *disjointness matrix* DIS_n over